Knowing Where and When to Look in a Time-Critical Multimodal Dual Task

Anthony J. Hornof, Yunfeng Zhang, and Tim Halverson

Computer and Information Science

University of Oregon

Eugene, OR 97403 USA

{hornof, zywind, thalvers}@cs.uoregon.edu

ABSTRACT

Human-computer systems intended for time-critical multitasking need to be designed with an understanding of how humans can coordinate and interleave perceptual, memory, and motor processes. This paper presents human performance data for a highly-practiced time-critical dual task. In the first of the two interleaved tasks, participants tracked a target with a joystick. In the second, participants keyed-in responses to objects moving across a radar display. Task manipulations include the peripheral visibility of the secondary display (visible or not) and the presence or absence of auditory cues to assist with the radar task. Eye movement analyses reveal extensive coordination and overlapping of human information processes and the extent to which task manipulations helped or hindered dual task performance. For example, auditory cues helped only a little when the secondary display was peripherally visible, but they helped a lot when it was not peripherally visible.

Author Keywords

Auditory displays, cognitive strategies, eye tracking, multimodal, multitasking, visual displays.

ACM Classification Keywords

H.5.2. [Information Interfaces and Presentation]: User Interfaces – Auditory (non-speech) feedback, evaluation/ methodology, screen design, theory and methods; H.1.2 [Models and Principles]: User/Machine Systems – Human factors; Human information processing.

General Terms

Experimentation, Human Factors, Measurement.

INTRODUCTION

An important task domain for human-computer interaction (HCI) is high-performance, time-critical, real-time systems intended to support multiple tasks in parallel. This domain is of great interest to designers of devices that might be used in high-performance life-critical multitasking situations such as air-traffic control, in-car navigation, and emergency evacuation. Such devices present users with the challenge of accomplishing a secondary task, such as

providing manual responses to visual information, so that they can quickly return to the primary critical task.

Computational cognitive models for predicting human performance in HCI tasks (such as [8, 13, 14]) need to accurately account for multitasking performance. Cognitive models often focus on cognitive strategies, the deliberate though often subconscious plans for coordinating and overlapping perceptual, decision, memory, and motor processing to do a task. Yet there is limited human data available to (a) reveal the strategies that people develop in high performance multitasking environments, (b) show how various information processes are overlapped, and thus (c) guide the development of predictive modeling techniques.

One approach to modeling skilled human-computer interaction uses constraint satisfaction to build strategies based on individual information processing steps (such as single eye movements) or substrategies (assemblies of several steps) [14]. More data is needed to explore how well a constraint-based approach such as this can predict dual task performance. Another modeling approach, intended for predicting dual task performance, posits that multiple tasks are threaded together through a greedy but polite sharing of processing resources [13]. More human data, including eye movement data, is needed to develop techniques for predicting dual task performance.

A dual task experiment conducted in the early 1990s at the Naval Research Laboratory (NRL) [2] produced human speed and accuracy data that proved useful for developing detailed computational cognitive models of dual task performance [7, 8, 9]. In the NRL dual task, participants use a joystick to track a moving target on one display and, in parallel, key-in responses to objects that appear on a secondary "radar" display. This paper presents an experiment that extends the original NRL dual task in numerous important ways, including:

(a) Eye movements are recorded for a more detailed account of how people interleave the two subtasks.
(b) Eye tracking is used in some conditions to hide objects on the not-currently-looked-at display, to simulate two displays that are separated by a substantial visual angle.
(c) Auditory cues are enhanced to map more closely to required responses.
(d) Participants are rigorously trained, financially motivated, and given extensive feedback so that their performance approaches an expert level.

This paper presents extensive human data, including eye movement data, that reveal more than just summary statistics [12], but that instead tell a rich story of the cognitive strategies that people develop in complex multimodal (in this case auditory and visual) multitasking scenarios. Strategies include (a) substantial overlapping of human information processing within and across two tasks, (b) precise decisions regarding when and where to move the eyes, and (c) the use of auditory cues to augment and sometimes replace the need for visual processing.

METHOD

Two tasks (or subtasks) were performed in parallel: a tracking task and a radar classification task. The tracking task is considered the primary task because it requires continuous visual attention and manual responses, like steering a car, whereas the classification task is considered secondary because it permits intermittent visual monitoring and requires just fifty-seven responses across each eight- or nine-minute scenario.

The experiment has a 2x2x5 within-subjects design. Three factors were manipulated: (a) peripheral visibility on or off, (b) auditory cues present or absent, and (c) wave size. Each scenario maintained a unique combination of the first two factors, whereas the third was varied within each scenario. Figure 1 shows the visual displays for each of the two subtasks, which were presented side-by-side on a single computer monitor, the radar on the left and the tracking on the right.

Peripheral visibility manipulated whether participants could see the contents of the *other* display—radar or tracking—that they were not currently looking at. When peripheral visibility is on, all visual information is available all the time. When it is off, a participant can only see the information (blips or tracking icons) on the display that he or she is currently looking at, with each display updated within 40 ms of the eyes (point of gaze) arriving or leaving. This simulates a task environment in which visual displays are separated by enough distance such that they cannot be monitored with peripheral vision.

Classification Task

In the radar classification task, small icons referred to as *blips* appeared at random locations in the top half of the radar display, and moved slowly down the display. The participant's task was to key in, for each blip, the single numerical digit on the blip along with a single-key classification of *hostile* or *neutral*. Participants were trained to use the keypad without looking.

