VIRTUAL-AUDIO AIDED VISUAL SEARCH ON A DESKTOP DISPLAY

Clayton Rothwell

Infoscitex Corporation
4027 Colonel Glenn Hwy, Suite 210
Dayton, OH 45431, USA
crothwell@infoscitex.com

Griffin Romigh and Brian Simpson

Air Force Research Laboratory
2610 Seventh Street, Area B, Bldg 441
Wright-Patterson Air Force Base, USA
griffin.romigh@us.af.mil
brian.simpson.4@us.af.mil

ABSTRACT

As visual display complexity grows, visual cues and alerts may become less salient and therefore less effective. Although the auditory system's resolution is rather coarse relative to the visual system, there is some evidence for virtual spatialized audio to benefit visual search on a small frontal region, such as a desktop monitor. Two experiments examined if search times could be reduced compared to visual-only search through spatial auditory cues rendered using one of two methods: individualized or generic head-related transfer functions. Results showed the cue type interacted with display complexity, with larger reductions compared to visual-only search as set size increased. For larger set sizes, individualized cues were significantly better than generic cues overall. Across all set sizes, individualized cues were better than generic cues for cueing eccentric elevations (>± 8 °). Where performance must be maximized, designers should use individualized virtual audio if at all possible, even in small frontal region within the field of view.

1. INTRODUCTION

The complexity and clutter of visual displays, such as dynamic interactive map displays, has increased over the last decade. Visual alerts on maps, for instance, have to compete with: the map symbols, colors, contrast and motion that are represented in the map. In other words, a designer is challenged to create a visual pop-out effect in an already colorful and moving scene. Increases in visual complexity may reduce the effectiveness of the visual alerts that have previously been effective in simpler, less cluttered maps. Spatialized auditory alerts can point to a location in space, such as the location of a particular visual object on a map display, as an act of deixis [1]. Yet the visual modality has better spatial resolution than the auditory modality, so it has often been the case the visual alerts alone have been used to alert different spatial locations on the monitor. This research investigated if auditory spatial alerts can aid visual search in a small frontal spatial region and what their utility is as a function of the complexity of the visual display (here in terms of the number of visual distractors).

Auditory spatial acuity is relatively worse than visual acuity. Visual vernier acuity averages around 5 arc seconds (or 0.0014° ; [2]). Auditory acuity has been estimated in a variety of ways. Measurements of minimum audible angle (MAA; [3, 4]) suggest resolution around 1°. Measurements of localization error in the free field find ~11° absolute angular error (after removing front-back confusions) and many studies suggest that localization errors increase for virtual audio (See [5] for a discussion). Within virtual audio there are differences in accuracy as well; individualized spatial audio can be near free-field performance whereas generic (i.e., non-individualized) spatial audio is worse [6]. Still, auditory cues have been shown to benefit visual search tasks despite the auditory system's resolution, such as a pilot searching for nearby aircraft traffic on the ground [7] or in the air [8], or in the task paradigm of *aurally aided visual search*, a visual search in 360° space surrounding the searcher (e.g., [9, 10]). In the research of Perrott et al. [9] and Bolia et al. [10], a spherical search space comprised of 277 loudspeakers placed approximately 15° apart surrounded the participant. Each loudspeaker has a cluster of 4 LEDs that can be independently lit. A target was displayed along with varying numbers of distractors (i.e., different set sizes) and the target was present on every trial. The target was one of two possible configurations of LEDs and the participant's task was to find and identify the target configuration. Accuracy and response time were measured as a function of the availability of a cue and/or the type of cue and the set size.

The aurally aided visual search paradigm has been used to show the benefit (i.e., reduction in search times) of an audio cue compared to visual only search. Additionally, this research has been used to discriminate between the effectiveness of different types of auditory cues. For example, researchers have used free-field sounds played from the target location and compared those to non-individualized virtual sound sources for that location (e.g., [10]). They found that both free-field and non-individualized virtual cues provided a benefit, but non-individualized virtual cues did not provided as much of a benefit as free-field cues did. Additional research has manipulated free-field and virtual auditory cues further by changing cue reliability / precision, measured the impact of hearing protection devices on spatial hearing, and investigated potential for multi-sensory cues to facilitate search times [11, 12, 13]. The reduction in search times from spatial audio cues is unsurprising in part because of the discretization of the visual search area and the possibility for visual targets to appear outside of the field of view. For instance, an auditory cue that was localized within 11° of the target would orient a searcher within one or two visual stimuli from the target. Also, an auditory cue to a region outside of the current field of view would naturally improve search times. Neither of these circumstances hold true in a small spatial region represented by a computer monitor. It is unclear if being

