Cognitive Models of the Effect of Audio Cueing on Attentional Shifts in a Complex Multimodal, Dual-Display Dual Task

Derek Brock (brock@itd.nrl.navy.mil) and Brian McClimens (mcclimen@itd.nrl.navy.mil)
Naval Research Laboratory, 4555 Overlook Ave., S.W.
Washington, DC 20375 USA

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

Abstract

A comparative cognitive model of two manipulations of a complex dual task in which 3D audio cueing was used to improve operator performance is presented. The model is implemented within the EPIC cognitive architecture and describes extensions that were necessary to simulate gaze shifts and the allocation of attention between separated task displays. A simulation of meta-cognitive decision-making to explain unprompted, volitional shifts of attention and the effect of audio cueing on performance and the frequency of attention shifts are explored.

Keywords: cognitive modeling; EPIC; dual task; 3D auditory cueing; separated task displays; gaze shifts; volitional shifts of attention; simulated meta-cognitive decision-making; gamma distribution; sense of timing

Introduction and Background

System designers take numerous approaches to reduce the number of workstation operators necessary to accomplish complex decision-making tasks in Navy command-and-control centers. Approaches include (a) the automation of multiple tasks and (b) the adoption of supervisory rather than direct control. As workstation operators are asked to manage an increasing number of tasks, reliable techniques are needed to manage operator attention. The research presented here demonstrates how cognitive modeling can explain how operators manage conflicting attention demands and also how cognitive modeling can inform the design of human-machine interfaces that facilitate efficient and accurate multi-display, multi-task execution.

The Navy has developed a prototype decision support workstation that features three flat-panel monitors centered in a 135° arc in front of the user (Osga, 2000). With this configuration, the operator can access much data but loses peripheral access to all three monitors when his or her gaze is turned to look at either the right or left screen. This loss can reduce the speed and accuracy of critical decisions (Brock et al., 2002; Brock et al., 2004).

The Naval Research Laboratory (NRL) is developing techniques for directing attention in complex operational settings using three-dimensional (3D) or "spatialized" sound (Begault, 1994). Properly designed 3D sounds can be used to convey a variety of task-related information, including the onset, location, and identity of critical events.

Brock et al. (2004) demonstrated that the use of 3D sound can significantly improve dual-task performance. The research presented here describes recent cognitive modeling work that has been done to explain the effects of audio cueing observed by Brock et al., and the effect of audio cueing on the allocation of attention between tasks. The models are based on the human data observed by Brock et al. in (a) the "no sound" condition and (b) one particular sound condition (the "screen-centric" condition).

Cognitive modeling is a research practice that endeavors to build computer programs that behave in some way like human beings. The models presented here are implemented within the EPIC (Executive Process-Interactive Control) cognitive architecture (Kieras and Meyer, 1997), which is a computational framework for building models of human performance based on the constraints of human perceptual, cognitive, and motor processing.

The cognitive modeling presented in this paper specifically explores (a) a simulation of meta-cognitive decision-making to explain the volitional shifts of attention, (b) performance aspects of task-related audio cueing, a somewhat new domain for cognitive modeling, and (c) extensions to the perceptual-motor components of EPIC that were necessary to simulate a complex dual-display task.

The Attention Management Study

The Dual Task

Figure 1 shows the physical layout of the dual task modeled in this paper. The task is from Brock et al. (2004). Participants used a joystick to continuously track an evasive target on the right and, at the same time, used the keyboard to periodically assess and classify "blips" moving down the radar screen on the left. The right task is "tracking" and the left task is "tactical." The task displays were separated by 90° of arc, such that the unattended display could not be seen with peripheral vision. The task was originally developed by Ballas, Heitmeyer & Perez (1992) and is analogous in many ways to the level of multitask activity that future Navy workstation operators will be subjected to.

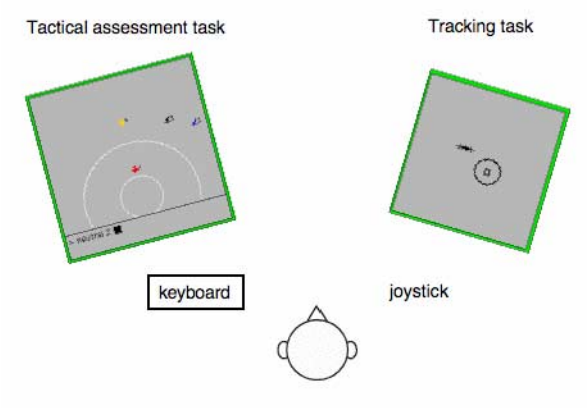

Figure 1: The physical layout of the dual task.

