Cleared for public release by ASC-99-0336, 19 February 1999

# Aurally Aided Visual Search in Three-Dimensional Space

**Robert S. Bolia**, Wright-Patterson Air Force Base, Ohio, **William R. D'Angelo**, University of Connecticut Health Center, Farmington, Connecticut, and **Richard L. McKinley**, Wright-Patterson Air Force Base, Ohio

We conducted an experiment to evaluate the effectiveness of spatial audio displays on target acquisition performance. Participants performed a visual search task with and without the aid of a spatial audio display. Potential target locations ranged between plus and minus 180° in azimuth and from −70° to +90° in elevation. Independent variables included the number of visual distractors present (1, 5, 10, 25, 50) and the spatial audio condition (no spatial audio, free-field spatial audio, virtual spatial audio). Results indicated that both free-field and virtual audio cues engendered a significant decrease in search times. Potential applications of this research include the design of spatial audio displays for aircraft cockpits and ground combat vehicles.

## INTRODUCTION

Many researchers have investigated the utility of three-dimensional (3D) auditory displays for reducing the workload associated with an overloaded visual channel and for enhancing the detection and identification of visual targets (Flanagan, McAnally, Martin, Meehan, & Oldfield, 1998; Nelson et al., 1998; Perrott, Cisneros, McKinley, & D'Angelo, 1996). Advances in digital signal-processing technology and the development of electromagnetic position trackers have enabled the construction of small, relatively inexpensive spatial audio displays that can be used to aid in visual target acquisition in aircraft cockpits, ground combat vehicles, and training simulators. Recent investigations have demonstrated the utility of spatialized audio cueing for some aircraft applications (Bronkhorst, Veltman, & van Breda, 1996; McKinley & Ericson, 1997). One approach to quantifying the performance advantage afforded by the use of spatial audio displays is to investigate the reduction in search times obtained when such displays are used in a target acquisition task. This can be accomplished using the aurally aided visual search paradigm developed by Perrott, Saberi, Brown, and Strybel (1990).

In a classical visual search task, a participant looks for a target stimulus among an array of nontarget stimuli, or *distractors*. In the usual paradigm, the participant reports either the presence or absence of the target, and his or her reaction time (RT) is recorded. The relationship between RT and the number of distractors is generally linear, with RT either increasing with set size or remaining approximately constant independent of the number of nontarget elements in the field. The former case is generally interpreted as involving some limited-capacity process, in which the observer must direct his or her attention to small sections of the visual field sequentially, and is hence referred to as a *serial* search. Conversely, searches for which reaction times do not vary with the number of distractors are presumed to involve some "preattentive" or parallel process and are hence termed *parallel* searches (Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989).

Perrott and his colleagues have investigated the effect of the addition of an audio cue colocated with the visual target on visual search times (Perrott, et al., 1990, 1996; Perrott, Sadralodabai, Saberi, & Strybel, 1991). In one experiment (Perrott et al., 1991) participants

Address correspondence to Robert S. Bolia, AFRL/HECP, 2255 H. Street, Bldg. 33, Wright-Patterson AFB, Ohio 45433; robert.bolia@wpafb.af.mil. **HUMAN FACTORS**, Vol. 41, No. 4, December 1999, pp. 664–669. Copyright © 1999, Human Factors and Ergonomics Society. All rights reserved.

searched for a visual target among an array of multiple distractors presented in the central visual field with and without spatial audio cueing. The results of this study indicated that the addition of a spatial audio cue provides a significant reduction in RT even for targets within a few degrees of the fovea.

More recently, Perrott et al. (1996) investigated the trivial case (i.e., one target, zero distractors) of an aurally aided visual search task in which the visual field was less restricted. In this study the target was presented on the interior surface of a geodesic sphere to an observer seated at the center. It was observed that the addition of a spatially correlated audio cue provided a significant reduction in target acquisition times – several hundred ms for targets in the rear hemifield.

Perrott et al. (1996) also investigated the utility of using a *virtual* audio cue to guide visual search. Spatialization of the signals was achieved by (a) convolving the audio cue with a transfer function representing the transformation of the sound source by the head, torso, and pinna – the so-called head-related transfer function (HRTF); and (b) delaying the signal in one ear relative to the other to simulate differences in arrival time for sound sources that were not located in the median sagittal plane of the listener. Results of the study indicated that the addition of a virtual spatial audio cue significantly reduced search times, although the magnitude of the reduction was not as great as was that achieved in the free-field case.