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at http://creativecommons.org/licenses/by-nc/4.0

DOI: 10.21785/icad2016.034

oriented within 11° of the intended location would reduce search times in a search area that may only subtend 50° x 30°, has many visual stimuli within that region, and is completely within the field of view. Yet other research has shown that free-field spatial auditory cues can speed target identification even in the frontal region and even in the absence of distractors (two frontal locations were measured: 0° and 15°; [14]), suggesting perhaps virtual auditory cues could speed search times.

The experiments presented here tested the utility of auditory spatial cues in a visual search task in a small frontal region. All audio cues were virtual, yet two different cue types were tested: cues created with individualized head-related transfer functions (HRTFs) and cues created with generic HRTFs (measured on a Knowles Electronics Mannequin for Acoustics Research, or KEMAR). This manipulation was to test if the previously found differences in localization accuracy between individualized and generic virtual audio would matter in this small region [6], similar to the effects found in previous work investigating spatial precision of free-field auditory cues on visual search [15]. Moreover, the comparison between generic-HRTFs and individualized-HRTFs was motivated by a practical issue: if auditory cues were shown to reduce search times and generic HRTFs were no different from individualized cues, then displays with spatialized auditory alerts could be deployed with one set of generic HRTFs rather than needing to measure HRTFs for every user and switch the HRTFs being used. The effectiveness of auditory cues was measured for many different levels of visual complexity, i.e., the number of distractors in the visual scene. Two experiments investigated different ranges of set size.

2. EXPERIMENT 1

The first experiment measured search times when there was no audio cue (visual only), when there were virtual audio cues rendered with individualized HRTFs, and when there were virtual audio cues rendered with generic HRTFs (KEMAR). Also, visual scene complexity was varied by manipulating set sizes, defined as the number of visual stimuli on the screen (including the target). The set sizes tested in Experiment 1 were: 1 (target only), 6, 12, and 24.

2.1. Method

2.1.1. Participants

Nine participants (4 female) with audiometrically-normal hearing and normal or corrected-to-normal vision were paid for their participation. All participants had previous experience with psychoacoustic tasks, including free-field and virtual audio localization experiments. All participants provided informed consent under a protocol approved by the Air Force Research Laboratory, 711th HPW Institutional Review Board.

2.1.2. Stimuli

Visual and auditory stimuli creation and experiment control was done within MATLAB (MathWorks), using the Psychtoolbox [16]. The visual search task was to indicate which one of two possible targets was present. The targets were similar to a Landolt C; they were circular rings with a diameter of 1.24° that had an opening of 0.13° on either the right or left side. The thickness of the circle's line (i.e., the stroke width) was 0.10°. The distractors were circles of the same diameter and stroke width that had no opening. Examples of both possible targets and the distractor are shown in Figure 1. To maximize the sensitivity to differences in the two spatial auditory cues, the target opening was made small such that the target could not be identified with peripheral vision, but had to be foveated [17]. Visual stimuli were presented on a monitor that subtended ±27° azimuth and ±16° elevation. Visual stimuli were presented against a black background and contrast of the visual stimuli was the same for the target and distractors. Pilot studies using a higher contrast value had pronounced perceptual tracers that were distracting to searchers. For each trial, the target was randomly placed and distractors, when present, were randomly placed such that they never overlapped the target or other distractors. When stimuli were immediately adjacent to each other, there was 0.7° between them.

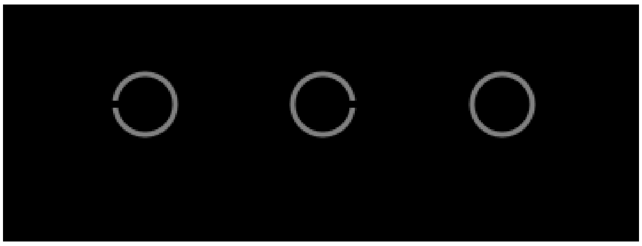

Figure 1: Examples of the visual stimuli that appeared in Experiments 1 and 2, not shown to scale. The leftmost stimulus is a target facing left, the middle stimulus is a target facing right, and the rightmost stimulus is a distractor.