There were three types of blips—fighter aircraft, missile sites, and cargo airplanes—each of which had a different shape. While on the display, each blip maintained a constant speed and direction, but went through several changes in appearance. Every blip started as black. While
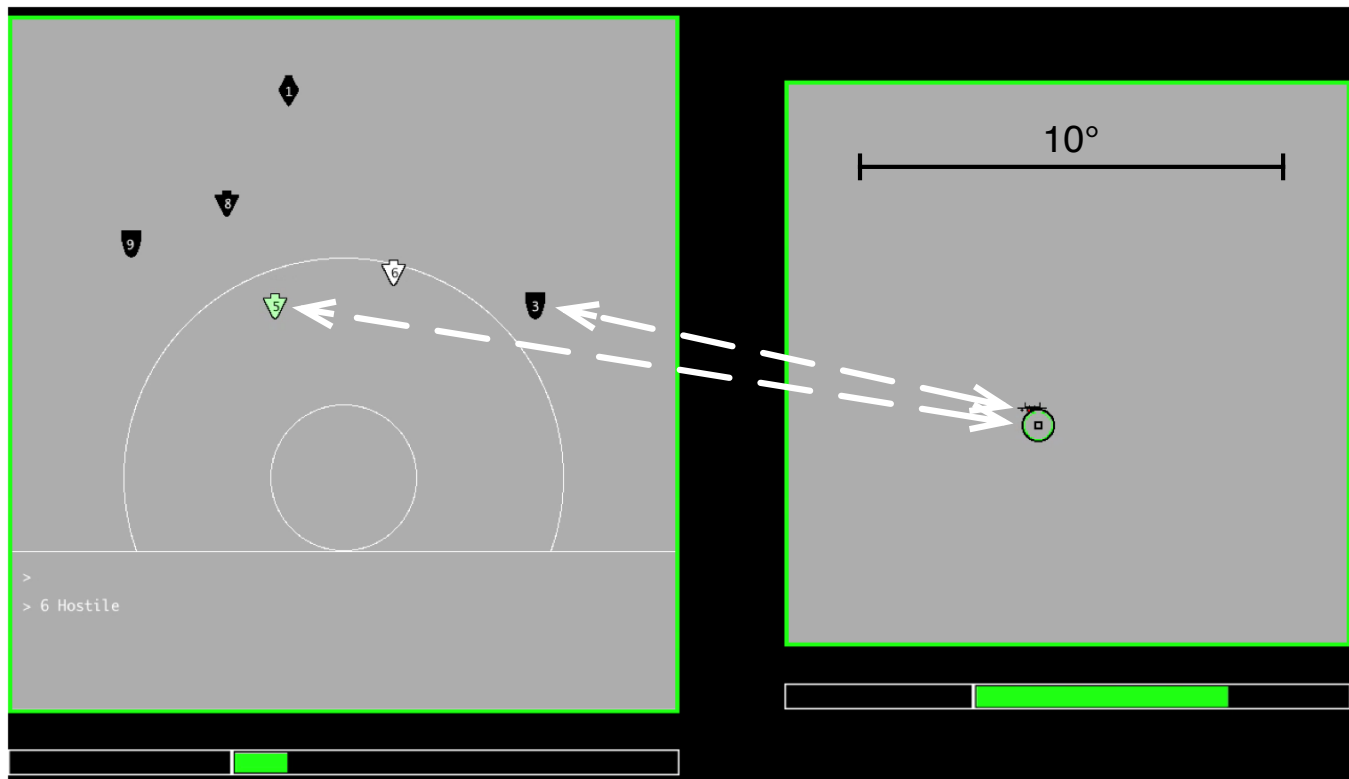

Figure 1. A screenshot of the task display annotated with two dashed lines that show the most common eye movements observed, between the radar blips on the left and the tracking target on the right. Radar blips were black before they were ready for classification; red, green, or yellow when ready-to-classify; and white after classification. Progress bars below the displays indicate that the participant is doing well on tracking, but needs to work harder on classification. A scale shows 10° of visual angle.

black, a blip's hostile or neutral status could be determined based on its shape, speed, and direction, following rules practiced before starting (the rules are detailed in [2] and [7]). But a blip's classification could not be keyed-in until it changed color.

After four to twenty-nine seconds (M = 13 s, SD = 5 s), each black blip changed to red, green, or yellow, at which point it was ready to be classified, and response time started. Red blips were hostile and green were neutral. Yellow blips had to be classified based on shape, speed, and direction (a harder task). After a classification response was keyed in for a blip, the blip turned white. All blips disappeared ten seconds after they became ready to classify.

When sound was present for the classification task, a blip's initial appearance and color change were indicated with spatialized auditory cues that were delivered via Sennheiser HD250 headphones. Blip appearance was cued with a (0.1 s) woodblock sound. The color change was cued with an alarm that indicated the blip's color. Alarms were from [4] as follows: Red was cued with seven pulses of 740Hz (that lasted 1.1 s); yellow with three pulses of 520 Hz (1.5 s); green with one pulse of 385 Hz (0.5 s). Auditory cues were mapped to visual locations using the most effective transformation discussed in [6]. All cue volumes were normalized before spatialization.

Blips appeared in *waves* containing 1, 2, 4, 6, or 8 blips. Each scenario included sixteen waves: three with one blip, five with two, four with four, two with six, and two with eight. Small waves (of one or two blips) had at least 4.5 s between blip color changes. Large waves (with four, six, or eight blips) had a blip changing color roughly every 2.7 s, with least 2 s between blip color changes (a harder task). Waves were separated by no blips on the radar display for 1 s, extended to 10 s twice per scenario. The ordering of waves and blips was varied across scenarios.

Extensive visual and auditory feedback motivated good performance. 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. A status bar below the radar display indicated how much money the participant earned with the previous ten blips. Financial incentives for the classification task were as follows: Each blip carried a bonus of up to six cents. Until it was classified or disappeared, one cent was lost per second. Every time a blip was incorrectly classified, all bonus plus an additional five cents was lost. Other errors cost one cent.

Tracking Task
In the tracking task, participants moved a joystick to keep a circle over a moving target. The task demanded constant attention because the target moved continuously in an unpredictable manner, and because the joystick controlled the velocity (as opposed to position) of the circle.

Financial incentives for the tracking task were as follows: Accuracy was calculated from center of the circle to the center of the target. For every second that the participant tracked well (error ≤ 20 pixels), the participant earned 0.6 cents. For every second that the participant did not track well (error ≥ 40 pixels), the participant lost 0.6 cents. Visual feedback helped to motivate good performance: When the participant was making money, the circle was highlighted in green; when loosing money, in red. A status bar below the tracking display reflected the average error of the past 40 s.

