Smith, Slavendy, Harris & Koubek (eds) ***Usability Evaluation and Interface Design: Cogntive Engineering, Intelligent Agents and Virtual Reality*** (Vol 1). 2001 Larwrence Erlbaum, (1363-1367).

The Effectiveness of Visual vs. Auditory Cues in Visual Search Performance: Implications for the Design of Virtual Environments.

Thomas. Z. Strybel and Diane L. Guettler

Department of Psychology, California State University Long Beach, Long Beach, CA 90840

ABSTRACT

In virtual environments, cuing the user to the location of important events may be necessary, and researchers are evaluating the effectiveness of auditory and visual cues in this regard. Although stimulus-driven visual cues produce fastest search times, the inability to ignore such cues could be disastrous in some work environments. Therefore, we evaluated the effectiveness of auditory spatial cues in orienting attention when stimulus-driven visual cues were also present. When the visual cue validity was 90%, invalid auditory spatial cues reduced search times. When the auditory cue validity was 90%, invalid visual cues increased search times. When neither cue was reliable (validity=50%), participants usually searched the visual cue area before the auditory cue.

1. INTRODUCTION.

Virtual environments can enhance the quality of human-computer interaction, particularly when they are applied to complex environments. A virtual environment is a computer-based generation of a natural or abstract environment; the user is immersed in a real or artificial world, and interacts directly with components of this world (Bullinger, Bauer, & Braun, 1997). For immersion, the virtual environment should have an extended visual field of view (often a full 360°), with the information displayed a function of the current position and orientation of the user. One particularly promising application of virtual environments is the aerospace cockpit. A virtual environment could provide configural displays about the status of the aircraft and the immediate environment outside the aircraft. Head Mounted Displays (HMDs), currently used in rotocraft cockpits, are examples of primitive virtual environments. With HMDs, flight and sensor data are projected on a lens attached to the pilot's helmet. Some information displayed depends on the pilot's current head position.

Performance costs can offset the potential benefits of virtual environments, however. If the information available to the user depends on current eye and head position, the user will miss critical information when he or she is not oriented properly. In cockpit applications, failure to detect and respond to critical information could be disastrous. Designers could reduce these consequences with multisensory displays that present information via the auditory and tactual, as well as visual, modalities. Of course, proper utilization of these capabilities places added burdens on the designers of the displays because the choices for display location, modality and format are expanded. For usable virtual environments, designers should incorporate the advantages and disadvantages of different sense modalities. For example, although auditory spatial resolution is poorer than visual spatial resolution, the auditory system has a greater field of view. This omnidirectional property of audition means that objects can be heard anywhere in the environment without the need for repositioning the auditory apparatus. Designers must also consider how information from each modality affects the user's selection of relevant sources from the environment. Research in this area, known as selective attention, has been focused primarily on the effectiveness of visual cues in covert orienting (the eyes and head remain stationary). The effectiveness of visual and multisensory cues in overt orienting (information selected by repositioning the eyes and head) has received less attention although overt orienting is multimodal in nature. For example, the time required to position the eyes on a visual target is reduced when a spatially-coincident auditory signal accompanies the target (e.g., Frens, Van Opstal, & Van Der Willigen, 1995). These multisensory effects are usually attributed to the role of the superior colliculus in regulating orienting behaviors. In the deep layers of this structure, multimodal maps of space have been observed and these direct the orienting behaviors of the organism. The multimodal maps are based on a single frame of reference, meaning that spatial maps specific to individual sensory systems have been transformed into a common reference frame in the superior colliculus (e.g., Stein & Meredith, 1993).