The auditory stimuli were 250-ms bursts of broadband noise (0.2-14.5 kHz) with a pink spectrum. Stimuli had 5-ms cosine ramps and were played at approximately 65 dB. These stimulus parameters were used in prior experiments on aurally aided visual search [9, 10] and were used for comparison. For each individual listener and a KEMAR acoustic mannequin, an HRTF was measured prior to the study according to the methods described in [18]. In short, subjects were outfitted with binaural microphones that blocked off, and sat flush with, the entrance of the ear canal while broadband signals (periodic chirps) were presented from 277 loudspeaker locations surrounding the listener and recorded binaurally. A similar process was used for the KEMAR mannequin, but utilized the built-in ear-canal microphones (GRAS 46AO). The resulting recordings were subsequently used to calculate a sample HRTF for each location in the form of 256 Discrete Fourier Transform magnitude coefficients for each ear and a corresponding ITD. ITDs were found by taking the difference in slope of the best-fit lines to the unwrapped low-frequency (300-1500 Hz) phase response of each ear. Headphone (Sennhiser HD-280) correction filters were also collected for each subject (and KEMAR) using a similar measurement technique (described in [18]). Final spatial filters were created for each measurement location by constructing a time domain filter using the headphone-corrected HRTF magnitude and a minimum phase assumption. ITDs were incorporated into the minimum phase filters by delaying the contralateral ear by the corresponding delay.

2.1.3. Procedure

Participants sat in a double-wall sound-isolated booth and used a chin rest. Their eyes were approximately 54 cm away from the

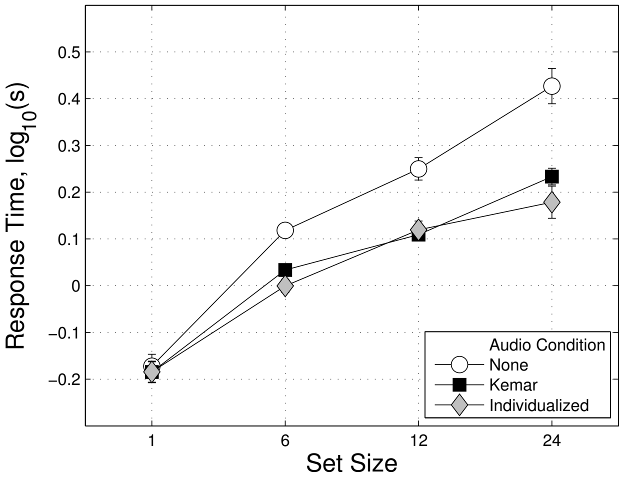

Figure 2: Log response times for Experiment 1, shown as a function of set size and audio cue type. Within-subject standard error bars are shown after Morey [19].

monitor and their eyes were approximately centered on the screen. Each trial began with a fixation point at the center of the screen that was present for 500-1000 ms then disappeared and the search display immediately appeared. Participants searched until they found the target and they used the keyboard arrow keys to indicate their response of left/right. Participants were instructed to find the target as fast as possible while maintaining accuracy. Block length was 50 trials, which varied in duration due to the variation in response times for different conditions. There were 2 blocks of each combination of cue type and set size (i.e., 24 blocks total) and the order of blocks was randomly determined.

2.2. Results

A repeated-measures analysis of variance (ANOVA) was conducted on log response times. Accuracy of target identification responses was quite high (98.9%) so all trials were included in the response time analysis. To assist in plotting data that are averaged across subjects, within-subject standard error bars are calculated after Morey [19] to visually represent the within-subject error term used in statistical tests.

There was a significant main effect of cue type (F(2, 16) = 47.14, p <.001). Both cue types led to shorter response times compared to the visual only condition, and the HRTF cues were not different from each other. Consistent with past research, there was a significant effect of set size (F(3, 24) = 184.70, p <.001), larger set sizes led to longer response times. In addition, there was a significant interaction between cue type and set size (Figure 2, F(6, 48) = 18.77, p <.001). As set size increases, auditory cues provide a larger reduction in search times.

2.3. Discussion

We found that that both the individualized-HRTF cues and the generic-HRTF cues reduced search times in comparison to the visual only condition and that the auditory cues provided a larger reduction as set size increased. The individualized-HRTF and generic-HRTF cues were not different from each other, suggesting that the increased localization accuracy of individualized HRTFs does not affect this task for simple displays with few visual objects. However, it is possible that there would be differences between the two different auditory cues for set sizes larger than those tested here.

