Task-Constrained Interleaving of Perceptual and Motor Processes in a Time-Critical Dual Task as Revealed Through Eye Tracking

Anthony J. Hornof (hornof@cs.uoregon.edu)
Yunfeng Zhang (zywind@cs.uoregon.edu)
Computer and Information Science, University of Oregon
Eugene, OR 97403 USA

Abstract

A multimodal dual task experiment that contributed to the original development and tuning of the EPIC cognitive architecture is revised and revisited with the collection of new high fidelity human performance data, most notably detailed eye movement data, that reveal the complex overlapping of perceptual and motor processes within and between the two competing tasks. The data permit a new detailed evaluation of assumptions made in previous models of the task, and contribute to the development of new models that explore opportunities for overlapping visual-perceptual, auditory-perceptual, ocular-motor, and manual-motor activities. Three models are presented: (a) A hierarchical task-switching model in which each task locks out the other; the model explains reaction time but does not account for eye movement data. (b) A maximum-perceptual-overlap model that maximizes parallel processing and predicts the trends in the eye movement data, but performs too quickly. (c) A moderately-overlapped model that introduces task-motivated constraints and predicts both reaction time and eye movement data. The best-fitting model demonstrates the complex task-constrained interleaving of perceptual and motor processes in a time-pressured dual task.

Keywords: Cognitive strategies, EPIC cognitive architecture, eye tracking, multimodal dual task, multitasking.

Introduction

A critical task domain for the research enterprise of cognitive modeling is that of multimodal (auditory and visual) multitasking. Psychologists and cognitive modelers puzzle over the question of how people engage in two or more time-pressured tasks that compete for perceptual, cognitive, and motor processes, such as for air-traffic control or in-car navigation (Byrne & Anderson, 2001; Howes, Lewis, & Vera, 2009; Meyer & Kieras, 1997; Salvucci & Taatgen, 2008). Gaining an understanding and ability to predict aspects of multimodal multitasking is of critical scientific and practical importance. This paper advances an understanding of such tasks by presenting cognitive models of time-critical multimodal multitasking and evaluates these models in detail using eye tracking data.

The Time-Critical Multimodal Dual Task

An earlier version of the experiment that forms the basis of this theoretical exploration was conducted in the early 1990s at the Naval Research Laboratory (NRL) (Ballas, Heitmeyer, & Perez, 1992). The experiment produced human speed and accuracy data that proved useful for developing detailed computational cognitive models of dual task performance (Kieras, Ballas, & Meyer, 2001). 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 that (a) eye movements are recorded, (b) eye tracking is used in some conditions to hide objects on the not-currently-looked-at display, (c) auditory cues relate more directly to required responses, and (d) participants are rigorously trained, financially motivated, and given extensive feedback so that performance approaches that of an expert.

Figure 1 shows an overview of the two displays used in the multimodal dual task modeled in this paper. Two tasks (or subtasks) were performed in parallel: a tracking task and a tactical classification task. The tracking task consisted of keeping a small circle on a moving target using a joystick. When the circle was positioned as such, it turned green, and the participant was financially rewarded at a constant rate. The tactical classification task consisted of monitoring groups of icons or "blips" (fifty-seven in a nine-minute scenario) that moved down a radar display, and keying-in the blip number and "hostile" or "neutral" as soon as the blip changed from black to red, green, or yellow, indicating that it was "ready to classify". When a blip became ready to classify, a financial bonus was awarded though it diminished at a constant rate until the blip was keyed-in, or classified. Red blips were hostile; green were neutral; yellow blips were classified based on their shape, speed, and direction, following practiced rules.