These features of the orienting system suggest that cueing critical information in virtual environments could be enhanced with auditory spatial cues because head and eye positioning are faster with multimodal targets, and the

Smith, Slavendy, Harris & Koubek (eds) ***Usability Evaluation and Interface Design: Cogntive Engineering, Intelligent Agents and Virtual Reality*** (Vol 1). 2001 Larwrence Erlbaum, (1363-1367).

auditory system is omnidirectional. Many behavioral studies are consistent with this observation. Auditory spatial cues substantially reduce the time required to locate and identify a visual target in a large search field. The benefits of auditory spatial cues depend on the characteristics of both the visual search task and the auditory cue. The important visual factors are target distance, target contrast, and the number of nontarget "distractors" in the local (area immediately surrounding the target) and global (remaining area) search field. The important auditory factors are cue precision and amplitude (e.g., Rudmann & Strybel, 1999; Strybel et al., 1995; Strybel, Vu, & Castagna-Osorio, 2000). Although providing auditory spatial cues to the target's location can improve search performance, complex environments might have multiple targets and cues, and the user must respond first to the most critical events. It is important, therefore, that designers know if the user can control his/her search strategy in the presence of multiple cues and targets. In selective attention research, a distinction is made between stimulus-driven and goal-directed cues (e.g., Egeth & Yantis, 1997). Stimulus-driven cues, such as abrupt visual onsets or feature singletons, capture attention; the user has difficulty ignoring them. The intentions of the observer control goal-directed cues. Although stimulus-driven cues produced the faster search times, the inability to ignore the cue could be disastrous in complex environments. Whether auditory spatial cues are stimulus-driven or goal-directed is debatable. With covert orienting, Spence and Driver (1997) showed that auditory cues to both auditory and visual targets exhibited stimulus-driven properties. However, visual cues appeared stimulus-driven only to visual targets. With overt orienting, Rudmann and Strybel (1999) showed that search times to displaced audio cues were sometimes longer than when no cue was present. Strybel Vu and Castagna-Osorio (2000) obtained similar results when the cue was presented in background noise. Fujawa and Strybel (1997) showed that uninformative, high-intensity auditory cues interfered with visual search performance. Although these findings suggest that auditory spatial cues are stimulus-driven, there has not been a direct comparison of visual and auditory spatial cues for overt orienting. In the present experiment we evaluated the stimulus-driven properties of auditory spatial cues when visual stimulus-driven cues were also present. We also varied the validity of the cues, or the percentage of trials on which the cue accurately signaled the target. For high validity conditions (90%), we investigated the extent to which an invalid cue could be ignored. For uninformative validity conditions (50%), we determined which cue the user would prefer.

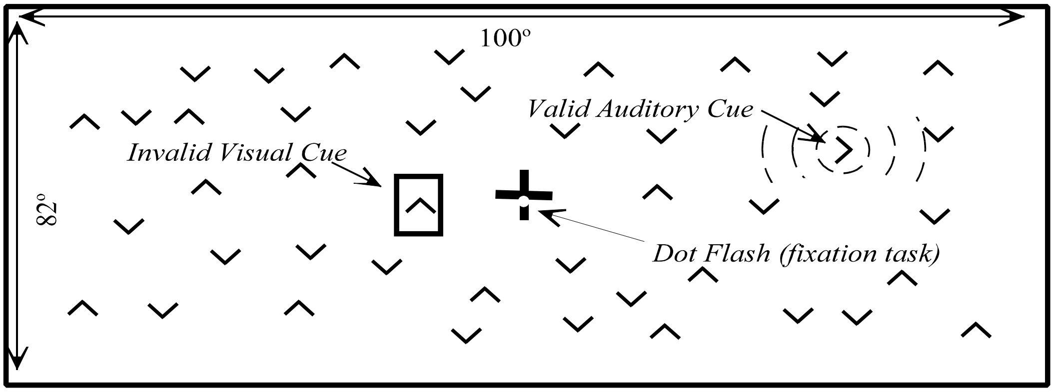

Figure 1. Search field and stimuli for a valid auditory cue and invalid visual cue that is closer to the fixation cross than the target (not to scale).

2. METHOD.

2.1 Participants. Six participants with normal hearing and normal or corrected-to-normal vision were tested. Two had participated in previous experiments on auditory spatial processing. Participants were paid $10.00/hour for their participation.