The purpose of the present study was to investigate the RT versus set size dependence for searches in a complete sphere with and without the presence of a spatially correlated audio cue. We discuss the results in terms of their implications for the design of virtual spatial audio displays.

## METHOD

### Apparatus

All testing was conducted in the Air Force Research Laboratory's Auditory Localization Facility (ALF) at Wright-Patterson Air Force Base, Ohio. During each testing session the participant was seated at the center of a geodesic sphere with a radius of 2.3 m and enclosed within a cubic anechoic chamber with a side length of 6.7 m. The aluminum struts of the sphere were covered with 2.5 cm acoustic foam to minimize reflections. A Bose 4.5 in. (11.43 cm) Helical Voice Coil full-range loudspeaker (Model 118038) was located at each of the sphere's 272 vertices. They were spaced approximately 15° apart and were positioned orthogonally to the sphere. A four-element square array of light-emitting diodes (LEDs) was mounted 5 cm above the anterior surface of each loudspeaker. Each diode emitted a 620-nm wavelength light at a luminance of 200 mL and subtended a visual angle of 0.5°. Target and distractor stimuli were approximately three log units above photopic threshold (Perrott et al., 1996).

### Experimental Design

Three auditory conditions (nonaudio, free-field audio, and virtual audio) were combined factorially with five set-size conditions (1, 5, 10, 25, or 50 visual distractors). In the nonaudio condition, the participant performed a visual search task with no audio cue. In the free-field condition, the participant performed the same task with a free-field audio cue colocated with the visual target. The third condition, which employed virtual audio cues, was the same as the free-field condition except that the audio cue was presented over circumaural headphones (Sennheiser HD-560, Germany). The virtual audio cueing was provided by an Air Force Research Laboratory 3D Audio Display Generator (3D ADG) built by Veridian (Dayton, Ohio) coupled to a Polhemus (Colchester, VT) 3Space electromagnetic head-tracker using HRTFs measured from a Knowles Electronic Manikin for Acoustic Research (KEMAR) (Itasca, IL) with 90th-percentile pinnae. The orientation information from the head tracker was employed by the 3D ADG to ensure that the audio image presented to the listener remained fixed with respect to the virtual acoustic environment (and, by extension, to the sphere containing the visual targets), rather than with respect to the listener's head.

### Participants

Three men and two women between the ages of 21 and 35 participated in the experiment.

Three of the participants were recruited from the pool maintained at the Air Force Research Laboratory, one was a second lieutenant in the U.S. Air Force, and the other was one of the principal investigators (Bolia). All participants had pure tone thresholds of less than 15 dB above audiometric zero (Goodman, 1965) and uncorrected 20/20 vision in both eyes.

### Procedures

At the beginning of each session, the observer was seated at the center of the ALF with the room darkened and all of the LEDs turned off. If the virtual audio condition was being tested, he or she would don headphones. At the inception of each trial, either two or four LEDs were energized at the fixation point (0° azimuth, 0° elevation). Before the observer was permitted to continue, he or she was required to correctly indicate the number of LEDs energized via a two-button response switch. This guaranteed that the observer was always facing forward at the beginning of each trial. The target and distractor LED clusters were then energized simultaneously, and the observer began the search. All targets fell between plus and minus 180° in azimuth and between –70° and +90° in elevation. The loudspeakers with an elevation coordinate of less than –70° were not used because either they were not equipped with an array of LEDs or the LEDs were not visible to the seated observer. As a result, the number of potential target/distractor locations was decreased from 272 to 266.

On a given trial, the target cluster always contained either two or four energized LEDs, and the search task involved finding the target and indicating the number of LEDs that were energized using a two-alternative, forced-choice response. The clusters at the distractor locations contained either one or three energized LEDs (see Figure 1). In the free-field audio condition, a continuous acoustic stimulus (pink noise, 70 dB SPL) emanated from the same location as the target. In the virtual audio condition, the same stimulus was presented over headphones and was digitally filtered so that it appeared to emanate from the direction of the target. In both conditions, the onset of the audio cue was simultaneous with that of the visual target. RT and correctness of response were stored for each trial.

