Using an Auditory Display to Assist in Multiple Visual Tasks

Tim Halverson and Anthony Hornof
Department of Computer and Information Science
1202 University of Oregon, Eugene, OR 97403-1202 USA
{thalvers, hornof}@cs.uoregon.edu

ABSTRACT

This research investigates the use of auditory displays in complex tasks with multiple visual displays. Participants were trained extensively in a dual task that requires time-pressured, manual responses to stimuli on two displays. Continuous responses were required on the first display. Complex decisions were intermittently required on the second display. The number of decisions required within a short time varied from one to eight. A 2x2 factorial design was used. In two conditions the participants could not see the other display in their periphery. In two conditions auditory cues were presented for visual events on the second display. Each participant completed sessions across three consecutive days. It was found that auditory displays allow for considerable strategy optimization, but only when peripheral information was not available and only after considerable practice. Additionally, the lack of peripheral information negatively affected performance more when fewer decisions were required. Design implications are discussed.

Author Keywords

Multitasking, auditory displays, gaze contingency

ACM Classification keywords

H.5.2 User Interfaces: Auditory (non-speech) feedback, evaluation/methodology; Experimentation, Human Factors

INTRODUCTION

Auditory displays are routinely used to augment human-computer visual interaction. Auditory alerts have been used particularly well to direct users' visual attention to information when the user is engaged in another task (e.g. chimes indicating a background process has completed on a desktop computer system or warning sirens in power plant control rooms). The benefit of auditory alerts is even greater when the visual information is located in or beyond the user's peripheral vision.

The human-computer interaction (HCI) community has long had an interest in multiple display technology, but many related research questions remain open. Grudin [4] observed that the HCI communities understanding of how people use multiple monitors "for secondary activities related to principal tasks, for peripheral awareness of information that is not the main focus, and for easy access to resources" (p. 458) is lacking. Researchers have investigated *distributed display environments* [6] and *large display environments* [3]. However, the research on distributed display environments has largely focused on using windowing systems and navigating across large or multiple monitors, or to use peripheral displays to display information for secondary or lower priority tasks (e.g. ambient displays, peripheral displays, etc.) and has largely neglected the use of multiple displays for executing multiple equally important tasks in parallel; that is, multi-tasking.

Multitasking research in HCI tends to investigate issues such as when people choose to switch tasks [7] and how the device affects people's execution of multiple tasks, such as how automation and response input methods affect task execution [1]. The research reported here investigates how and when people can use spatialized, auditory alerts to improve their performance in a complex dual task spanning two displays.

This research advances an understanding of human performance when engaged in multitasking behavior that requires interaction with complex multimodal auditory and visual displays. The research investigates how people utilize auditory cues based on (a) whether the visual display of the second task is peripherally available or not and (b) task difficulty.

METHOD

An experiment was conducted in which participants performed a complex, interleaved dual task that included two visual displays, time-pressured manual responses to stimuli on both displays, and an auditory display to enhance performance on the secondary (radar) display. The task used in this study builds on and extends a task used in previous HCI studies [1].

Twelve participants, seven female and five male, between the ages of 18 and 51 (M = 26.8) from the University of Oregon and surrounding communities successfully completed the experiment. Participants were paid based on their speed and accuracy. Participants took part in the experiment on three consecutive days for approximately two hours on the first day and one hour on each subsequent day.

The tactical and tracking tasks were preformed in parallel and presented on one physical monitor. Each task utilized a separate visual display separated by 2.5 degrees of visual angle. The tactical task was presented on the left and the tracking task on the right.

CHI 2009, April 4–9, 2009, Boston, MA, USA.

Tactical Task

A series of "blips", 32 x 32 pixel icons, occasionally appeared at random locations in the tactical display and moved from top to bottom. Three types of blips were used: fighter, support plane, and missile site. Each blip type had a unique shape. The blips appeared in *waves* in which 1, 2, 4, 6, or 8 blips were visible.

The tactical task included quickly and accurately "classifying" each blip as "hostile" or "neutral", and using a keypad to enter the blip icon's number (1 through 9) and classification. "Touch typing" was enforced. The classification of the blips was determined based on a set of rules involving a blips shape, color, and movement. The hostility of the blips was sometimes indicated with auditory cues, with different sounds used for hostile, neutral, and unknown. The sounds were adapted from Cabrera, Ferguson, and Laing [2] with intensities normalized.

Each blip progressed on the display as follows: When each blip first appears on screen, it is black. At this point, the blip displays the characteristics (heading and velocity) that indicate its hostility, but the participant can not respond to the blip. When auditory cues are used, a woodblock-like sound indicated the appearance of the blip. Four to twenty-nine seconds later, the blips changed color to either green, red, or yellow. At this point, the participants could identify the blips using the keypad. After ten second, the blip would disappear. If a blip was classified before it disappeared, the blip color would change to white.