Two important factors were manipulated in the experiment: (a) peripheral visibility on or off and
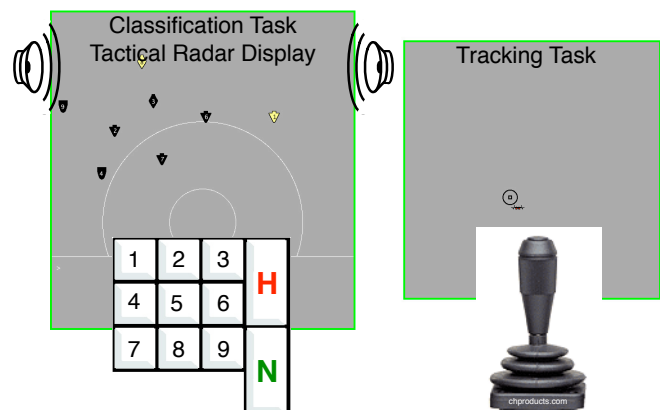

Figure 1: An overview of the visual and auditory displays and input devices used in the multimodal dual task.

(b) auditory cues present or absent. *Peripheral Visibility* manipulated whether participants could see the contents of the other display—radar or tracking—that they were not currently looking at. This simulates a task environment in which visual displays are separated by enough distance such that they cannot be monitored with peripheral vision. *Auditory Cues (Sound On)* indicates that a blip's initial appearance (as black) and color change (to red, green, or yellow) were indicated with spatialized auditory cues. Each nine-minute scenario maintained a constant setting of peripheral visible or not-visible and sound on or off.

Figure 2 summarizes the most important eye and hand movement data from the experiment, which is described in more detail in Hornof, Zhang, Halverson (2010). Figure 2 shows the time required for the four consecutive stages of classifying a blip: (a) Initiate the eye movement from the tracking display to the tactical display; (b) once on the tactical, find the target and move the eyes to it; (c) keep the eyes on the blip long enough to identify it and then move the eyes back to tracking; and (d) after the eyes are back on tracking, key-in the blip (keying-in was consistently performed *after* the eyes were back on tracking). These data serve to reveal the complex interleaving of perceptual, cognitive, and motor processing, and provide a basis for the current modeling endeavor.
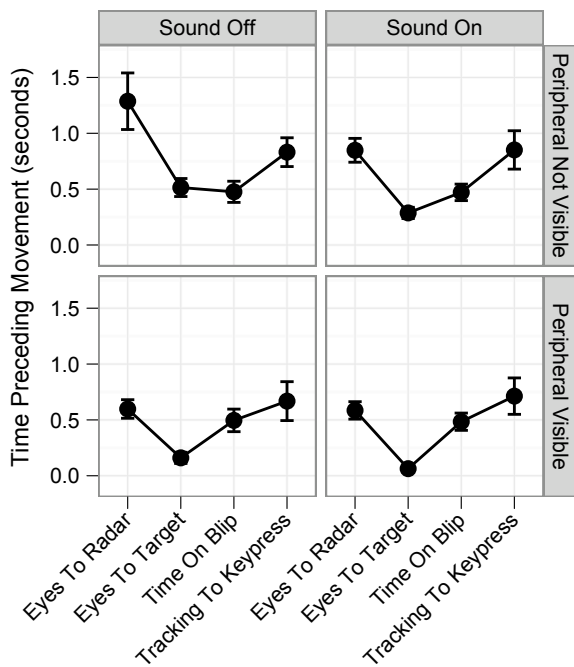

Figure 2. Time preceding eye movements across the lifetime of a colored blip. Each panel shows a unique combination of the factors of peripheral visibility and sound on/off. The *x*-axis shows a sort of timeline of the stages involved in classifying a blip.

The EPIC Cognitive Architecture

The EPIC cognitive architecture (Executive Process-Interactive Control; Kieras & Meyer, 1997) was used to model the multimodal dual task, as it was used previously to model the earlier version of the same task (ibid.; Kieras, Ballas, & Meyer, 2001). EPIC is particularly well-suited for exploring a range of explanations of multitasking performance because of its specific commitment, at the architectural level, to only enforcing sequential processing for motor activities, such as to constrain the eyes to rotate to only one point at a time, and the hands to only execute one sequence of movements at a time. Perceptual information can flow into the auditory and visual processors in parallel, and multiple production rules—IF-THEN statements that represent the strategy used to do a task—can fire in a single 50 ms cycle. Strategies can be written to permit only one rule to fire at a time (as in our initial model) or to explore the full potential of overlapping (as in our second model).

