USING AN AUDITORY DISPLAY TO MANAGE ATTENTION IN A DUAL TASK, MULTISCREEN ENVIRONMENT

Derek Brock, Janet L. Stroup, and James A. Ballas

Naval Research Laboratory
Washington, DC 20375, USA
`brock|stroup|ballas@itd.nrl.navy.mil`

ABSTRACT

Spatialized sound technology is under consideration for use in future U. S. Navy watchstation systems as a technique to manage attention. In this study, we looked at whether spatialized sound would reduce head movements. The subjects used a simulated watchstation that had three displays, one forward and one on each side. A dual task paradigm was used that included a continuous tracking task in one window and an intermittent task in another window. These two windows were presented adjacent to each other in the center display or opposite each other on the side displays. Subjects performed the dual task with and without sound. Head turns were recorded manually and were found to be significantly fewer in number when sound was present. Further, when sound was present, subjects used its cessation as an indication of a successfully entered response. This aural feedback reduced head movements that would normally be made to confirm the successful data entry. Together with other results on reaction time and accuracy, these results provide persuasive support for the use of spatialized sound to direct attention.

1. INTRODUCTION

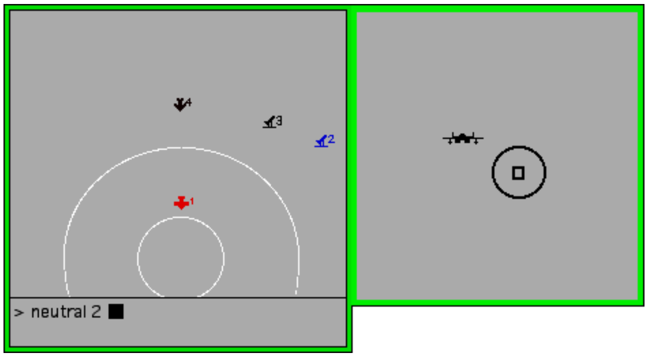
Spatialized, or 3-dimensional (3D), auditory display technology is being seriously considered for incorporation in future U. S. Navy watchstation systems for a number of compelling reasons. On the basis of new defense realities, it appears that greater levels of tactical, situational, and systems information will soon have to be managed by smaller numbers of working personnel in command and control centers. Recent work on this problem has resulted in an advanced multi-modal watchstation design [1] that proposes to give operators responsibility for as many as three separate tasks at a time, with each task displayed on its own flat panel monitor (see Fig. 1). In time-driven situations, where cognitive load and competing performance demands are likely to be high, system level aids for managing attention are certain to be needed. Among its numerous information capacities, advanced auditory display technology has a unique potential for directing visual attention using spatialized sound (auditory deixis) and affording nonvisual monitoring of processes [2]. If this potential is realized in terms of measurable performance gains, auditory display has a core role to play in the more fully human-centric user interfaces that are needed for tomorrow's tactical decision environments.

2. BACKGROUND

The present research makes use of a dual task paradigm that was initially developed at the Naval Research Laboratory in the early '90s for direct manipulation interface research [3]. Modified versions of this testbed have been used extensively for purposes of cognitive modeling and auditory display research [4]. Fig. 2 shows the visual presentations of the two tasks, which were originally conceived to resemble a class of partially automated tasks performed in combat aircraft.

Figure 1. Illustration of multi-modal watchstation (MMWS).

The task shown on the right in Fig. 2 simulates an aerial pursuit. Subjects are asked to continually track the target (an animation of an airplane, as seen from behind) with a right-handed joystick that controls a pipper and reticle (the circle-shaped cursor with a square in its center). Keeping the pipper on or near the target is sufficiently difficult by itself to demand the subject's full visual attention.