Procedure
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 completed the experiment. Each participated on three consecutive days, for roughly one and a half hours per day, and completed four scenarios per day. Each scenario lasted eight or nine minutes and presented a unique combination of the two factors of peripheral visibility and presence or absence of auditory cues; orderings were counterbalanced. Participants were trained to criteria for each of the two tasks individually before starting a scenario. Rewards for each subtask were reported after each scenario. Participants earned a base payment of ten dollars per hour plus an average of eleven dollars in bonuses per day.

Visual stimuli were presented on a 1280x1024 LCD display attached to a Dual 2.5GHz PowerMac G5 running OS X. The original experimental source code was acquired from NRL and rewritten (in C++ using Apple XCode) to interface the system with a VRSonic SoundSimCube and an LC Technologies dual-camera eye tracker. A chinrest was used to maintain a constant eye-to-screen distance. Two technicians staffed the three systems for all data collection.

RESULTS AND DISCUSSION
A tremendous amount of data was collected and analyzed. We believe that we have distilled this data down to a number of useful probe points that reveal how people did the task. We organize the data around (a) top-level results, (b) task strategies that participants developed for the different task conditions, (c) how participants determined *when* to move their eyes, (d) how they determined *where* to move their eyes, and lastly (e) specific evidence of overlapped perceptual, motor, and subtask performance in response to dual task challenges.

Statistical analyses were conducted using the R statistical environment [11]. Most used linear mixed-effects models and treatment contrasts for planned comparisons of simple effects. An alpha value of .05 was used for all analyses. Post hoc eye movement data error correction was conducted using the RFL (required fixation detection) technique [5]. The technique was extended to incorporate moving RFLs (the blips), and to incorporate multiple error-correction signatures across each scenario.

Top Level Results
A high-level first pass of the top-level performance across the three days and four conditions demonstrates that the

experiment successfully captured highly-motivated time-critical multitasking performance.

Figure 2 shows the mean blip classification time across all participants and all days.[1]   Some important trends in the data are as follows: (a) Peripheral-visible conditions are faster than peripheral-not-visible conditions, $F(1, 105) =$ 98.9, $p < .001$. (b) Whether sound is on or off only makes a difference when peripheral information is *not* visible, $t(105) = -2.95$, $p = .003$.   When peripheral visual information *is* available, sound does not improve blip classification times, $t(105) = 0.28$, $p = .78$. (c) Performance improves across the three days, $F(2, 22) = 36.1$, $p < .001$.
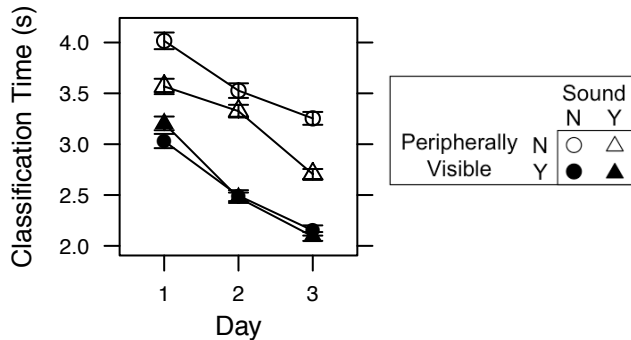

Figure 2. Blip classification time as a function of day, for each of the four conditions. Error bars (in all graphs) indicate standard error.

Three of the four conditions shown in Figure 2 show a negatively accelerating downward slope, as is typically observed as people learn a task over time, whereas one condition (peripheral-not-visible and yes-sound) shows a positively accelerating downward slope that does not resemble a typical learning effect.   This unique trend suggests that, in this one condition, participants did not just make small improvements to their strategy from Day 1 to Day 3, but rather that along the way they explored some different strategies. It appears as if participants may have learned how to use the sound to better monitor what was happening in the radar display when it was not visible.

Figure 3 shows how the percentage of blips classified correctly increased with each day, and that accuracy was better in the peripheral-visible condition, $F(1, 127) = 13.4$, $p < .001$.   Improved speed *and* accuracy in the same conditions demonstrate that participants did not simply shift their speed-accuracy operating criterion [10] based on the condition.

Figure 4 shows the mean tracking error—the average distance between the tracking circle and the tracking blip. Important trends are as follows: For all conditions, tracking error decreased across the three days, $F(2, 127) = 143$, $p < .$
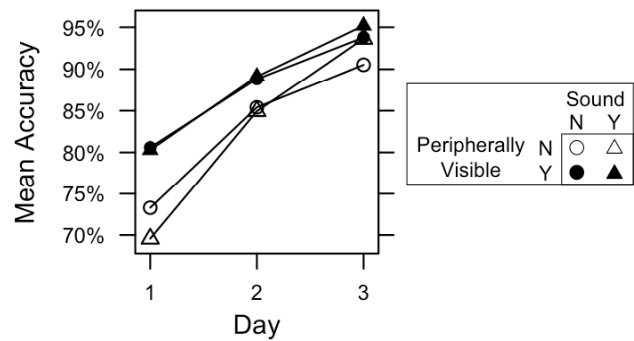
---

[1] A consistent set of plot symbols is used for all of the graphs that show the four combinations of factors. Perhaps use the following mnemonic to recall the symbols: Filled symbols represent that peripheral visual information was available because filled symbols have more visual information than unfilled.   Triangles represent that auditory cues were available because triangles have more "points" of information than circles.

Figure 3. The percent of blips classified correctly as a function of day for each of the four conditions.

001. Tracking error was always greater when no peripheral visual information was available ($F(1,127) = 48.3$, $p < .001$) but the difference decreased on Days 2 and 3.

Considering blip classification time in conjunction with the tracking error demonstrates that the experimental design, performance feedback, and payment scheme all worked well together to motivate participants to (a) work hard to improve their performance for *both* the tracking and classification tasks across all three days and (b) balance their performance appropriately between tasks—this is demonstrated by participants performing *both* the tracking *and* classification tasks slightly worse when peripheral visual information was not available.   The degraded performance in the classification task for peripheral-not-visible was not simply because participants spent more time and resources on the tracking task.   Clearly both tasks became harder and performance suffered in both.