A common finding in the literature on virtual audio is that elevation error is larger than azimuth error, particularly with generic HRTFs [20]. This likely is due to the nature of the spectral monaural cues used for elevation, which vary more between individuals than the interaural time and level cues used for azimuth. Therefore, we tried to examine if there was a difference between individualized and generic HRTFs when the azimuth and elevation of the target was considered. We separated the trials where the target appeared at an eccentric azimuth or elevation and compared that to trials where targets appeared at a central azimuth or elevation. The screen subtended ±27° in azimuth and ±16° in elevation. Therefore, eccentric azimuth was defined as targets that appeared at an absolute azimuth greater than 13.5° and eccentric elevation was defined as targets that appeared at an absolute elevation greater than 8°. The left panel of Figure 3 shows eccentric azimuths for the three cue types. Eccentric azimuths were not slower for generic or individualized cues, though they were slower for visual only conditions (p <.01). Eccentric elevations, shown in the right panel of Figure 2, were slower compared to the central elevations (p <.001). In addition, there was an interaction between cue type and eccentricity (p <.05). The KEMAR cue was not as fast as the individualized cue for eccentric elevations, though they are not different for the central elevations

3. EXPERIMENT 2

Experiment 2 investigated a larger range of set sizes, while using the same cue types as Experiment 1. We hypothesized that larger set sizes would introduce an overall difference in response times between the individualized-HRTF cues and the generic-HRTF cues in addition to the differences for eccentric elevations found in Experiment 1.

3.1. Method

The same participants from Experiment 1 were used for Experiment 2. The same stimuli and equipment from Experiment 1 were used for Experiment 2, with the exception that the set sizes tested were: 24, 48, 96, and 1092 (filled screen). When the screen was filled, the stimuli comprised a grid. No stimuli overlapped and there was 0.7° between them.

3.2. Results

The same data analysis was conducted in Experiment 2 as had been conducted for Experiment 1 on the log response times for target identification. Again, target identification accuracy was quite high (98.8%), so all trials were included in the response time analysis. There was a significant main effect of cue type (Figure 4, F(2, 16)= 67.87, p <.001). Performance in the visual-only condition was slower than when KEMAR-HRTF cues were present, and search times with KEMAR-HRTF cues were slower than search times with individualized-HRTF cues. There was a significant main effect of set size (F(3, 24) = 53.90, p <.001). Response times increased as set size increased. There was a significant interaction

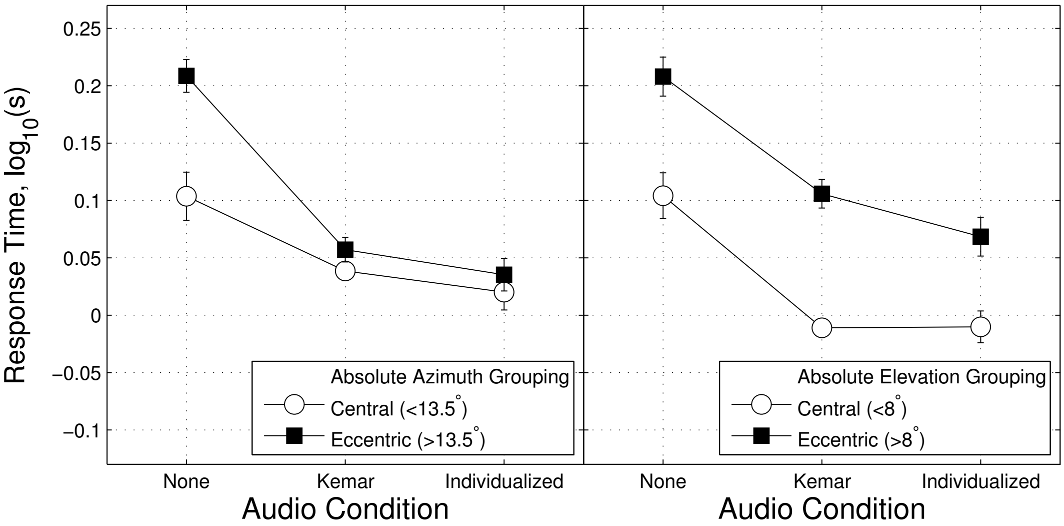

Figure 3: Log response times as a function of eccentricity in azimuth (left panel) and elevation (right panel) for Experiment 1. Within-subject standard error bars are shown after Morey [19].

between cue type and set size (Figure 5, $F(6, 48) = 3.0$, $p < .05$). As set size increased, the reduction in response times provided by individualized-HRTF cues in comparison to visual only search increased ($p < .001$). There was no interaction between KEMAR-HRTF cues and individualized-HRTF cues as a function of set size or between KEMAR-HRTF cues and visual-only search as a function of set size (both $p > .30$).