When the auditory cues were used, the location of each blip was indicated by the location of a spatialized auditory cue. To enable the participants to more accurately map the locations in auditory space to the locations in the visual space, auditory cues were transformed using the horizontal arc transformation from Hornof, Halverson, Isaacson, & Brown [5]. All sounds were broadband to facilitate spatial perception.

Participants were monetarily motivated as follows: Every time a blip appeared, they received six cents. For every second that passed before a blip was classified, one cent was lost. Every time a blip was incorrectly classified or not classified before it disappeared, the participant lost any possible bonus for that blip and was penalized an additional five cents. Any time another error occurred discussed in the last paragraph, the participant was penalized one cent.

Participants were provided with extensive visual and auditory feedback to motivate good performance. A status bar below the tactical display indicated how much the participant had earned with the previous ten blips.. Additionally, every time a blip was correctly identified, a pleasant "cha-ching" sound was played. Every time a blip was incorrectly identified or the participant entered an invalid key (e.g. a blip number that is currently not on screen), an annoying buzzer sounded. Other errors (e.g. taking too long to enter both blip number and hostility) resulted in a distinct but less annoying "bloop" sound.

Tracking Task

The tracking task, performed in parallel with the tactical task, required the participants to keep a circle over a moving target. The target was difficult to track for two reasons: First, the target moved in an unpredictable manner. Second, the joystick controlled the velocity of the circle. The joystick was industrial quality, from CH Products.

Participants were monetarily motivated to track well. Tracking accuracy was calculated from center of the circle to the center of the target. For every two seconds that the participant was tracking well (within 0.5 degrees of visual angle), the participant received slightly more than one cent. For every two seconds the participant was not tracking well (greater than 1.3 degrees of visual angle), the participant lost slightly more than one cent.

Participants were provided with visual feedback to motivate good performance. Whenever the participant was receiving money, the circle was highlighted in green. Whenever, the participant was loosing money, the circle was highlighted in red. Further, a status bar below the tracking display represented their average error over the past 40 s.

Procedure

Three factors were manipulated in this study to investigate how and when people can and do utilize spatialized, auditory cues. These three factors were: (a) gaze contingency, (b) the presence of the auditory cues discussed earlier, and (c) task difficulty, as determined by wave size..

Gaze contingency was used to simulate physically separated displays. Gaze contingency could be on or off. When it was on, the participant could only see the information (blips or tracking icons) on the display that an eye tracking device reported that the participant was looking at. Only the background of the opposite display was shown. Gaze contingency was used instead of physically separated displays to allow the collection of eye movement data with a single eye tracker.

The first two factors – gaze contingency and auditory display – were varied across four unique "scenarios", each lasting about eight minutes. In each of the four scenarios, gaze contingency was either on or off, and the auditory display either on or off. The third factor, task difficulty, was varied *within* each scenario.

On each of three consecutive days, participants completed four scenarios, each lasting about 8 minutes. In each scenario, sixteen waves were presented. There was usually a one second pause between each wave, with a 10 second pause after fifth wave and eleventh waves. The participants in this study were trained to criteria on all aspects of the tasks before starting actual data collection.

Visual stimuli were presented on a 1280x1024 LCD display with a Dual 2.5GHz PowerMac G5 running OS X 10.4.11. The auditory display was presented through Sennheiser HD250 headphones and generated by a VRSonic SoundSimCube. Eye movements were recorded using an LC Technologies dual-camera eye tracker. A chinrest was used to maintain a constant eye-to-screen distance.

RESULTS

Classification Task Reaction Time

The time for participants to correctly classify blips on the radar display were analyzed using a mixed model ANOVA with repeated-measures. Figure 1 shows the mean time to classify a blip as a factor of sound, gaze, and session day. Main effects of gaze contingency and session day were found. When people could only see the display they were gazing at, blip classification took longer, F(1, 6541) = 115.0, $p < .001$. The more experience a person had, the less time it took them to classify a blip, F(2, 6496) = 52.3, $p < .001$.

The use of sounds to indicate the appearance of a blip, and that a blip was ready to be classified, had no effect on the time to classify blips consistently across all conditions, F(1, 6561) = 0.5, $p = .478$. However, there were two and three way interactions involving sound that were significant. When the appearance and readiness of blips to be classified were sonified, gaze contingency slowed blip identification less, F(1, 3207) = 20.5, $p < .001$. Additionally, the time to classify blips when sounds were used decreased more with experience than when there were no sounds, F(2, 3197) = 5.99, $p = .003$.

More specifically, as seen in Figure 1, the participants' performance improved more over time when both gaze contingency and sounds were on, F(2, 1772) = 3.6, $p = .029$. That is, performance improved more across sessions in the condition where gaze contingency and sounds were used than in any other condition.