Extensions to the EPIC Cognitive Architecture

Initial sets of production rules that were constructed to put the eyes and hands through the tasks revealed two extensions to the EPIC cognitive architecture that would be needed to model this task: (a) a computational solution to the binding problem, which is the question of how people assemble perceptual stimuli to maintain a seamless conscious experience, and (b) a temporal processor to determine, entirely from within the simulated organism, when a certain amount of time has elapsed.

To address the binding problem, the visual processor in the EPIC cognitive architecture was updated (by EPIC's creator David Kieras) so that psychological objects in EPIC's visual working memory maintain their identity even as they disappear and reappear in the physical environment. In other words, if the simulated human moves its eyes so that a blip disappears (as in the peripheral-not-visible conditions), and then moves its eyes so that the same blip reappears, EPIC would previously have created a new psychological object for the reappeared blip. Now, provided that the initial psychological object associated with the blip did not fully decay, the reappeared blip is reconnected to the already-existing psychological object.

The second extension to EPIC was to add a temporal processor that replicates the temporal processor added to the ACT-R cognitive architecture (Taatgen, van Rijn, & Anderson, 2007). This gives the models a way to make self-motivated periodic checks of the tactical display when there was no peripheral visibility or auditory cuing.

Modeling Overview

Each of the models below were presented with the exact same auditory and visual stimuli in identical nine-minute scenarios that were presented to our human participants.

The following parameters were used in the models: The time required to determine the classification of a yellow blip based on its speed and direction was set to 800 ms. Alarm sounds are identified 300 ms after their onset in auditory perception rather than with their onset, to give enough time to distinguish the alarm from the blip appearance sound.

A common element within all strategies include that tracking adjustments (by moving the joystick with a Ply) were made only when the tracking circle was *not* green, consistent with a strategy that maximizes payoff.

The model development presented here follows the "bracketing" approach advocated by Kieras & Meyer (2000) in which the analyst attempts to "bracket" the human data with a slowest-reasonable and fastest-reasonable strategies. Three corresponding task strategies are developed: (a) Hierarchical task-switching (the slowest-reasonable model); (b) Maximum-perceptual-overlap (the fastest-reasonable model); and (c) Moderately-overlapped (the fastest-reasonable model slowed down based on task constraints). Models based on these three strategies, and comparisons of each model's predictions with the human data, are presented next.

Hierarchical Task-Switching Model

The hierarchical task-switching (the slowest-reasonable) model represents a straightforward translation of the multimodal dual task into a hierarchical task with strict serial processing of each subtask. Figure 3 shows the corresponding hierarchical task analysis. The production rules were generated by first creating a GOMS model (John & Kieras, 1996) of the task, and then translating that model into the corresponding production rules. Parallelism existed in the model primarily in terms of auditory and visual information getting deposited in EPIC's perceptual stores.

A key characteristic of the model includes that, once it determines that a blip is ready to classify, it holds the eyes on that blip until the keystrokes for that blip are initiated. During this period, the cognitive processor is dedicated to just classifying the blip. Tracking is completely locked out. This aspect of the model resembles the original EPIC models of the task, in which "the dual-task executive enforces mutual exclusion between the tracking task and the tactical task." (Kieras, Ballas, & Meyer, 2001, p.10)

Figure 4 shows the mean blip classification times across the four combinations of peripheral-visibility and sound-on-or-off, and for red/green versus yellow blips. The model explains the overall reaction time data very well across all eight conditions, with an average absolute error (AAE) of 4.6%. (Note that all AAEs presented in this paper are calculated using the overall observed mean as the denominator for each percentage calculation, to reduce the distortion that would otherwise result from observed and predicted values that are very close to zero.)