> neutral 2 ■

Figure 2. Illustration of the two task windows

The task shown on the left in Fig. 2 presents the subject with an intermittent tactical decision task in which potential threats ("tracks") must be classified as hostile or neutral based on a set of rules for their behavior. The tracks are depicted as enumerated icons that represent fighters, cargo planes, and surface-to-air missile sites, which move down the simulated radar screen as the subject's own aircraft supposedly moves forward. The decision making task is assisted by a supposedly automated assessment mechanism that color codes each track as hostile (red) or neutral (blue), or unknown (yellow) when a definitive assessment cannot be made. When a track is yellow, subjects must decide whether it is hostile or neutral on the basis of its behavior. Otherwise, the task is simply to confirm the automation's assessment of a given track. Subjects indicate their decisions and confirmations with their left hands on the numeric keypad of a standard computer keyboard. A complete response requires two keystrokes: one to indicate hostile or neutral and a second to indicate the track number.

In the dual task setting, subjects are expected to try to carry out both the continuous tracking task and the tactical decision task at the same time.

In natural environments, deictic sound—sound that comes from and, hence, indicates a particular location—has been shown to improve visual monitoring and search performance [5]. Given this result, it is reasonable to ask if the use of an auditory display for similar purposes in a visually dispersed task environment will have practical and commensurate effects on task performance. To find out, a configuration of displays similar to the U. S. Navy's current multi-modal watchstation specification was assembled at the Naval Research Laboratory, and an experiment with a two-factor, repeated measures design was conducted using an appropriately modified version of the dual task paradigm described above.

3. METHOD

For the present study, a Crystal River Engineering Acoustetron was used to augment the tactical decision task with a 3D auditory display. The auditory design utilized three easily distinguished sound loops—a siren, the sound of an airplane propeller, and a diesel truck horn—which were assigned to the fighter, cargo plane, and surface-to-air missile site tracks, respectively. Each track's associated sound loop was spatialized and then played when the track changed color; this was also when subjects were expected to enter a response. The spatialization indicated deictically where the corresponding track was located in the tactical display. The visual positions of the tracks were mapped to left, right, and center positions in the auditory display's 3D listening space, which was presented with headphones. In effect, space in the auditory display was designed to correspond to an egocentric mental representation of the tactical decision world. Sound loops were presented one at a time and always corresponded to the oldest unacknowledged track in the visual display. Each sound loop was immediately stopped when its corresponding tactical decision was entered.

Sixteen volunteers (4 women and 12 men) from among the Laboratory's staff participated in the study. Subjects were trained to perform the continuous tracking and tactical decision tasks separately and then carried out four dual task exercises (in diagram balanced order). In each of these exercises, subjects saw a different scenario of 65 tactical decision events under a different treatment within a two-factor design. The first factor manipulated the distance between the visual presentations of the two tasks; the tasks were either presented beside each other on the center screen, as they are depicted in Fig. 1, or on the left and right screens opposing each other. The angular distance between these two screens is enough to prevent peripheral visual monitoring of one screen while looking at the other, and it is essentially on this opposing condition that the work described in this paper is focused. The other factor manipulated in the experiment was the use of the auditory display, i.e., sound vs. no sound.

As subjects engaged in the dual task exercises, their head movements were passively observed and recorded by the experimenter. In pilot studies, it was found that manual coding of head movements using the system described below was preferable to the much greater cost of gleaning the same information from an automated source such as a head tracking and/or eye-tracking device. An extremely low manual coding error rate in the full experiment (< 0.15%) affirmed the merit of this choice[1]. Head movements were recorded with a handheld personal digital assistant (PDA) running a program called Flexible Interface Technique System (FIT-System) from SmileDesign <http://www.smiledesign.ch>. In this data collection technique [6], a template of event boxes is created as an overlay for the PDA's screen. For this study, a template was created to record head movements directed at five locations: the three monitors (left, right, and center), the joystick, and the keyboard (see Fig. 3). As the subject moved his or her head, the experimenter tapped the PDA screen in the box corresponding to where the subject was looking, and the FIT-System software recorded the tap location and time stamped it. No attempt was made to record subjects' eye movements when the tasks were presented beside each other on the center monitor. (While it may have been possible to gather such information with a video camera, this was not done since evaluating performance in the opposing screen task condition was the real focus of the study.)