2.2 Apparatus. The experiment was conducted in large (approximately 3 m x 4.3 m) dimly-lit, semianechoic room. All surfaces of the room were covered by Martek 10.16-cm acoustic foam, except for a small window separating the test room from the control room. The participant was seated 1.3m in the front and center of a large (100° horizontal by 82° vertical) screen in the middle of the room. The search field was projected through the window onto the screen by two projection panels and overhead projectors in the control room. Visual stimuli (targets, distractors, visual cues,

Smith, Slavendy, Harris & Koubek (eds) ***Usability Evaluation and Interface Design: Cogntive Engineering, Intelligent Agents and Virtual Reality*** (Vol 1). 2001 Larwrence Erlbaum, (1363-1367).

and fixation cross) were white symbols projected on a dark background with a contrast of 73%. Auditory cues were provided by 45 Blaupunkt 7.6-cm loudspeakers positioned behind the screen in eight concentric circular rings with radii between 12° and 45°. The number of speakers in each ring varied between two and eight. The acoustic cue was high-pass noise (lower frequency cutoff = 2000 Hz.) at 70 dB A-weighted, pulsed at 10 Hz. A microprocessor with Tucker-Davis Technologies' programmable modules controlled all visual and auditory stimulus generation, trial sequencing, and response collection.

2.3 Procedure. The stimulus arrangement is shown in Figure 1. Participants were instructed to locate and identify a visual target (left or right pointing arrow) in a distractor field (up/down facing arrows) by pushing one of two buttons attached to a hand-held box. Targets were presented at one of three distances from fixation, 18°, 30° or 42°. The density of the distractors was either 16% or 32%. Each participant was tested in five cue/validity conditions. In two of these conditions, only one cue was available on each trial (validity = 100%). These conditions were considered baseline conditions because they provided a measure of the effectiveness of the auditory and visual cue in signaling target location when presented alone. In the remaining three conditions, two cues were provided on each trial, one auditory and one visual. These conditions varied the percentage of trials on which the cues were valid (90% auditory-10% visual, 90% visual-10% auditory, and 50% visual-50% auditory). For the 90% validity conditions, the participants were instructed to ignore the invalid cue because it rarely signaled the target. For the 50% validity condition, the participants were told that neither cue was reliable and there was no best strategy. A valid auditory cue was the noise presented from a loudspeaker directly behind the target. A valid visual cue was a rectangle surrounding the target. An invalid cue was either the sound or rectangle presented at a distractor in the opposite hemifield. The distance of the invalid cue, relative to the target was also varied (closer, equidistant or farther from fixation). For example, Figure 1 presents a valid auditory cue in the right hemifield and invalid visual cue in the left hemifield. The invalid visual cue is closer to the fixation cross relative to the target.

The specific sequence of events on each trial was as follows. At the beginning of the trial, a fixation cross and pattern of "X"s corresponding to the location of targets and distractors was projected. These remained on the screen for a variable foreperiod (1000 - 2000 ms). To ensure that the participant remained fixated on the cross until the trial began, a fixation task was presented in the last 100 ms of the foreperiod. A filled circle was briefly flashed either above or below the center of the cross, as shown in Figure 1. At the end of the foreperiod the cross was removed and the targets, distractors and cues were presented. The distractors were displayed by removing the upper/lower half of each distractor "X." The target was displayed by removing the left/right half of the target "X." The visual and auditory cues were also turned on. Because the visual target and distractors were already on the screen, the visual cue was seen as an abrupt onset. The participant located and identified the target direction by pushing the corresponding response button. After the target response, the participant indicated whether the fixation flash occurred on the top or bottom half of the cross. The participant was told to respond to the target as quickly as possible, but accuracy was more important than speed. The fixation response was not timed.