Eye movement data were first analyzed at a high level to explore fundamental aspects of how participants did the task. Two results that are true across all conditions include: (a) Participants made 98% of their eye movements (saccades) to either the classification or tracking display (as opposed to the text output, status bars, or elsewhere), demonstrating that participants maintained strict focus on these two tasks.   (b) Participants spent about 86% of the time looking at the tracking display. Participants treated the tracking task as an important primary task such as steering a car, and the classification task as secondary task such as operating an onboard navigation system.
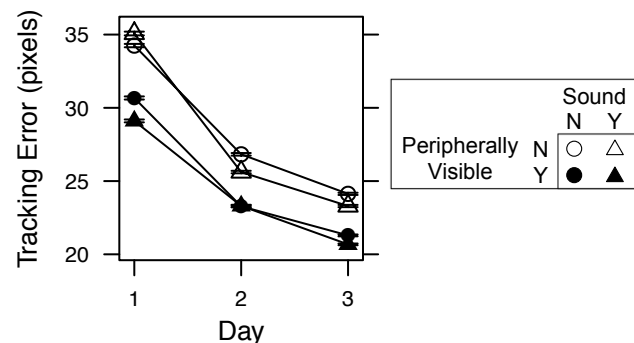

Figure 4. Mean tracking error (in pixels) as a function of day for each of the four conditions.

Overall performance data demonstrate that participants took the task seriously and maximized their performance. Day 1 performance is that of a novice whereas Day 3 performance starts to approach that of an expert. If we assume that people, when given sufficient feedback and motivation, will develop and refine task strategies to perform optimally, it is reasonable to conclude that the strategic decisions (such as eye movements) made on Day 3 are more optimal than those made on Day 1, and that expert multitasking strategies are emerging.

Task Challenges and Strategy Adaptations
The experiment varies task difficulty along a number of dimensions to reveal how different task strategies will evolve to integrate perceptual and motor processing to maximize performance. This section presents data that demonstrate how participants developed strategies to respond to the following task challenges: (a) Peripheral visual information was sometimes not available. (b) Auditory cues regarding blip status were sometimes not available. (c) Wave size varied from one to eight blips. (d) Yellow blips were harder to classify than red or green.

Task Challenges 1&2: No peripheral visual or auditory cues.
The experimental factors of peripheral visibility and sound affected task performance. Figure 5 shows the time interval from when a blip changed color (and became ready for classification) to when the eyes (the point of gaze) moved from the tracking to the radar display, for the 75% of the blip color changes that occurred when the eyes were on the tracking display. Important trends are as follows: With peripheral-visible, eye movements to the radar display were initiated much earlier than with peripheral-not-visible ($F(1, 105) = 239$, $p < .001$), leading to a faster classification (as seen earlier in Figure 2). With peripheral-visible, sound made little or no difference, $t(105) = -0.76$, $p = .45$. With peripheral-*not*-visible, sound made a significant difference, $t(105) = -4.27$, $p < .001$. The time to move the eyes to the radar display decreased across the three days, $F(2, 22) = 13.3$, $p < .001$. Note how Figure 5 resembles Figure 2, suggesting that the time required to get the eyes to the radar display largely determined the classification time.

Task strategies were developed to handle the experimental factors of peripheral visibility and sound. Shorter eyes-to-

radar times for peripherally-visible blips, as in Figure 5, suggest a strategy that responds directly to a peripherally-visible blip color change. The eyes moved to the radar an average of 674 ms after a peripherally-visible blip changed color regardless of whether auditory cues were available, $t(105) = -0.76$, $p = .45$. For color changes that were not peripherally-visible but signaled with auditory cues, participants used the cues to look at the radar an average of 984 ms after the color change; this use of auditory cues improved across the three days.

In the peripheral-not-visible and no-sound condition, participants made more eye movements to the radar display ($\chi^2(1) = 6.76$, $p = .009$), as would be needed to anticipate and respond to blip color changes. Figure 6 shows the mean number of fixations on each black blip for small and large waves. As can be seen, participants tended to look at each blip roughly once when it was black, except for small waves in the peripheral-not-visible no-sound condition, in which black blips tended to be fixated twice. Separating the counts by wave size helps to isolate the self-motivated glances (brief visits with the eyes) to the radar in the peripheral-not-visible no-sound condition because, for small waves in this condition, the gaze is less likely to be on the radar display for a previous blip color change than for large waves in this condition, and these self-paced glances are more necessary.
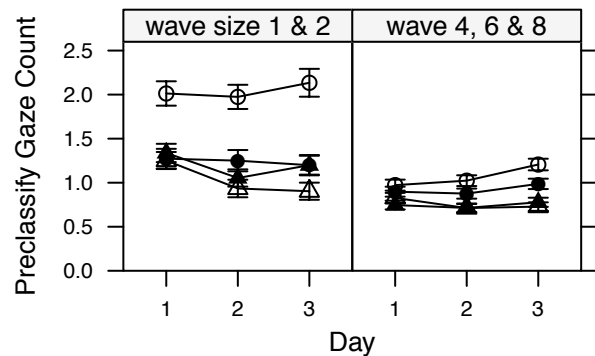
Figure 6. The mean number of fixations on black blips, for each of the four conditions, as a function of day. The left frame shows small waves; the right, large.