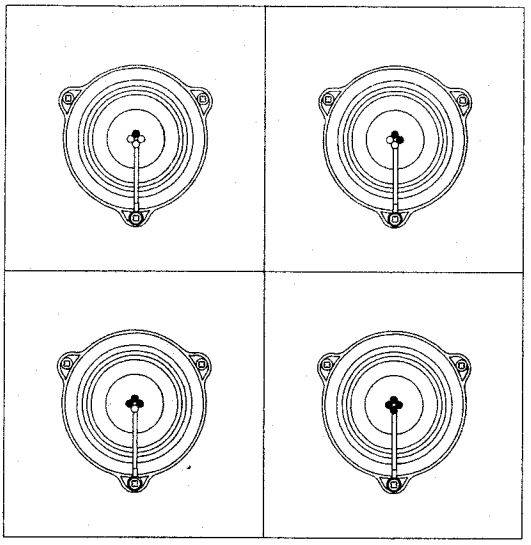

*Figure 1.* Schematic diagram of the loudspeaker + LED array configuration for all possible targets and distractors. Filled circles indicate energized LEDs. Target configurations are depicted on the right side of the figure and distractor configurations are depicted on the left.

Participants were exposed to each of the conditions several times prior to testing. Subsequently, each of the participants completed 5 blocks of 266 trials (one trial per target location) for each of the 15 possible treatments (3 Audio Conditions × 5 Set Sizes). The order in which the conditions were run was randomized session by session to minimize order effects.

## RESULTS

### Percentage Correct

In order to examine the possibility that variations in search time occasioned concomitant variations in target identification accuracy, mean percentages of correct responses were analyzed using a 3 (audio condition) × 5 (set size) repeated-measures analysis of variance (ANOVA). The main effect of audio condition, $F(2, 8) = 3.15$, $p > .05$, the main effect of set size, $F(4, 16) = 1.55$, $p > .05$, and the interaction between them, $F(8, 32) = 1.16$, $p > .05$, were not found to be statistically significant. The percentage of correct responses was

greater than 95% for all combinations of audio condition and set size.

### Reaction Time

Mean reaction times for all of the experimental conditions were analyzed using a similar 3 × 5 repeated measures ANOVA. This revealed significant main effects of audio condition, $F(2, 8) = 99.65$, $p < .05$, and set size, $F(4, 16) = 242.24$, $p < .05$, and a significant Audio Condition × Set Size interaction, $F(8, 32) = 83.18$, $p < .05$. This interaction is illustrated in Figure 2, in which reaction time is plotted as a function of set size for each of the three audio conditions.

The interaction was further investigated by tests of simple main effects of the audio conditions as a function of set size and of the set sizes as a function of audio condition. This was done in order to examine (a) the effect of set size on search times for each of the tasks (visual search vs. aurally aided visual search) and (b) whether or not there were performance differences between the audio conditions for all set sizes or only for visual fields of sufficient complexity. All simple main effects were statistically significant ($p < .01$). Post hoc $t$ tests corrected for family-wise $\alpha$ error were performed for a pairwise comparison of the mean reaction times for all of the set sizes within each audio condition and for all of the audio conditions within each set size. Furthermore, simple linear regression analyses were performed on RT as a function of set size for each audio condition. None of the RT versus set size functions were found to differ statistically from linearity ($p < .05$).

Within the nonaudio condition, all effects of set size were significant ($p < .01$). As Figure 2 illustrates, reaction times in this condition increased linearly with set size. The regression analysis revealed a rate of increase of 248 ms per distractor ($t = 105.11$; $p < .05$).

Under the free-field audio condition, none of the effects reached significance at the .01 level. This implies that when a spatialized free-field audio cue is added to a visual search task, reaction times are approximately constant regardless of set size. This result was supported by the regression analysis, which showed