Each participant ran three sessions per cue-validity condition. Within a session, ten trials at each of three target distances, two distractor densities and three invalid cue relative locations, were presented in random order, making a total of 180 trials per session. The number of valid and invalid trials was determined by the validity condition. Three sessions were run at one condition, before going to the next, with the order of conditions randomized. The first trial block in each condition was considered practice, and was not included in subsequent analysis.

Smith, Slavendy, Harris & Koubek (eds) ***Usability Evaluation and Interface Design: Cogntive Engineering, Intelligent Agents and Virtual Reality*** (Vol 1). 2001 Larwrence Erlbaum, (1363-1367).

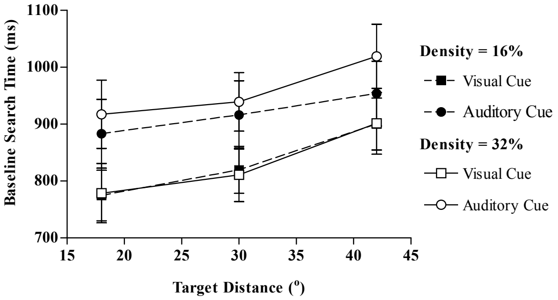

Figure 2: **Baseline Condition.** Mean search times as a function of target distance and distractor density for auditory and visual cues presented alone.

3. RESULTS

Mean search times were computed for each subject at each condition after incorrect-response trials and trials producing reaction times more than two standard deviations from the mean were omitted. We analyzed the validity conditions separately because in each validity condition, the number of valid trials varied. For the baseline conditions (validity = 100%), a three-way, Cue Modality x Distractor Density x Target Distance repeated measures analysis of variance was performed on the mean search times for each subject. Significant Cue x Distance and Cue x Distractor Density interactions were obtained [F(1,5) = 19.78; p = 0.007; F(2,10) = 4.35; p = .04, respectively]. These interactions are illustrated in Figure 2. The visual cue produced faster overall search times and was unaffected by distractor density. Distractor density affected search times with the auditory cue, with 32% density search times being on the average 40 ms longer than 16% density. Search times with both visual and auditory cues were affected by distance, although the effect of distance was more pronounced in the visual cue condition. In Figure 2 it also appears that the effect of distance in the auditory modality was lessened at 16% distractor density. The three-way interaction of cue modality, distractor density and distance was only marginally significant, however.

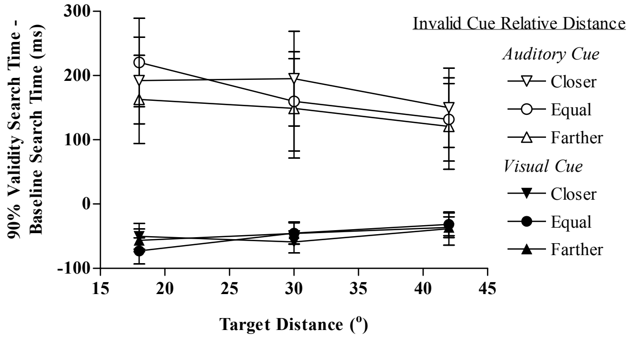

Figure 3A: Validity = 90%. Change in search times relative to baseline as a function of target distance and invalid cue relative distance.

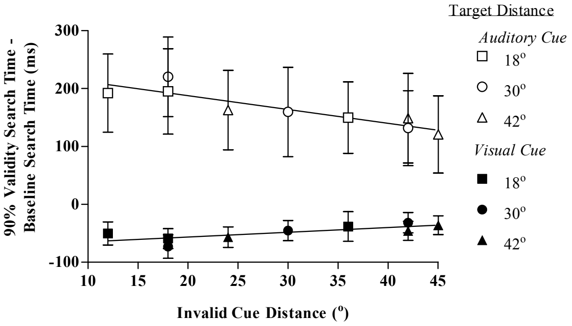

Figure 3B: Validity=90%. Change in search times relative to baseline as a function of invalid cue and target distance from fixation.