Task Challenge 3: Wave Size.
Increasing the wave size increased the task difficulty. Figure 7 shows blip classification times as a function of wave size, for each of the three days. The positive slopes in the curves shows how, in general, classification time increases with wave size, $\beta = 130$, $t(574) = 10.4$, $p < .001$. This is especially true in peripherally-available (filled plot symbol) conditions on Day 1, but the slope starts to level off by Day 3. In large waves, blips became ready-to-classify (roughly one every 2.7 s) at a slightly faster rate than the overall mean classification time of 2.9 s, and so some blips likely had to wait for others, sort of piling up. Small waves had at least 4.5 s between color changes, and would have had fewer such delays.
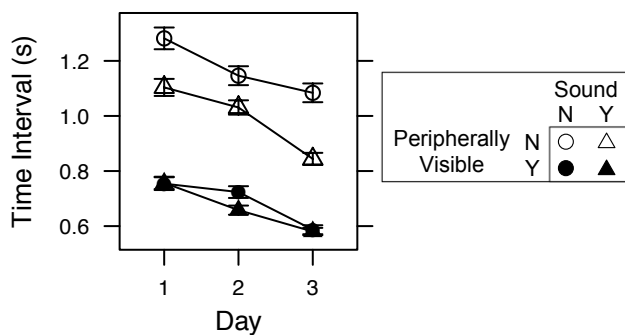

Figure 5. The time interval from when a blip changed color (and became ready to classify) to when the eyes moved from the tracking to the radar display, for each of the four conditions, as a function of day.
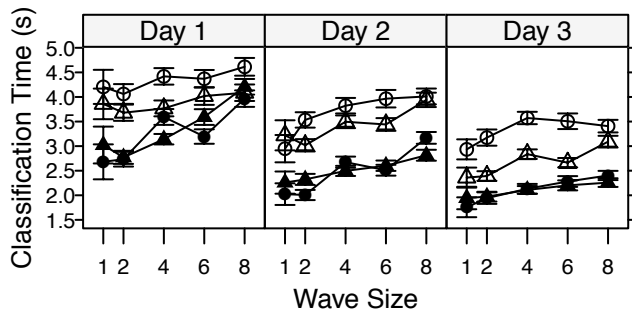
Figure 7. Blip classification time as a function of wave size, for each of the four conditions. The three panels show performance across the three days.

The similarity of the plots and gradual improvement across the three days in Figure 7 suggest that performance improved primarily from practice rather than refining strategies. But some data also suggest strategy refinements. For example, the changing relative distance between the unfilled circles and unfilled triangles across the three days suggest that participants, with practice, develop a strategy that better utilizes the auditory cues to keep track of what is happening on the radar when its blips are not visible.

Task Challenge 4: Yellow Blips.
Yellow blips presented a task challenge (in comparison to red and green blips) in that they required the participant to apply a set of rules to determine whether a blip was hostile or neutral based on its shape, speed, and direction. Figure 8 shows the time spent fixating and classifying blips as a function of blip color. Participants fixated yellow blips for an additional 237 ms (in comparison to red and green blips), perhaps to perceive the visual features such as speed. Yellow blip classification time took 483 ms longer than red and green blips. After moving the eyes off a blip, yellow blips required an additional 246 ms, relative to red and green blips, to translate the features into hostile or neutral.

To respond to the various task challenges, participants developed task strategies that (a) respond directly to visual stimuli when they are peripherally visible, (b) respond to auditory cues when visual stimuli are not peripherally visible, (c) periodically check on the secondary task display when no visual periphery or sound is available, and (d) over time improve the pacing of these self-motivated eye movements.
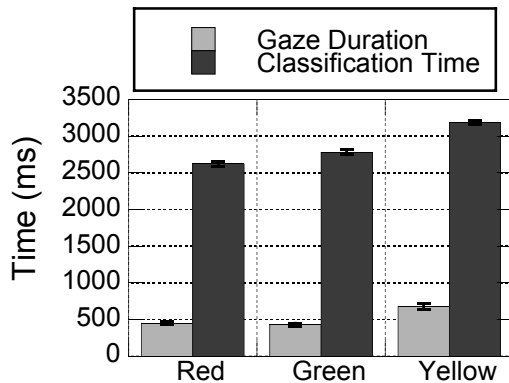

Figure 8. The time spent fixating and classifying blips.

Knowing When to Look
The data reveal how people decide *when* to move their eyes to accomplish multiple time-pressured tasks in parallel. In this task, participants primarily focus on tracking (spending 86% of the time on that display) but periodically need to move their eyes to classification. The main motivation to look at the radar is to get the information needed to classify a blip after it changed color. Knowing *when* to move the eyes requires participants to develop strategies to determine when this has happened or is likely to happen.

Two or three eye movements are required to classify a blip and resume tracking. Fifty-five percent of all blips were classified with the following two or three consecutive eye movements: (1) From the tracking to the radar (and usually directly to the target colored blip if peripherally-visible), (2) to the target colored blip (if not peripherally-visible), and (3) back to tracking. The other forty-five percent of the blip classifications varied slightly (such as to examine a black blip before returning to tracking) but these three movements are prototypical and reveal the fundamental overlapping of human information processing used throughout the task. Figure 9 shows the times (from all three days) that precede each of the three eye movements: eyes to radar, eyes to target, and eyes to tracking.
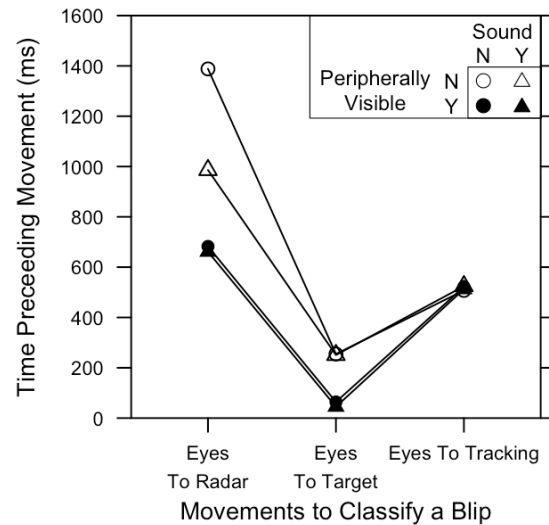

Figure 9. Time preceding eye movements across the lifetime of a colored blip. The *x*-axis shows the movements: From tracking to radar; to the target blip; and back to tracking.