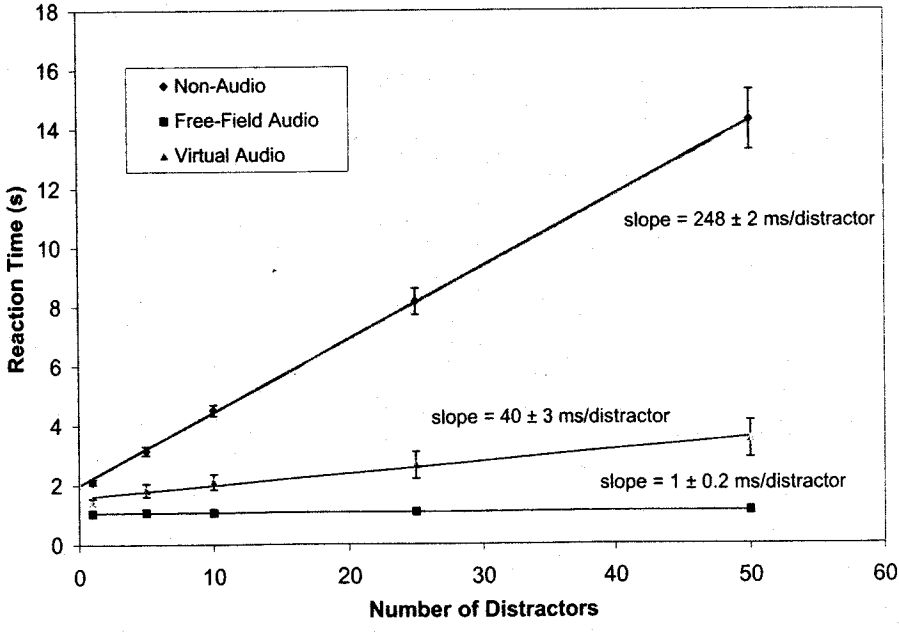


*Figure 2.* Reaction time (RT) as a function of set size for each of the three audio conditions. The markers represent actual data, ± standard errors (the error bars for the free-field audio data are not visible at this scale). The lines are the best-fit lines determined by simple linear regression on the mean RT data. The slopes of the regression lines (± standard errors) are also given and represent the increase in RT for each additional distractor added to the set.

that RT increased with set size at a rate of only 1 ms per distractor ($t = 5.48$; $p < .05$).

Only two of the comparisons within the virtual audio condition were significant at the .01 level: Reaction times in the 10-distractors condition were significantly different from those in both the 1-distractor and the 5-distractors conditions. The results of the regression analysis exhibited an RT versus set size function with a slope of 40 ms per distractor ($t = 11.38$; $p < .05$).

Within each set size, all differences in reaction time between audio conditions were statistically significant ($p < .01$), with the exception of the differences in the free-field versus virtual comparisons for the 25 and 50 distractors conditions, which were marginal ($p = .019$ and $p = .015$, respectively).

## DISCUSSION

Results indicated that the addition of a spatial audio cue significantly decreased reaction times in a visual search task without a corresponding decrease in the percentage of correct responses. In the free-field audio condition, these results point to a parallel search(i.e., a search for which time to complete the search does not vary significantly with the number of distractors). In the virtual audio condition, a serial search was implicated, but with a slope one sixth of the size of the slope obtained in the nonaudio condition. This represents a substantial improvement in performance over that obtained in the unaided condition.

The difference in the rate of increase of RT with set size between the free-field and virtual audio conditions was presumably attributable to imperfect replication of the free-field cues under headphone listening conditions. This might have been caused by one or more factors, including (a) defects inherent in the HRTF collection technique (i.e., there might have been physical differences between the manikin and the individual, the sum of which have perceptual consequences; e.g., acoustic absorption properties of the head and torso); (b) the use of nonindividualized HRTFs; and (c) the spatial resolution of the HRTFs.

The first point could be addressed by a psychophysical validation experiment comparing performance with HRTFs collected from a human listener with those collected from a manikin with that listener's individualized pinna models. This hypothesis is currently under investigation.

It is also possible that the discrepancy between performance in the free-field condition and performance in the virtual condition was attributable to the use of nonindividualized HRTFs in the generation of the virtual cues. Wenzel, Arruda, Kistler, and Wightman (1993) demonstrated that although listeners typically localize as well in azimuth with nonindividualized transfer functions, they make fewer front-back confusions and are more accurate in their elevation judgments when using their own HRTFs. Accordingly, significant degradations along either of these dimensions could result in longer search times for aurally aided visual searches conducted in an environment that includes targets in the rear hemifield or outside the horizontal plane.