If an analyst were primarily interested in the classification task and hence did not proceed to model the tracking task with any degree of fidelity, and if the analyst did not have any eye movement data to work with, the modeling project would likely be done at this point, and we might declare victory—we modeled the primary data of interest with good
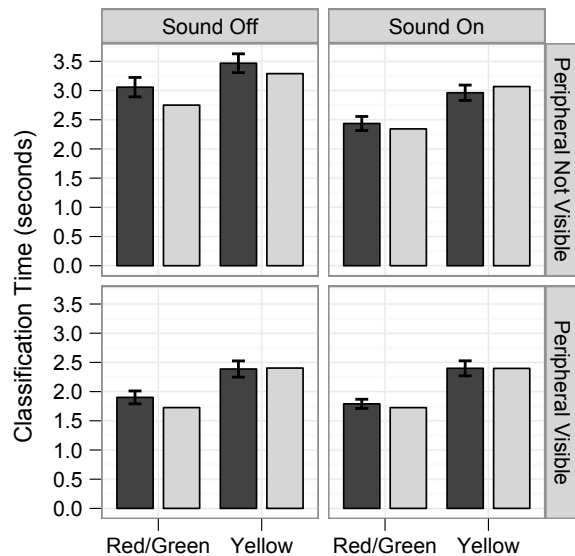

Figure 4: The mean classification time of blips as a function of blip color, observed (dark bars) and predicted (light bars) by the hierarchical task-switching model. The average absolute error (AAE) of the prediction is 4.6%.

accuracy. But a deeper look at the data that are available in this modeling exercise reveal a dark truth—the model is not accounting for the complex overlapping of visual and motor processes that participants are exhibiting with their eye movements. As well, a look at the tracking task data show that the model is performing far worse than skilled participants, predicting an overall mean tracking error of 42 pixels compared to the observed tracking error of 29 pixels.

Figure 5 shows the same observed data presented in Figure 2, along with the eye movement times predicted by the hierarchical task-switching model. As can be seen in Figure 5, the model is spending far too long looking at each blip. The tracking-to-keypress is negative (and hence a value of zero is used) because the model returns the eyes to tracking *after* the classification. Participants spent far less time on each blip, and spent substantial time with the eyes back on tracking before keying-in a classification.

The hierarchical task-switching model, though intended as a slowest-reasonable bracket, does a good job of predicting the mean classification times. But the model does not capture the interleaving of perceptual and motor processes that people clearly exhibited. The next model attempts to capture and maximize such an interleaving.
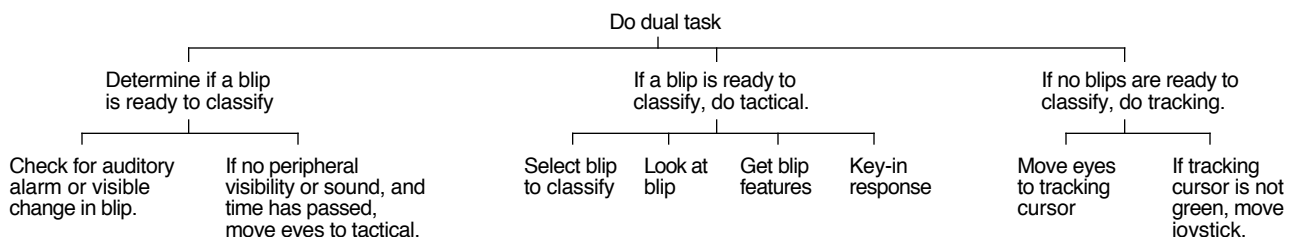