1. Eyes to Radar (to Classify a Blip)
There are primarily three events that get the eyes to the radar: Participants see the color-change in the periphery, hear an auditory cue, or just decide that it is time to switch tasks. The overall eyes-to-radar response times are shown above the first tick mark on the *x*-axis in Figure 9. Participants respond most quickly when color-change events are peripherally visible (672 ms), more slowly to just-auditory color-change cues (986 ms), and slowest (1.4 s) when there is no visual or auditory indicator that a blip has changed color. As with many real-world situations, a visual stimuli in the near periphery is the quickest way to

notify when a secondary task needs attention, and an auditory cue is second-fastest. Self-paced task switching is slowest.

2. Eyes to Target

After the gaze arrives on the radar display, unless it arrived right on the target blip, the next step is to look at the target. As shown in Figure 9, the average eyes-to-target time is just 55 ms for peripherally-visible blips because the eyes typically move directly to the target, and 253 ms for peripherally-not-visible blips—just enough time to plan and execute a single eye movement to the just-appeared target blip.

3. Eyes to Tracking.

Once the gaze lands on the target blip, it stays there just long enough to perceive the visual features needed to classify the blip, and then typically returns directly the tracking. The duration is constant across all conditions, as shown in the convergence of all lines onto a single eyes-to-tracking point in Figure 9. Figure 10 shows how participants reduced their time spent studying a blip by about 200 ms across three days of practice, and appear to be approaching a somewhat optimal duration by Day 3.
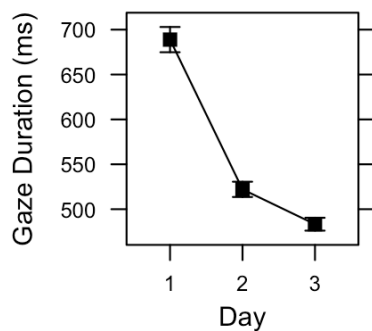

Figure 10. The time spent looking at a blip while it was in its colored state (and thus ready for classification) as a function of day.

Examining Blips Prior to Classification

Aside from moving the eyes to the radar to classify a blip, another reason to visit the radar when there is nothing to classify is to maintain situational awareness that could reduce subsequent classification times, either by remembering the location of black blips or determining their classifications in advance. Figure 11 shows the mean number of fixations on black blips as a function of day. Most blips were examined once while still black, suggesting that participants tried to maintain some awareness of blips prior to classification. In the peripheral-not-visible no-sound condition, the number of pre-classification glances increased across the three days, perhaps as participants converged on an optimal strategy.

Figure 11 shows evidence of audio cues affecting task strategies. Participants made fewer pre-classification glances when auditory cues were available ($\chi^2(1) = 14.7$, $p < .001$), evidently relying more on the audio than the visual to determine when blips needed classification. This holds true even for the two peripherally-visible conditions, providing the first graph in which the solid circles and solid
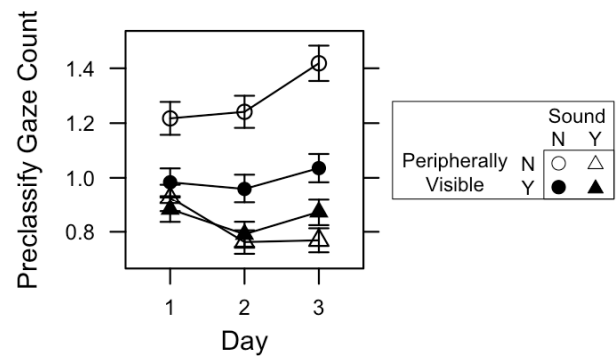

Figure 11. The mean number of fixations on black blips, for each of the four conditions, as a function of day.

triangles are not overlapped—the first evidence that participants used sound in the peripherally-visible conditions.

It is encouraging, given the goal of predicting human performance in time-critical dual task situations, to see that the timing of many of the eye movements, such as the eyes-to-target times in Figure 9, could likely be predicted based on straightforward task analysis and an understanding of fundamental human perceptual-motor constraints (such as built into [14]).

Other timings are more puzzling, such as the time required to get the eyes to the radar display after a blip changed color, roughly 700 ms if the color change was visible in the periphery or 1000 ms if announced by an auditory cue. The eyes can move much more quickly, on the order of 250 ms, in response to such events. A small part of this delay likely comes from participants taking some time to improve their tracking accuracy before moving to the radar: Eye-to-radar movement times correlate to tracking error, but with the very small slope of 2.3 ms per pixel of error, $t(5312) = 10.5$, $p < .001$. The timing of some eye movements in this dual task are based on not-yet-determined strategic decisions and information processing constraints.

Knowing Where to Look

This section discusses how people determine *where* to move their eyes to accomplish two tasks in parallel. In this task, there was a clear performance benefit if, after a blip changed color, the eyes could move directly to that blip with a single movement. There are perhaps three ways for a participant to know the location of a blip that just changed color: peripheral vision, knowledge from earlier glances at the radar, and information in the auditory cues. This section presents evidence that participants used all three.

Peripherally Visible Color Change

Participants clearly used peripheral vision to move the gaze directly to targets. Figure 12 shows the likelihood that an eye movement to the radar would land on (within 2° of visual angle of) the blip that just changed color. Participants moved directly to the target blip more often in peripherally-visible conditions than in peripherally-not-visible conditions, $z = -21.9$, $p < .001$ (analyzed using a generalized linear mixed model with a binomial response distribution).
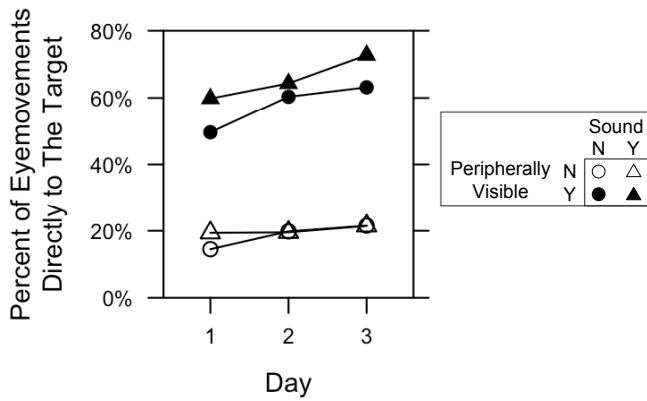
Figure 12. The percentage of eye movements to the radar display in which the gaze lands on the blip that just changed color.
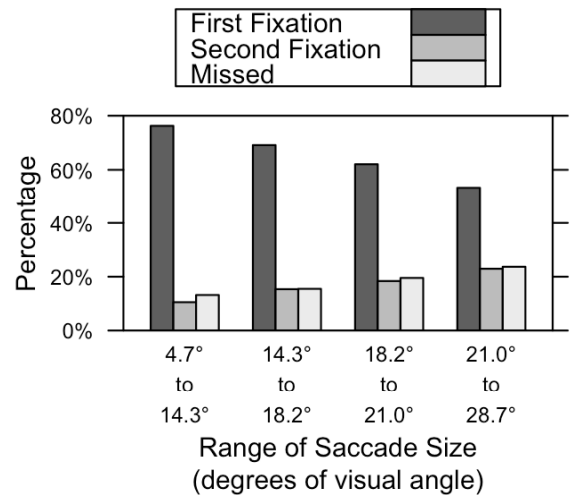