Finally, the difference in performance between the free-field and virtual audio conditions might have been caused by the spatial resolution of the HRTF set. The 266 loudspeakers used in this experiment were fixed in space such that the separation between them averaged about 15° in both azimuth and elevation. When a continuous broadband sound was played through one of the loudspeakers, a listener turning his or her head received continuous updates as to the spatial location of the source. Conversely, a listener turning his or her head while listening to the 3D ADG received only discrete updates – the HRTFs shifted suddenly halfway between two loudspeaker locations. A small amount of spatial uncertainty was thereby created in the virtual audio condition, which could account for longer reaction times than those achieved in the free-field condition. This hypothesis could be tested by collecting a denser set of transfer functions and examining performance using sets of transfer functions of different densities. The results of such a study would certainly have implications for the design of spatial audio displays.

## CONCLUSION

The results of this experiment extend the work of Perrott et al. (Perrott et al., 1990, 1991,

1996) in demonstrating the utility of the auditory system for the redirection of gaze. Indeed, the results suggest that visual searches of arbitrary complexity in a nearly complete sphere can be completed as quickly as can much simpler searches given a spatial audio cue colocated with the target. This offers empirical support for the hypothesis that one of the primary functions of the auditory modality is to assist the visual system in the detection and identification of targets in extracorporeal space (Heffner & Heffner, 1992; Perrott et al., 1990). Thus it makes sense to consider the utility of spatial audio cueing when designing displays for use in environments in which efficient acquisition of visual targets is important (aviation, combat vehicles, etc.).

The results of the present investigation demonstrate that virtual spatial audio cueing can reduce search time by a factor of six or more for high-complexity searches without a corresponding reduction in the accuracy of target identification. An improvement of this magnitude could be the difference between life and death for a pilot flying a high-performance aircraft or for a member of a tank crew on a battlefield.

## ACKNOWLEDGMENTS

The authors acknowledge the technical contributions of Ronald C. Dallman and Michael L. Ward of Veridian; Dennis L. Allen of Sytronics, Inc., who provided technical support as well as illustrations; W. Todd Nelson and Mark A. Ericson of the Air Force Research Laboratory; Charles W. Nixon of Veridian; and two anonymous reviewers, who provided valuable comments on earlier drafts.

## REFERENCES

Bronkhorst, A. W., Veltman, J. A., & van Breda, L. (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors, 38*, 23–33.

Flanagan, P., McAnally, K. I., Martin, R. L., Meehan, J. W., & Oldfield, S. R. (1998). Aurally and visually guided visual search in a virtual environment. *Human Factors, 40*, 461–468.

Goodman, A. (1965). Reference zero levels for pure-tone audiometers. *American Speech and Hearing Association, 7*, 262–263.

Heffner, R. S., & Heffner, H. E. (1992). Visual factors in sound localization in mammals. *Journal of Comparative Neurology, 317*, 219–232.

McKinley, R. L., & Ericson, M. A. (1997). Flight demonstration of a 3-D auditory display. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments.* Mahwah, NJ: Erlbaum (pp. 683–699).

Nelson, W. T., Hettinger, L. J., Cunningham, J. A., Brickman, B. J., Haas, M. W., & McKinley, R. L. (1998). Effects of localized auditory information on visual target detection performance using a helmet-mounted display. *Human Factors, 40*, 452–460.

Perrott, D. R., Cisneros, J., McKinley, R. L., & D'Angelo, W. R. (1996). Aurally aided visual search under virtual and free-field listening conditions. *Human Factors, 38*, 702–715.

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. (1990). Auditory psychomotor coordination and visual search behavior. *Perception and Psychophysics, 48*, 214–226.

Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. (1991). Aurally aided search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors, 33*, 389–400.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America, 94*, 111–123.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 419–433.

Robert S. Bolia is a computer scientist in the Human Interface Technology Branch, Human Effectiveness Directorate, Air Force Research Laboratory. He received his B.A. in mathematics from Wright State University in 1997.

William R. D'Angelo is a research assistant in the Department of Anatomy at the University of Connecticut Health Center, Farmington, CT. He received his M.S. in biomedical engineering from Boston University in 1992.

Richard L. McKinley is the technical director of the Aural Displays and Bioacoustics Branch, Human Effectiveness Directorate, Air Force Research Laboratory. He received his M.S. in bioengineering/digital signal processing in 1988 from the U.S. Air Force Institute of Technology, Dayton, OH.

*Date received: April 1, 1999*
*Date accepted: June 21, 1999*