At 90% validity, we subtracted the baseline search times from the valid cue search times for each subject, because of the relative advantage of the visual cue, and the difference in the number of valid and invalid trials. Significant Cue x Target Distance [F(2, 10) = 4.33; p = .04] and Cue x Target Distance x Invalid Cue Relative Distance [F(4,20) = 4.20; p = .01] interactions were obtained, as shown in Figure 3A. Most surprising was the effect of invalid auditory cues on search times with the valid visual cues: search times were shorter than with the visual cue presented alone. This benefit of invalid auditory cues was reduced with target distance. Search times with valid auditory cues, on the other hand, increased by 100 - 300 ms over the auditory baseline condition. The interference was highest at the shortest target distance, and decreased with distance. A consistent effect of the invalid cue relative distance was not evident in Figure 3A, so we replotted these data in Figure 3B to account for the actual distance of the invalid cue from fixation. From this figure it appears that the benefits of the invalid auditory cue to searching with the valid visual cue, and the costs of the invalid visual cue to searching with the valid auditory cue are linearly related to the distance of the invalid cue from fixation. The absolute value of the slope of the valid auditory function (2.4 ms/^{o}) is more than double the slope for the visual function (.82 ms/^{o}).

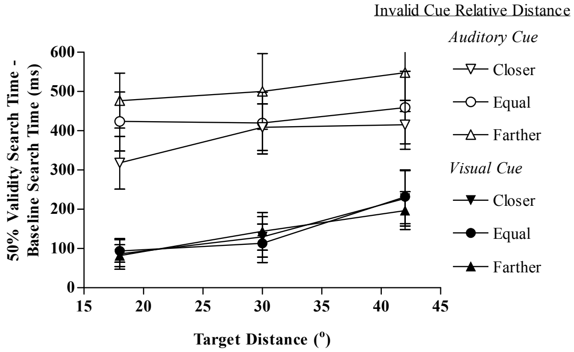

Figure 4A: Validity=50%. Change in search time relative to baseline as a function of target distance and invalid cue relative distance.

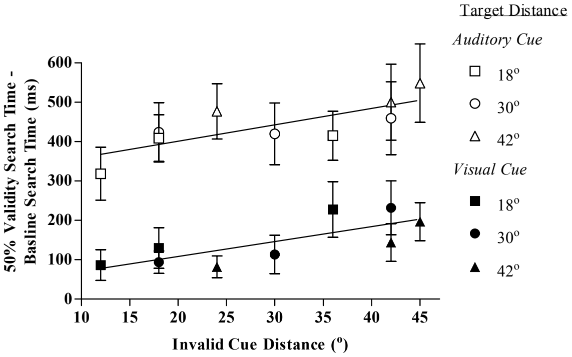

Figure 4B: Validity=50%. Change in search times relative to baseline as a function of invalid cue and target distance from fixation.

Figure 4 presents the change in search times for the 50% validity condition. Significant main effects of distance [F(2,10) = 13.36; p = .001] and invalid cue relative distance [F(2,10) = 3.84; p = .05] were obtained. A significant Cue x Invalid Cue Relative Distance interaction was also obtained [F(2,10) = 11.67; p = .002]. As shown in Figure 4A, search times in each modality increased relative to the baseline search condition; the cost was greatest for valid auditory cues. The relative distance of the invalid cue affected search time for valid auditory cues most. We obtained the greatest increase in search times when the invalid visual cues were farther from fixation. When we replotted the search times against the actual invalid cue distance in Figure 4B, most of the search time variability was accounted for. Here, the slopes are nearly identical (4.2 ms/^{o} and 3.8 ms/^{o} for auditory and visual cues respectively.)