Figure 13. The percentage of eye movements to the radar display that land on the target in one fixation, in two fixations, or that miss the target in that visit to the radar; as a function of the distance from the gaze position to the target location. Only data from peripherally-visible conditions are included.

Participants were more likely to move their eyes directly to a peripherally-visible ready-to-classify blip if it was closer to the current eye position. Figure 13 accounts for all eye movements to the radar screen when peripheral visibility was available, and shows the likelihood that the target was fixated with a single eye movement. The data in Figure 13 are separated into four ranges of distances so that every range includes an equal number of eye movements. As the blips move from the nearest to the furthest range, the chance of moving the eyes directly to the target decreases from roughly 75% to 50%. It is difficult to know how much of the 25% decrease results from long saccades undershooting and requiring corrective saccades [1] and how much from vision degrading for objects further from the point of gaze [3]. A 10% increase in *second*-fixation acquisitions from the nearest to the furthest range, also shown in Figure 13, suggests that it is a combination of the two. (A 25% increase would suggest it was all from undershooting.) The data clearly indicate that participants used peripherally-visible color changes to move the eyes directly to the target.

Memory of Blip Location
There is some evidence that participants maintained situational awareness, or memory, of blip locations. As can be seen in Figure 12, the percentage of eye movements directly to the target for peripherally-not-visible conditions is low, around 20%. But we believe this to be above chance, which we estimate to be around 10% based on the dimensions of the display and the distribution of blip locations. This suggests that memory of blip locations was sometimes used to move the eyes directly to the target even when it was not visible in the periphery.

Auditory Cues
Auditory cues helped to get the eyes to the target. Evidence was presented in Figure 7, which showed that classification times improved in the peripheral-not-visible conditions when auditory cues were present, especially on Day 3. Figure 12 illustrates that the auditory cues did not, however, help the peripheral-not-visible conditions by getting the eyes from tracking to target in a single movement; this is illustrated in the overlapping unfilled plot symbols. The vertical distance between the filled plot symbols illustrate

how, in the peripherally-visible conditions, sound did help, $z = 4.92$, $p < .001$. We suspect that the auditory cue may have slightly reinforced the peripherally visible change, helping to draw attention to the object. It might also be that the color information encoded in the sound (high pitch for red, medium pitch for yellow, low pitch for green) helped to reinforce which blip was the target.

Auditory cues were spatialized so that the sound appeared in virtual space at a location that indicated the blip's location on the radar display. Our analysis of the data has yet to reveal any clear benefit from the spatialization. This is consistent with the marginal benefits of spatialized audio from a similar task [6].

Eye movement data demonstrate that participants used peripheral vision, situational awareness, and auditory cues to determine when and where to move their eyes in a time-critical dual task. Surprising results here include that the audio cues helped the participants to move the eyes directly to the target even when it was peripherally visible. But this was not because the sounds appeared at the exact same physical location as the blips, because the locations were transformed to improve resolution, as in [6]. Somehow, the audio cues helped to get the eyes directly to the target, though this does not seem to help overall classification time, the performance measure that really matters. It is also always interesting, though rarely surprising, to find situations in which people can reliably move their eyes directly to a needed piece of information with a single eye movement because the exact location of that information is clearly visible in an uncluttered periphery.

Overlapped Performance
Consistent with theories and approaches to predictive modeling [8, 13, 14], participants exhibit evidence of perceptual and motor processing that is overlapped and interleaved to maintain very good parallel performance on

two demanding tasks. The high-speed coordination of eye movements and perceptual processing between the two visual displays, both to classify blips but also to maintain situational awareness, demonstrates a complex interleaving of processing between the two tasks. Figure 14 reproduces Figure 9 but now adds the observed keypress time to the movements included in the graph; the dashed lines indicate how much additional time is required to key-in a blip more than a second after the eyes have moved back to the tracking task. This can be seen in the vertical distance between the plot symbols for the keypress and, directly below, for eyes-to-tracking; this vertical distance represents an overlap between various processes required for the classification and tracking tasks.

Participants clearly move their eyes off of target blips as quickly as possible, staying only long enough to gather the needed visual features, and then process these features to produce a response choice after moving the eyes back to tracking.

Four participants, three of whom were the best-performing participants, demonstrated a somewhat complex task strategy that permitted them to classify blips without looking at those blips after they changed color. We call these *blind classifications*. Four participants each exhibited an average of eighty-six such classifications (12.6% of all classifications). Blind classifications can only occur if a participant studies a blip when it is black and, at the very least, commits that blip's number to memory. These classifications represent a parallel execution of the classification task without even moving the eyes from the tracking task.
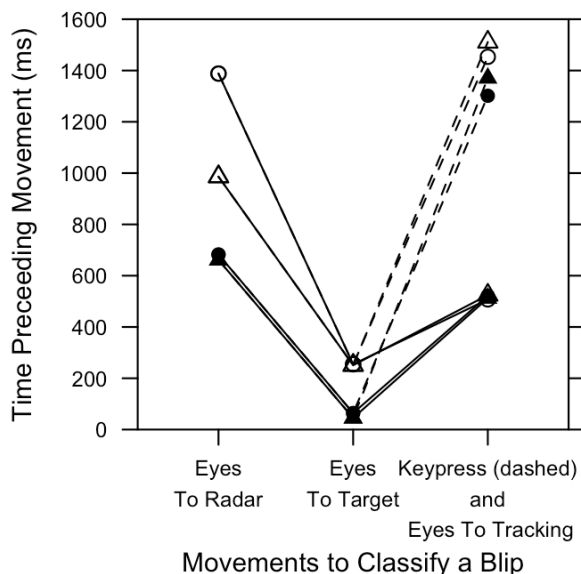

Figure 14. Times preceding eye movements and keystrokes (dashed lines) to classify a colored blip.