Figure 3: The hierarchical task analysis used to generate the hierarchical task-switching model.
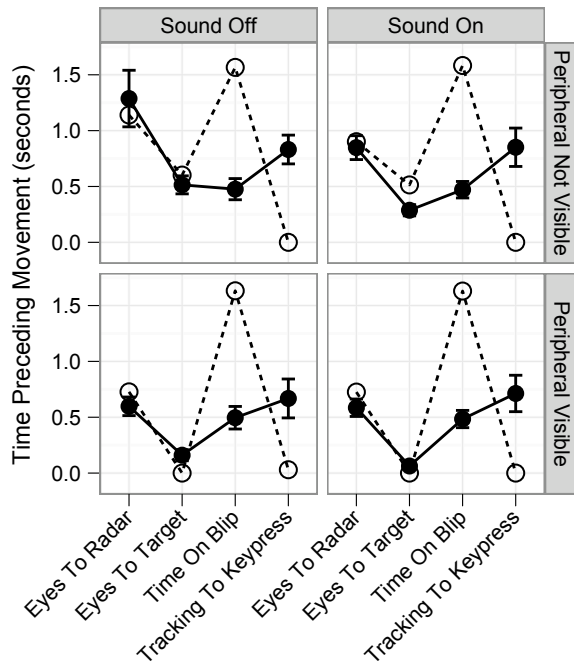
Figure 5: The time preceding eye movements observed (solid lines) and predicted (dashed lines) by the hierarchical task-switching model. (AAE = 91.4%)

Maximum-Perceptual-Overlap Model

The maximum-perceptual-overlap (fastest-reasonable) model is written to maximize all aspects of parallel processing that are built into the EPIC cognitive architecture. The production rules are written such that ocular-motor and manual-motor processing proceed entirely independently of each other, with manual-motor processing resulting from visual-perceptual features that become available based on ocular-motor activity.

Figure 6 shows two state transition diagrams that represent how one set of production rules moves the eyes between tracking and tactical to acquire visual information, and another set of rules independently shifts manual motor activity between tracking and tactical. When the model runs, both sets of rules—ocular-motor and manual-motor—spend most of their time on tracking. When a blip appears, the ocular-motor rules shift to tactical just long enough to perceive blip features, which become available to the manual-motor rules, which switch briefly to tactical to key-in a response. Each set of rules returns to tracking as soon as its tactical subtask is completed.

Figure 7 shows the classification time predictions of the maximum-perceptual-overlap model. As can be seen, the model is too fast, as would be expected for a fastest-reasonable model. Looking at the predicted eye movement times in Figure 8, however, reveals that the model does a good job predicting the overall trends in how long the eyes took to move through the stages involved in classifying a blip, especially in the peripheral-visible conditions. The comparably good fit of the eye movement data, especially when compared to the first model's poor fit with the same data, suggest that participants may truly have developed
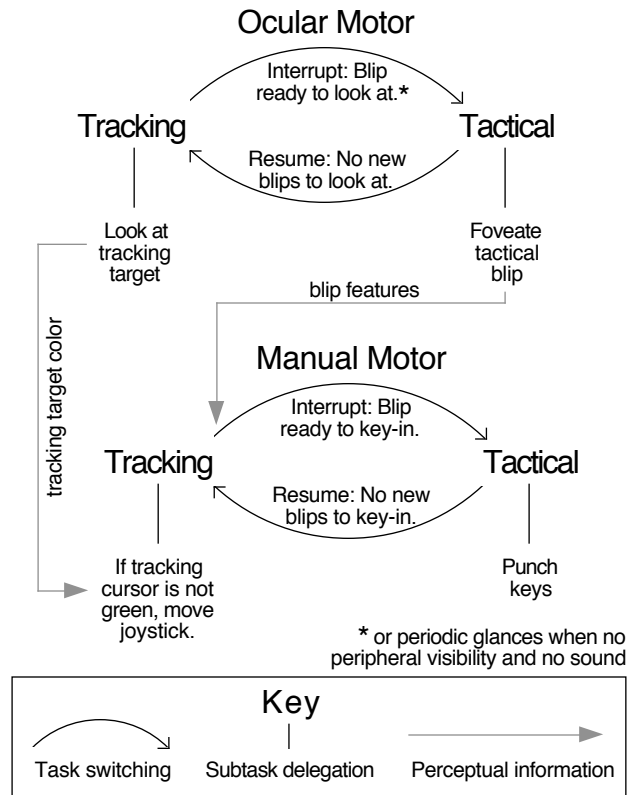