Figure 2 shows the mean time to classify a blip as a factor of gaze contingency and wave size. A main effect of wave size was found. The larger the wave in which a blip appeared, the longer it took people to identify the blip, F(4, 4683) = 96.2, $p < .001$. However, there was an interaction between gaze contingency and wave size, F(4, 4608) = 6.2, $p < .001$. Gaze contingency slowed responses less when there were more blips.

Classification Task Errors

Figure 3 shows classification errors. The number of errors decreased as the participants gained experience. Additionally, the numbers of incorrectly classified blips are similar across all four conditions. This suggests that the participants were not unduly sacrificing accuracy for speed across the conditions. The number of blips *not* classified was greater in the gaze contingent conditions, as would be expected when the participants cannot see them appear. However, by the third day, the number of blips not classified approached zero. This is especially true when the auditory cues were present. Task compliancy errors (such as stray keystrokes) are not shown due to lack of space, but are quite low, particularly after the first day suggesting that participants were well trained and motivated.

Tracking Data

Figure 4 shows the mean bonus money received by participants for the tracking task as a function of condition. The bonus money is a measure of overall tracking performance. The tracking performance trends are similar to the classification time trends. The participants performed better when gaze contingency was off and performance increased with experience. This suggests that the differences in classification time were indeed a result of the conditions affecting the classification task, and not simply an effect of the participants varying the time they spend on each task as a function of the condition.

DISCUSSION

This research investigated the usefulness of auditory displays to help monitor a second, critical task that is not always visually available. As with previous research, it was found that people can use an auditory display to effectively monitor a display that is not peripherally visible. This research replicated aspects of these previous findings, but with the use of a gaze contingent display rather than a physically separated display.

Additionally, the current research shows that people's utilization of the auditory display allows for considerable strategy optimization. This can be seen in Figure 1 by the larger decrease in classification time in the gaze contingency and sound condition (the dashed, gray line). The participants were able to improve their classification times on days two and three in all conditions. However, on

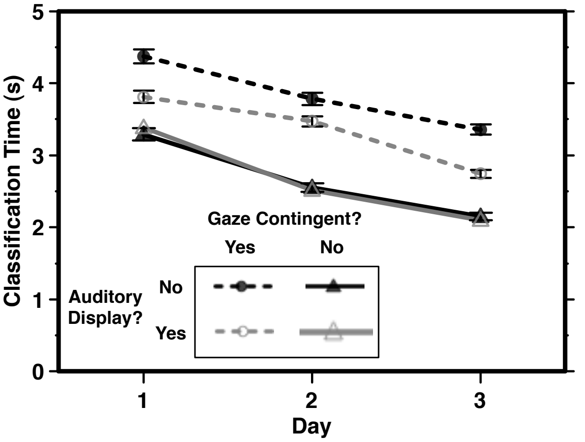

Figure 1. Classification time as a factor of sound, gaze and session day. Error bars ±1 SE of all blip classification times.

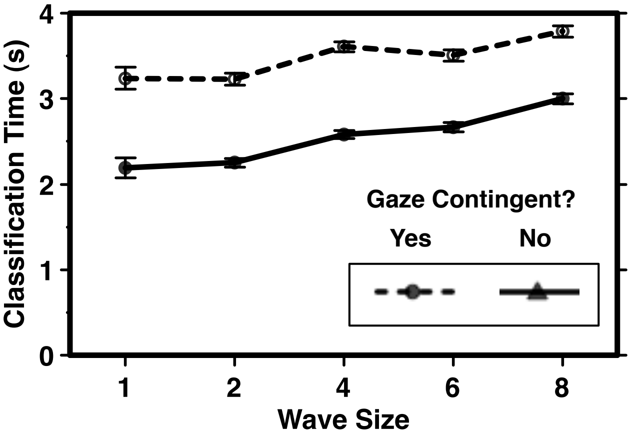

Figure 2. Classification time as a factor of wave size and gaze contingency. Error bars ±1 SE of all blip classification times.

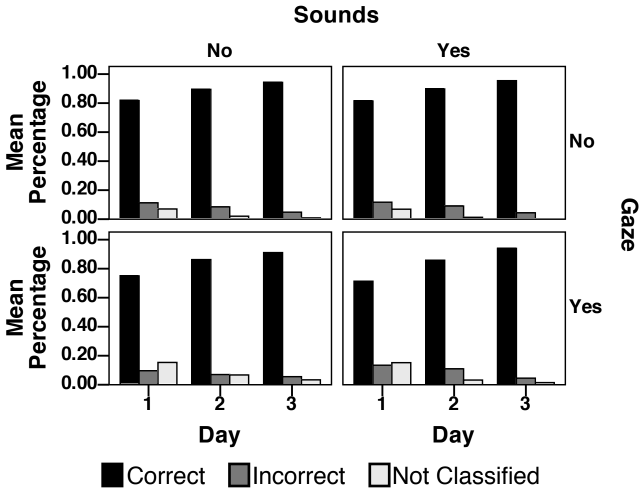

Figure 3. Accuracy as a factor of sound, gaze and session day.

day three, they were able to improve their performance even more when the auditory display was used.