4. DISCUSSION.

We obtained no adverse effects of invalid auditory cues when the validity of the visual cues was 90%. Instead, the auditory cues reduced the time to locate and identify the target, contrary to previous experiments on auditory spatial cueing. Rudmann and Strybel (1999) showed that search times with auditory cues that were displaced from the target were longer than when no cue was provided in some conditions. Spence and Driver (1997) showed that auditory cues can be stimulus-driven in that they increased reaction time when inaccurate. We believe the difference in experimental outcomes between our findings and previous research is due to either the absence or inadequacy of the visual cue in past works. No visual cue was provided with the auditory cue in Rudmann and Strybel, although the pattern of distractors sometimes cued the local target area. Spence and Driver tested each cue in separate sessions; the cues were never presented simultaneously. Our results at 90% validity are consistent with Frens et al.'s (1995) findings regarding eye movements to multimodal stimuli. The latency of the eye movements to high intensity auditory and visual stimuli were reduced by approximately 50 ms, relative to the saccadic latencies to the visual stimuli in isolation. This effect diminished as the stimuli were spatially separated. Note in Figure 3B that the benefits obtained here are diminishing slightly with increasing distance, as predicted by Frens et al.

When the validity of each cue was 50%, our participants usually searched the visual cue before the auditory cue because the cost of the invalid cue was less, as shown in Figures 4A and 4B. With both cues the cost of the invalid cue increased with its distance. We believe this is due to the time required to move from an invalid cue to a valid cue. As the distance of the invalid and valid cue increased, it took longer to reorient to the cue in the opposite hemifield. Moreover, given that the visual cue produced faster search times in isolation (as shown in Figure 2), participants probably oriented to the visual cue more quickly after searching an invalid auditory cue than they oriented to the auditory cue after searching an invalid visual cue.

In summary, our results suggest that stimulus-driven visual cues interfere more with auditory cues even when users know the auditory cues are valid. Users prefer visual cues to auditory cues when neither cue is reliable. Designers of multisensory environments could probably provide auditory spatial cues without concern about visual cue interference.

On the other hand, designers should be careful about using auditory spatial cues to signal the most critical events, because visual onsets may interfere with them. These recommendations are limited, however, to the specific stimulus characteristics tested here. From Figure 2, it appears that our visual cue was more conspicuous and accurate than the auditory cue. Degrading either the contrast or the accuracy of the visual cue might produce a greater effect of auditory spatial cues. Suggesting this possibility, Frens et al. (1995) showed that saccade latencies to low-intensity visual cues increased considerably, but saccade latencies to low intensity auditory cues were not affected.

REFERENCES.

Bullinger, H., J., Bauer, W., and Braun, M. (1997).Virtual Environments. In G. Salvendy (ed.) Handbook of Human Factors and Ergonomics 2nd Edition (pp. 1725-1760). New York: Wiley.

Egeth, H.,E., and Yantis, S. (1997). Visual attention: control, representation, and time course. Annual Review of Psychology, 48. 269-297.

Frens, M., A., Van Opstal, A., J., and Van der Willigen, R., F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Perception and Psychophysics, 57, 802-816.

Fujawa, G., and Strybel, T., Z. (1997). Auditory spatial facilitation with valid and invalid cues. Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, (pp. 556-560). Santa Monica, Ca: Human Factors and Ergonomics Society.

Stein, B., E., and Meredith, M., A. (1993). The Merging of the Senses. Cambridge, MA: MIT Press.

Rudmann, D., and Strybel, T., Z. (1999). Auditory spatial facilitation of visual search performance: Effect of local vs global visual distractor density. Human Factors, 41, 146 - 160.

Spence, C., and Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. Perception and Psychophysics, 59, 1-22.

Strybel, T., Z., Boucher, J., Fujawa, G., and Volp, C. (1995). Auditory spatial cueing in visual search tasks: effects of amplitude, contrast and duration. Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, (pp. 109-114). Santa Monica, Ca: Human Factors and Ergonomics Society.

Strybel, T., Z., Vu, K.-P., L., and Castagna-Osorio, J., T. (2000). Auditory spatial cueing of visual search in the presence of noise (Abstract), Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting. Santa Monica, Ca: Human Factors and Ergonomics Society.