Figure 6: State transition diagrams that represent the independent ocular-motor and manual-motor processing in the maximum-perceptual-overlap model.
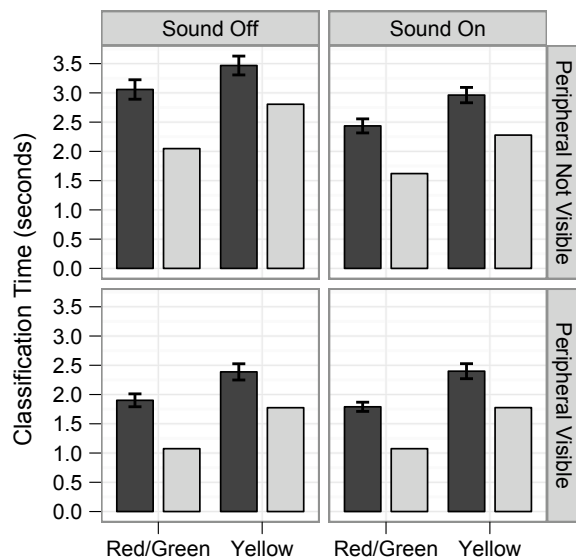

Figure 7: Classification times observed (dark bars) and predicted (light bars) by the maximum-perceptual-overlap model. (AAE = 29.2%)

expert strategies that include independent parallelism between ocular-motor and manual-motor decision making. But, as might be expected, the fastest-reasonable model is overall too fast. The predicted mean tracking error is also substantially better (20 pixels) than the observed (29 pixels).
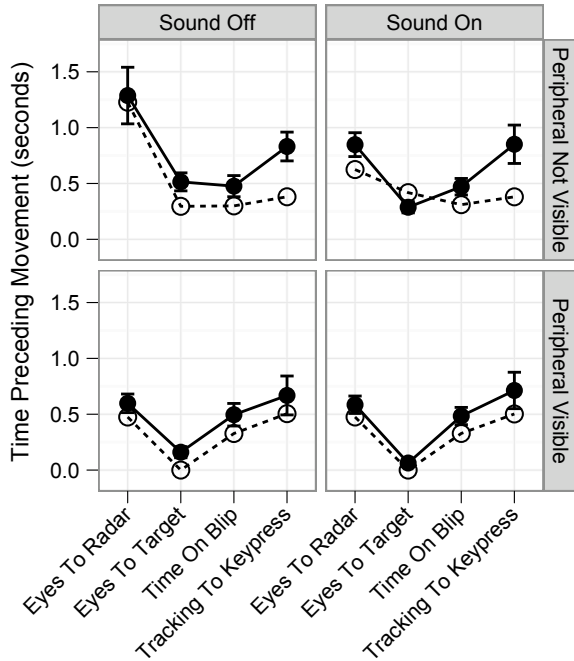
Figure 8: The time preceding eye movements observed (solid lines) and predicted (dashed lines) by the maximum-perceptual-overlap model. (AAE = 32.6%)

The final strategy explores constraints that can be introduced to the fastest-reasonable model.

Moderately-Overlapped Model

The moderately-overlapped model was constructed by starting with the maximum-perceptual-overlap (fastest-reasonable) model, presented in the previous section. Three analyses were conducted. First, the model traces and observed data were studied side-by-side to reveal subtle differences between the predicted and observed eye and hand movements. Second, opportunities were explored to adjust strategies to maximize payout (see Howes et al., 2009). Third, the manual-motor devices were examined to improve the fidelity of their simulation.