Users appear to require more practice in order to form efficient strategies that capitalize on the use of auditory cues. The participants were trained to criteria on all aspects of the dual-task and then performed the task on three consecutive days. It was not until the third day that participants were able to tune their strategies to effectively utilize the auditory display. This suggests that any evaluation of auditory displays, especially in time-pressured, secondary tasks, should include extensive training and practice before evaluating the effectiveness of the auditory display.

The results of this research also suggest that audio should not be used to signal every event in a peripheral task. While the lack of peripheral visibility slowed blip classification time overall, gaze contingency did not slow blip classification in larger waves as much as it did in smaller waves. This can be seen in the interaction of gaze contingency and wave size shown in Figure 2. That is, when there is more happening on the secondary display, people tend to attend to the secondary display more. Therefore, it is more likely that the user will be looking at the secondary display when information appears, making the use of sound as a cue for this new information unnecessary. If only key events (for example, the start of a wave of blips) were cued by the auditory display, it's possible that people could perform just as well as when all blips are sonified. This conclusion is supported by participant comments. Several participants said that they tended to ignore the auditory cues after a few had sounded within a short time. One participant commented, "The sounds are annoying... especially with lots of blips. It's annoying to hear all of [the blips] announced." Further research is required to determine whether a more "sparse" auditory display will result in equivalent or better performance and less "annoyance."

CONCLUSION

In addition to advancing an empirical understanding of how people conduct multiple visual tasks in parallel with the help of auditory displays, this research found empirical support for two design recommendations related to the use of auditory displays. First, it may take users longer to optimize their strategies to use auditory displays. Therefore, evaluations of auditory displays should include extensive training and practice for the users. Second, although auditory alerts are useful for improving users performance in some circumstances, the efficacy of individual auditory alerts decrease when they are used to indicate multiple visual events that are spaced closely in time. Therefore, it may be more useful to use a single auditory alert for temporally contiguous visual events.

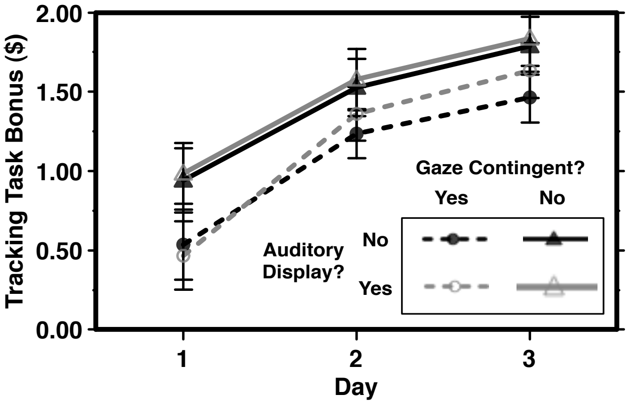

Figure 4. Tracking task bonus as a factor of sound, gaze and session day. Error bars ±1 SE of the 16 participant bonuses.

The next stage of this research project will investigate *when* and *how* people use auditory alerts by building computation cognitive models of the user interaction and analysis of users' eye movements.

REFERENCES

1. Ballas, J. A., Heitmeyer, C. L., & Pérez, M. A. (1992). Evaluating two aspects of direct manipulation in advanced cockpits . *Proceedings of the Conference on Human Factors in Computing Systems,* 127-134.
2. Cabrera, D., Ferguson, S., & Laing, G. (2005). Development of auditory alerts for air traffic control consoles. *119th Audio Engineering Society Convention,*
3. Czerwinski, M., Robertson, G., Meyers, B., Smith, G., Robbins, D., & Tan, D. (2006). Large display research overview. *Proceedings of the Conference on Human Factors in Computing Systems,* 69-74.
4. Grudin (2001). Partitioning digital worlds: Focal and peripheral awareness in multiple monitor use. *Proceedings of the Conference on Human Factors in Computing Systems,* 458-465.
5. Hornof, A. J., Halverson, T., Isaacson, A., & Brown, E. (2008). Transforming object locations on a 2D visual display into cued locations in 3D auditory space. *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society,*
6. Hutchins, D. F., Stasko, J., & Czerwinski, M. (2005). Distributed display environments. *Interactions, 12,* 50-3.
7. Kushleyeva, Y., Salvucci, D. D., & Lee, F. J. (2004). Deciding when to switch tasks in time-critical multitasking. *Cognitive Systems Research, 6*, 41-9.