The trials were separated into eccentric azimuths or elevations and central azimuths or elevations using the same criteria as in Experiment 1 (Figure 6). For azimuth (left panel), target identification times were faster for individualized-HRTF cues than KEMAR-HRTF cues for both central and eccentric target azimuths. There was a significant interaction ($p < .01$), which was due to the slower responses for visual-only eccentric locations compared to the visual-only central locations. For elevation (right panel), there was no interaction between eccentricity and cue type ($p = .07$). Because of the particular hypothesis that eccentric locations might reveal differences between individualized-HRTF cues and KEMAR-HRTF cues, another ANOVA was conducted without the visual-only data. This test was significant ($p < .01$). For central elevations, KEMAR cues and individualized cues were not different but for eccentric elevations, KEMAR cues were slower than individualized cues.

3.3. Discussion

Experiment 2 showed that auditory cues led to faster response times compared to visual only search. In addition, it showed that individualized-HRTF cues were faster than KEMAR-HRTF cues. Further analyses showed that this enhancement was found in eccentric elevations, and surprisingly found also in azimuth (central and eccentric). The difference in azimuth may be attributed to the large main effect of cue type that appears in elevation.

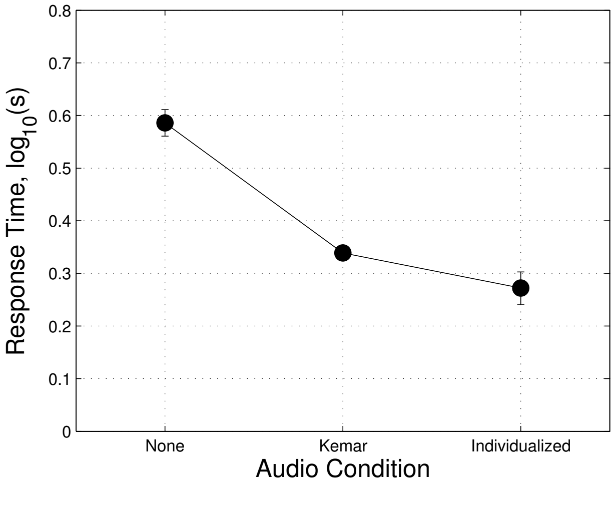

Figure 4: Log response times for Experiment 2, shown as a function of audio cue type. Within-subject standard error bars are shown after Morey [19].