These analyses led to the following five adjustments to the model, all of which are represented by the bold italic additions in Figure 9: (a) Eye-to-radar time is delayed by having the tracking task finish any joystick Ply underway, waiting for the tracking circle to turn green, to leave that task in a money-making mode. (b) The time on yellow blips is extended to permit identification of speed and direction (set to 250 ms). (c) Tracking-to-keypress time is extended by assuming that, when moving the eyes from tactical back to tracking, people make one joystick adjustment before keying-in the blip classification; this increases tracking payment while further considering the classification. (d) The timing for a Ply was increased (to a coefficient of 300 and minimum time of 400 ms) assuming that the Ply effectively requires separate joystick movements to start and then stop the tracking circle. (e) The Punch was replaced with a Keypress to represent how the fingers are not positioned directly above the keys, but need to travel.
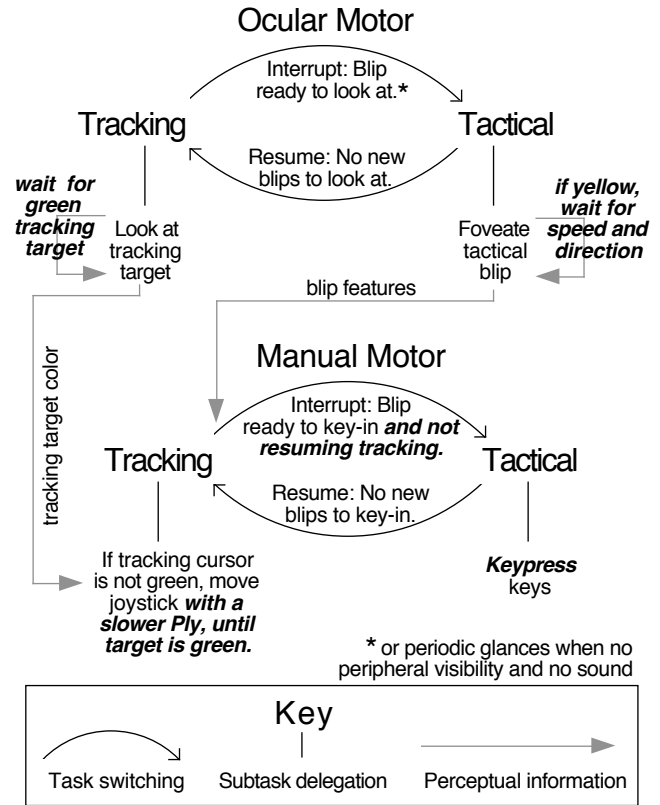

Figure 9: The moderately-overlapped model, with additions to the previous model shown in bold italics.
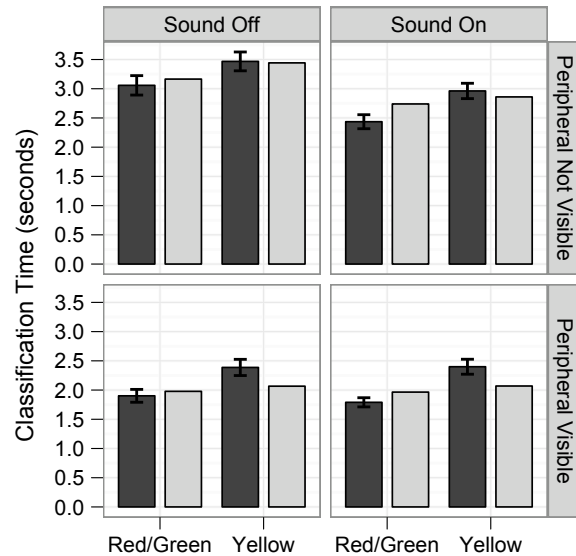

Figure 10: Times observed (dark bars) and predicted (light bars) by the moderately-overlapped model. (AAE = 7.1%)