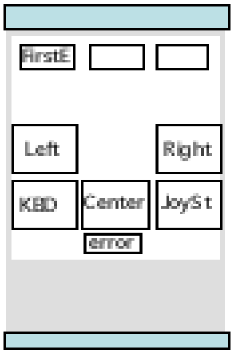


Figure 3. *FIT-System PDA template used by the experimenter.*

At the conclusion of the exercises, the FIT-System files were transferred from the PDA to a desktop computer and processed to generate event frequencies, event timing, and link frequencies, which show the number of times a head movement went from one area to another. Some of this processing is done using Excel macros that come with the FIT-system software.

4. RESULTS

Since it was not possible for the experimenter to record subjects' eye movements when the tasks were presented together on the center monitor, only head movement data that was collected for the two treatments in which the tasks were displayed on the right and left monitors are evaluated here. In addition, an equipment failure invalidated the data collected for one of the subjects. Tallies of the number of head movements made by all 15 subjects from one task-related location to another under these treatments are shown in Tables 1 and 2. Table 1 shows subjects' head movements when the auditory display was present and Table 2 shows the corresponding numbers when it was absent. Since the numbers of head turns to and from the center monitor (which was not used in either of these treatments) and the joystick were negligible, these data are not shown.

sound	>Right	>Left	>Keybd	
Right>	0	2068	130	2198
Left>	1908	0	456	2364
Keybd>	289	297	0	586
	2197	2366	586	5150

Table 1. *Head movements from one location to another when the continuous tracking and tactical decision tasks were displayed on the right and left monitors, respectively, and the auditory display was used.*

No sound	>Right	>Left	>Keybd	
Right>	0	2888	142	3030
Left>	2790	0	478	3268
Keybd>	244	376	0	620
	3034	3264	620	6928

Table 2. *Head movements from one location to another when the continuous tracking and tactical decision tasks were displayed on the right and left monitors, respectively, and the auditory display was not used.*

A one way, repeated measures ANOVA of total head movements for each subject under the two conditions (sound vs. no sound) shows that the differences between these two tables is highly significant ($F(1, 14) = 23.517$, $p < .001$). The mean number of head movements per subject was 343.2 when sound was used and 461.2 when it was not.

Table 1 and 2's marginal totals for head turns away from the right monitor, where the continuous tracking task was displayed, specifically, 2198 when sound was used and 3030 when it was not, are of particular interest. Since the angular distance between the opposing screens prevented peripheral visual monitoring of the tactical decision task on the left while attending to the tracking task on the right (and vice versa), subjects in the no sound condition had no choice but to turn repeatedly to the left monitor to discover when tactical decisions needed to be made. In the sound condition, though, the auditory design was such that it was possible for subjects to rely on the onset of any of the three sounds to alert them to a decision event. Evidently, they did. This auditory monitoring explanation also accounts for the roughly equivalent difference between the two tables' marginal totals for head movements away from the monitor on the left (compare 2364 with 3268).

In fact, all of the numbers in Table 1 are less than their counterparts in Table 2 with the exception of head movements from the keyboard to the right monitor. In the sound condition, this number is 289, and in the no sound condition, it is 244. These numbers represent the number of times subjects stopped looking at the keyboard and then returned directly to the tracking task. (It is a fair assumption that the subjects were looking at the keyboard to be sure of the accuracy of their key presses.) The most plausible explanation for this particular difference turns out to be another instance of auditory monitoring. Even though the left monitor and the keyboard were near each other in the experimental setup, subjects in the no sound condition had to return their visual attention to the tactical decision display (i.e., the left screen) if they wanted to confirm the effect of their keystrokes. In the sound condition, though, each sound loop immediately stopped playing when its corresponding tactical decision was entered. As a consequence, it was also possible for subjects to rely aurally on the cessation of sound as a means of confirming their entry. They then could bypass the tactical display and return directly to the tracking task. This explanation is borne out by the substantially larger number of head movements from the keyboard to the left monitor in the no sound condition (i.e., 376 vs. 297) and the slightly larger marginal total for all head movements from the keyboard in the same table (620 vs. 586).