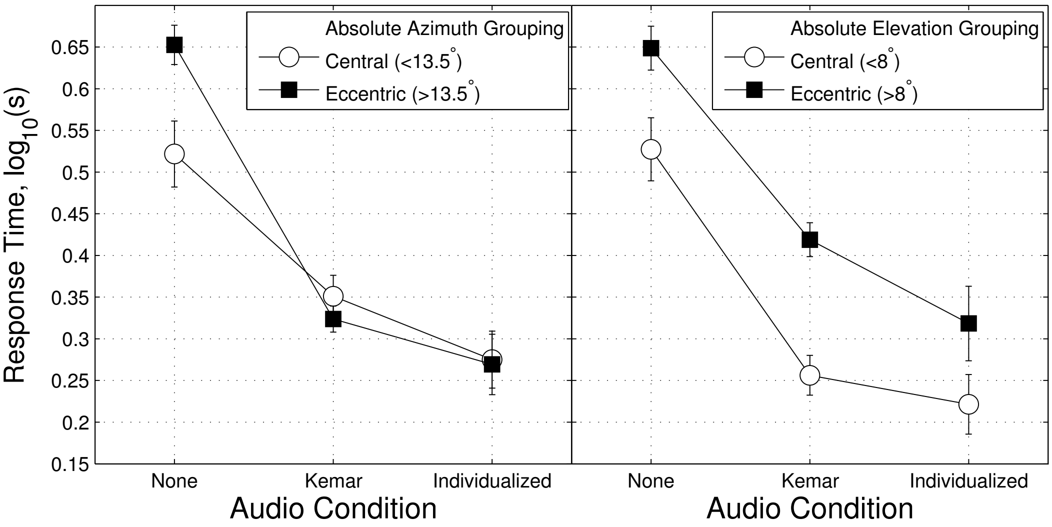

Figure 6: Log response times as a function of eccentricity in azimuth (left panel) and elevation (right panel) for Experiment 2. Within-subject standard error bars are shown after Morey [19].

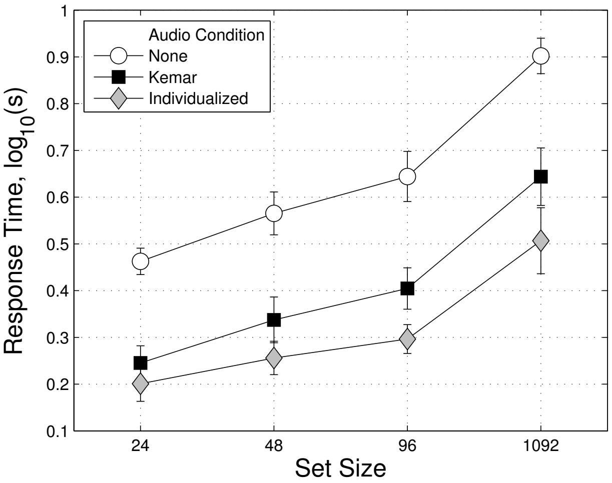

Figure 5: Log response times for Experiment 2, shown as a function of set size and audio cue type. Within-subject standard error bars are shown after Morey [19].

4. GENERAL DISCUSSION

This study explored whether or not spatialized auditory cues could provide effective spatial cues in a visual search task in a small frontal region. Auditory cues were either absent, created using generic HRTFs or created using individually-measured HRTFs. Experiments 1 and 2 tested performance in two different ranges of visual display complexity. Experiment 1 used set sizes from 1 (target only) to 24, and Experiment 2 used set sizes from 24 to 1092 (filled screen).

In general, the presence of a spatial auditory cue reduced search times and interacted with visual scene complexity. In simple visual displays, there was no overall difference between search times with individualized cues and generic cues, though individualized cues to eccentric elevations were faster than generic cues. In Experiment 1, there was no difference between visual only search times and search times with an audio cue when the set size was 1 (no distractors present). This is in contrast to the research done by Perrot et al [14] that found that free-field audio reduced target identification times even on target-only trials. We did not find this, perhaps because virtual audio was used here or perhaps because the identifying features of their visual targets may have been more easily detected using peripheral vision than our stimuli were. Perrot el al's target subtended a visual angle of 0.97°and its orientation was the identifying feature whereas we had a small feature by comparison (0.13°).

For complex displays, there were overall differences between the two types of virtual audio cues, with individualized cues providing faster response times than generic cues. The eccentricity effects were found in complex displays as well, suggesting that individualized cues would become more and more important with the use of larger displays. The findings for both Experiments agree with a study by Vu et al. on free-field audio cues to a visual target that investigated cue displacement [15]. They found that non-displaced cues (cues at the target location) were better than displaced cues (cues off the target location in either horizontal or vertical dimension) which were better than a non-informative cue at reducing search times, and the magnitude of these effects varied with the number of distractors and the amount of displacement. In the present work, the virtual cues created using KEMAR HRTFs may have been displaced or more displaced than the cues created with individualized-HRTFs, and therefore may have been perceived at a spatial location that was not the visual target's location. The displacement may have been mostly in elevation as indicated by search times for eccentric elevations, and Vu et al. found that elevation displacement reduced search times more than horizontal displacement. However, the indication of displacement here is only indirect, no measure of localization was conducted for the virtual stimuli. Future work using virtual audio in visual search should measure localization of the stimuli, and perhaps an *in situ* measurement of localization could be accomplished through eye tracking.