Figures 10 and 11 show how the moderately-overlapped model does a good job of predicting both classification and eye-movement timings. The model also accurately predicts tracking error, predicting 26 pixels compared to the observed 29 pixels. Table 1 shows how this model provides the best overall fit with the observed data.
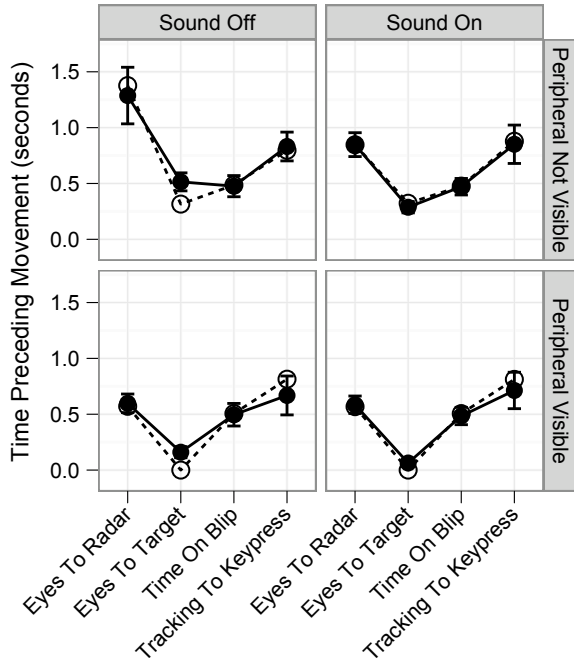
Figure 11: The time preceding movements observed (solid lines) and predicted (dashed lines) by the moderately-overlapped model. (AAE = 10.1%)

Table 1. Average absolute error of each model's predictions.

Model	Classification Time	Time Preceding Movements	Tracking Error
Hierarchical Task-Switching	4.6%	91.4%	43.6%
Maximum-Overlap	29.2%	32.6%	31.2%
Moderately-Overlapped	7.1%	10.1%	13.9%

Conclusion

The models presented here demonstrate the difficulty in accurately modeling complex multitasking behavior. First, there is the challenge of collecting enough data to evaluate the accuracy of a model; the initial hierarchical task-switching model accurately predicted the classification time, but not eye movements. Then, there is the challenge of correctly identifying opportunities for expert, overlapped behavior; the maximum-perceptual-overlap model presented here relied on the massive parallelism of the EPIC architecture's cognitive processor to demonstrate that expert strategies might manage ocular-motor and manual-motor processes largely independently. Lastly, there is the challenge of determining which task-based constraints should be introduced to govern the use of perceptual information that passes within and between two tasks that compete for motor processing; those presented for the moderately-overlapped model may or may not accurately capture the true constraints that governed behavior.

The models presented here do not clearly subscribe to the notion of an independent process that actively coordinates between two task strategies, whether that process be an executive process, as in the original models for a similar task (Kieras, Ballas, & Meyer, 2001) or an independent mechanism, as in Salvucci and Taatgen (2008). This paper explores the possibility that a dual task strategy is perhaps an altogether new, carefully interleaved strategy.

Acknowledgments

References

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.

Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review, 108,* 847-869.

Hornof, A. J., Zhang, Y., Halverson, T. (2010). Knowing where and when to look in a time-critical multimodal dual task. *Proceedings of ACM CHI 2010: Conference on Human Factors in Computing Systems,* New York: ACM, 2103-2112.

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review, 116*(4), 717-751.

John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction, 3*(4), 287-319.

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.

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.

Kieras, D. E., & Meyer, D. E. (2000). The role of cognitive task analysis in the application of predictive models of human performance. In J. M. C. Schraagen, S. E. Chipman & V. L. Shalin (Eds.), *Cognitive Task Analysis* (pp. 237-260). Mahwah, NJ: Lawrence Erlbaum.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*(1), 3-65.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review 115,* 101-130.

Taatgen, N. A., van Rijn, H., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychological Review, 114*(3), 577-598.