5. CONCLUSIONS

These results confirm the adage that a "function of the ears is to point the eyes" [7]. While there is a large amount of evidence for this adage, research that shows the effectiveness of spatialized sound within the specifics of new systems concepts is useful in building a case to move the finding from the laboratory to the field. Somewhat novelly, this study shows that spatialized sound can significantly reduce head movements between two visually dispersed task displays when a person must monitor both tasks nearly continuously. And the reduction in head movements comes without a reduction in accuracy [8] on either task. In fact, not only were there fewer head movements between the displays when the sound was present, but response times were also significantly quicker [8]. Without the auditory display, average response time increased from 2655 to 3302 ms when the task displays were separated; with the auditory display, response time increased from 2411 to 2677 ms. And note that with the presence of the auditory display, subjects were almost fully able to compensate for the increase in response time that was due to separating the tasks (compare 2677 ms with 2655 ms). The reduction in head movements, together with the improved response time and stable accuracy, shows that spatial sound supports an improved efficiency in attention. Sound usually supports multiple functions simultaneously. In the present case, the reduced head movement came about because of the sound's alerting and the deictic functions. That is, the sound in this experiment indicated both when a response was due and where to look to find the visual track. And because the sound ended when a successful data entry was produced, its cessation provided an indicator that the subject had successfully entered a response. This aural feedback reduced head movements that would normally be made to visually confirm that the data entry had been successful. This is event status monitoring, and while we often think of using auditory displays to monitor continuous background events, it is also useful to monitor discrete events. These results provide compelling support for the use of spatialized sound to manage attention.

6. ACKNOWLEDGEMENTS

This work was conducted with sponsorship from the Office of Naval Research.

7. REFERENCES

[1] G. Osga, “21st Century Workstations: Active partners in accomplishing task goals,” in *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, San Diego, CA, August 2000.

[2] J.A. Ballas, “Delivery of information through sound,” in G. Kramer (Ed). *Auditory Display: Sonification, Audification and Auditory Interfaces*. SFI Studies in the Sciences of Complexity. Proceedings Volume XVIII, Addison-Wesley, Reading, MA, 1994.

[3] J.A. Ballas, C.L. Heitmeyer, and M.A. Pérez, “Evaluating two aspects of direct manipulation in advanced cockpits,” in *Proceedings CHI `92,* Monterey, CA, 1992.

[4] D. Kieras, D. Meyer, and J. Ballas, “Towards demystification of direct manipulation: Cognitive modeling charts the Gulf of Execution,” in *Proceedings CHI `2001*, Seattle, WA , 2001.

[5] D.R. Perrot, J. Cisneros, R.L. McKinley, and W.R. D’Angelo, “Aurally aided visual search under virtual and free-field listening conditions,” *Human Factors*, pp. 702-715, 1966.

[6] J. Held and H. Krueger, “The FIT-System: A new hand-held computer tool for ergonomic assessment,” *Medical & Biological Engineering & Computing*, Suppl. 2, Vol. 37, pp. 862-863, 1999.

[7] E.W. Wenzel, “Spatial sound and sonification,” in G. Kramer (Ed). *Auditory Display: Sonification, Audification and Auditory Interfaces*. SFI Studies in the Sciences of Complexity. Proceedings Volume XVIII, Addison-Wesley, Reading, MA, 1994.

[8] D. Brock, J.L. Stroup, and J.A. Ballas. “Effects of 3D auditory cueing on dual task performance in a simulated multiscreen watchstation environment.” *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting,* Submitted

---

[1] Altogether, only eighteen manual coding errors were identified. Of these, fifteen could be characterized as double entries, wherein the experimenter inadvertently double-tapped the PDA and created two entries for the same event. In each of these instances, the second entry was eliminated and its duration and ending time incorporated into the first entry. The remaining three errors were specifically noted as "error" events by the experimenter using the error box on the PDA template (see Fig. 3). In each of these cases, the event immediately prior to the tap in the error box was taken to have been incorrect. Each of these events was eliminated and its beginning time and duration (including the duration of the error entry) were incorporated into the event immediately following the error box tap.