The data from Experiments 1 and 2 suggest that the primary benefit of individualized-HRTF cues is found in the elevation dimension, consistent with previous localization research [20]. In both experiments, individualized-HRTF cues to targets at eccentric elevations resulted in faster searches than generic-HRTF cues. Cues to targets at central elevations resulted in search times that were not different between the two cue types. This finding suggests that other alternative auditory cues that do not indicate elevation, such as stereo panning or interaural level differences alone, would not perform as well as individualized-HRTFs. We did not test these alternative auditory cues but future work could investigate if they may reduce search times compared to visual-only search and if they provide comparable search times to generic-HRTFs or if generic-HRTFs still perform better perhaps due to providing some elevation information.

In conclusion, these data support the notion that spatial audio cues are useful spatial cues to visual displays. Furthermore, individualized spatial audio is functionally superior to generic spatial audio with eccentricity and display complexity.

5. REFERENCES

[1] J. A. Ballas, "Delivery of information through sound," in *Santa Fe Institute Studies in the Sciences of Completex-Proceedings Volume*, vol. 18. Addison-Wesley Publishing, 1994, pp. 79–79.

[2] D. M. Levi, S. A. Klein, and A. Aitsebaomo, "Vernier acuity, crowding and cortical magnification," *Vision Research*, vol. 25, no. 7, pp. 963–977, 1985.

[3] A. W. Mills, "On the minimum audible angle," *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, 1958.

[4] D. W. Grantham, "Detection and discrimination of simulated motion of auditory targets in the horizontal plane," *The Journal of the Acoustical Society of America*, vol. 79, no. 6, pp. 1939–1949, 1986.

[5] G. D. Romigh, D. S. Brungart, and B. D. Simpson, "Free-field localization performance with a head-tracked virtual auditory display," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 9, no. 5, pp. 943–954, 2015.

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

[7] D. R. Begault, "Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 35, no. 4, pp. 707–717, 1993.

[8] B. D. Simpson, D. S. Brungart, R. H. Gilkey, J. L. Cowgill, R. C. Dallman, R. F. Green, K. L. Youngblood, and T. J. Moore, "3d audio cueing for target identification in a simulated flight task," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 16. SAGE Publications, 2004, pp. 1836–1840.

[9] D. R. Perrott, J. Cisneros, R. L. McKinley, and W. R. D'Angelo, "Aurally aided visual search under virtual and free-field listening conditions," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 38, no. 4, pp. 702–715, 1996.

[10] R. S. Bolia, W. R. D'Angelo, and R. L. McKinley, "Aurally aided visual search in three-dimensional space," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 41, no. 4, pp. 664–669, 1999.

[11] J. C. Mateo, B. D. Simpson, R. H. Gilkey, N. Iyer, and D. S. Brungart, "Spatial multisensory cueing to support visual target-acquisition performance," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, no. 1. SAGE Publications, 2012, pp. 1312–1316.

[12] B. D. Simpson, R. S. Bolia, R. L. McKinley, and D. S. Brungart, "The impact of hearing protection on sound localization and orienting behavior," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 47, no. 1, pp. 188–198, 2005.

[13] J. M. Haggit, "Cued visual search and multisensory enhancement," Master's thesis, Wright State University, 2014.

[14] D. R. Perrott, T. Sadralodabai, K. Saberi, and T. Z. Strybel, "Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 33, no. 4, pp. 389–400, 1991.

[15] K.-P. L. Vu, T. Z. Strybel, and R. W. Proctor, "Effects of displacement magnitude and direction of auditory cues on auditory spatial facilitation of visual search," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 48, no. 3, pp. 587–599, 2006.

[16] D. H. Brainard, "The psychophysics toolbox," *Spatial Vision*, vol. 10, pp. 433–436, 1997.

[17] J. P. McIntire, P. R. Havig, S. N. Watamaniuk, and R. H. Gilkey, "Visual search performance with 3-d auditory cues: Effects of motion, target location, and practice," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 2010.

[18] D. S. Brungart, G. Romigh, and B. D. Simpson, "Rapid collection of hrtfs and comparison to free-field listening," in *International Workshop on the Principles and Applications of Spatial Hearing*, 2009.

[19] R. D. Morey, “Confidence intervals from normalized data: A correction to Cousineau (2005),” *Tutorials in Quantitative Methods for Psychology*, vol. 4, no. 2, pp. 61–64, 2008.

[20] G. D. Romigh and B. D. Simpson, “Do you hear where i hear?: isolating the individualized sound localization cues,” *Frontiers in Neuroscience*, vol. 8, 2014.