CONCLUSION

This paper presents a time-critical multimodal dual task experiment that successfully demonstrates and measures how (a) auditory cues can complement visual information to improve performance especially when visual information is not peripherally visible, (b) subtasks can pile up and interfere with each other unless individual subtasks can be accomplished within intermediary performance deadlines, (c) making secondary task information visible in the near periphery of the primary task display can improve dual task performance, and (d) carefully controlled performance feedback and incentives can guide the development of effective cognitive task strategies that integrates across tasks and modalities. Though these findings are not necessarily new, it is perhaps unusual to see them all illustrated so vividly in a single study as is done here.

Human-computer systems intended for high performance multitasking should be designed to support the streamlined, overlapped processing illustrated in this paper. Rigorous evaluation such as described here can be conducted to determine whether optimal dual task strategies and performance are likely to be achieved. Designers need to understand what people are doing in these complex, time-critical tasks, and it is perhaps only through studies such as theses that an accurate and detailed understanding can be achieved.

Design Guidelines

Three specific observations and related design guidelines for time-critical multimodal multitasking systems flow from this study:

Observation. A single subdued auditory alert (in our case, a 110 ms wood block sound whenever a blip appeared) seemed to help a lot to keep participants aware of what was happening in the secondary display even when it was not visible in the periphery. It helped them to know, for example, when a large wave of blips was starting to appear and hence more eye movements to the secondary display would soon be needed.

Guideline. Use auditory alerts to deliver information for secondary tasks when the primary task has high visual demand. This will create the strategic opportunity for interleaving perceptual modalities across multiple tasks. Even subtle auditory alerts as to the status of a secondary display can go a long way towards optimal interleaved performance.

Observation. Our device supported a non-visually-assisted manual response for the secondary task. After participants acquired a blip's information on the secondary display, they moved their gaze back to the primary display before keying in the blip classification.

Guideline. For high-performance interfaces to support frequent task switching and highly streamlined multitasking, design components to display all task information in parallel, and to permit simultaneous parallel input for all critical tasks. For example, do not require users to select a window with a mouse to make a response. Task time can be reduced by permitting users to look at one display while making responses for the other.

Observation. Different performance measures evolved along different trends across the three days. Some demonstrate unsurprising practice and learning effects, but others suggest evolving task strategies, such as new decisions on how to use perceptual information. Perhaps most notable are adjustments to the use of auditory cues from Day 2 to Day 3.

Guideline. For complex, critical tasks in which any additional information may help—or hurt—performance, a prolonged evaluation is required to permit people to develop complex interleaved tasks and exhibit the improved performance made possible by such strategies.

The experiment presented here demonstrates how a first-pass analysis of the data might support a simple conclusion, such as that auditory cues did not affect performance when peripheral information was available. Deeper probing reveals that this is not entirely correct. While there is no clear evidence that participants used the spatialized component of the auditory cues, the cues did permit resourceful participants to develop strategies in which they (a) sometimes use auditory cues in place of eye movements to maintain situational awareness, (b) constrain search once the eyes arrived on the secondary task display, and (c) use the multimodal interface to sometimes accomplish secondary subtasks without even looking at the secondary display, as with the blind classifications.

REFERENCES
[1] Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance,* 15(3), 529-543.

[2] Ballas, J. A., Heitmeyer, C. L., & Perez, M. A. (1992). Evaluating two aspects of direct manipulation in advanced cockpits. *Proceedings of ACM CHI '92: Conference on Human Factors in Computing Systems,* 127-134.

[3] Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of effective stimulus in visual search. *Perception & Psychophysics,* 62(3), 576-585.

[4] Cabrera, D., Ferguson, S., & Laing, G. (2005). Development of auditory alerts for air traffic control consoles. *Proceedings of the 119th Audio Engineering Society Convention,* New York, USA.

[5] Hornof, A. J., & Halverson, T. (2002). Cleaning up systematic error in eye tracking data by using required fixation locations. *Behavior Research Methods, Instruments, and Computers,* 34(4), 592-604.

[6] Hornof, A. J., Halverson, T., Issacson, 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,* 1170-1174

[7] Kieras, D. E., Ballas, J., & Meyer, D. E. (2001). *Computational Models for the Effects of Localized Sound Cuing in a Complex Dual Task (No. EPIC Report No. 13).* Ann Arbor, Michigan: University of Michigan, Department of Electrical Engineering and Computer Science.

[8] Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction,* 12(4), 391-438.

[9] Kieras, D., Meyer, D., & Ballas, J. (2001). Towards demystification of direct manipulation: Cognitive modeling charts the gulf of execution. *Proceedings of ACM CHI 2001: Conference on Human Factors in Computing Systems,* 128-135.

[10] Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance and Cognition.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

[11] R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

[12] Russo, J. E. (1978). Adaptation of cognitive processes to the eye movement system. In J. W. Senders, D. F. Fisher & R. A. Monty (Eds.), *Eye Movements and the Higher Psychological Functions* (pp. 89-109). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

[13] Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. *Proceedings of ACM CHI 2009: Conference on Human Factors in Computing Systems,* 1819-1828.

[14] Vera, A.H., Howes, A., McCurdy, M., Lewis, R.L., (2004) A constraint satisfaction approach to predicting skilled interactive cognition. *Proceedings of ACM CHI 2004: Conference on Human Factors in Computing Systems,* 121-128.