Whereas the tracking demands continuous attention, the tactical task can be accomplished with brief intermittent glances and bursts of activity. The procedure for tactical assessment and classification is somewhat complex. The tactical radar deals in three shapes of blips. All initially appear as black, numbered icons, and move at varying slow speeds, and in different patterns, from the top to the bottom of the screen. A few seconds after each blip first appears, its color is changed from black to red or blue or yellow. This color-coding lasts for about ten seconds, and during this time, a two-keystroke combination must be entered—the blip's number and its classification. Red and blue blips are classified as, respectively, hostile and neutral. Yellow blips must be classified based on their onscreen behavior and rules learned in advance. After assessment (or ten seconds), the blip disappears.

Auditory Cueing In some conditions (from Brock et al., 2004), the tasks were augmented with 3D auditory cues. This paper discusses models of the baseline "no sound" condition and one of the "sound" conditions (the screen-centric condition). A unique sound loop was played for each of the three different blip shapes on the tactical radar screen. Audio cues were sounded when blips were color-coded. Only one audio cue was played at a time. Thus, if a new blip became color-coded while another blip's auditory cue was already playing, the new one had to wait until the previous blip was classified and/or disappeared. The apparent 3D source of the sounds was located forward and 45° to the left of the orientation of participant's head. Headphones were used, but not head tracking.

Empirical Measures of Performance The performance measures discussed here (originally presented in Brock et al., 2004) are derived from (a) tactical-response timing data logged by the dual-task software, and (b) counts of participant gaze shifts between the left and right displays and the keyboard, logged by an experimenter on a Palm Pilot. The actual observed shifts were head turns, which are assumed to correspond to gaze shifts and attentional shifts based on the large visual angle among the devices.

Figure 2 shows the mean number of gaze shifts and tactical assessment response times in the sound and no-sound conditions. Response time is the total time to classify a blip (with two keystrokes) after it changes color.

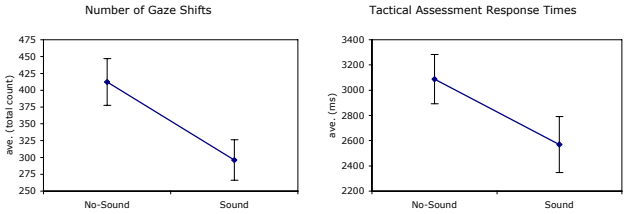

Figure 2: The total number of gaze shifts (based on head turns) and the response times observed in the no-sound (baseline) and sound conditions. (Note the nonzero *y*-axis.)

As seen in Figure 2, audio cueing reduced the number of head turns that were needed, and at the same time improved the tactical assessment response times (with no difference in response accuracy). Both differences are significant ($p < .001$). These results demonstrate that screen-centric audio cues can improve complex dual-task performance.

Table 1 shows counts of the attentional shifts made by participants, and provides a more detailed picture of how participants allocated their attention in the no-sound and sound conditions. The no-sound condition, for example, required 44% more shifts from the right (tracking) display to left (tactical) display (174 compared to 121). The data also indicate that, even in the sound condition, participants did not rely solely on audio cues to monitor the status of blips on the tactical radar screen. Even though only 65 blips were presented in each manipulation, participants in the sound condition averaged 121 looks from right to left; this means that participants made, on average, 56 additional self-motivated shifts to the tactical task. These 56 inspections are perhaps analogous to the 174 self-motivated right-to-left shifts in the no-sound condition. In each condition, this is an estimate of the total number of meta-cognitively-prompted volitional shifts from right to left.

Table 1: Counts of attentional shifts (based on head turns) observed in the no-sound and sound conditions. Location key: right = tracking and left = tactical.

No-Sound - Mean Count of Attentional Shifts

Shift from	Right	Left	Keybd	
Right to	0	174	7	180
Left to	170	0	27	197
Keybd to	10	23	0	34
	180	197	34	411

Sound - Mean Count of Attentional Shifts

Shift from	Right	Left	Keybd	
Right to	0	121	6	127
Left to	116	0	23	139
Keybd to	11	19	0	29
	127	139	29	295

The role and frequency of these meta-cognitive decisions to shift from tracking to tactical assessment is explored in the model through parametric manipulations of a gamma distribution. The distribution is used to represent the time that elapses before the participant motivates an internal decision to switch back to the tactical task in the absence of any external motivator to do so.

Modeling Dual-Task Performance

The Modeling framework

The EPIC cognitive architecture (Executive Process-Interactive Control; Kieras and Meyer 1997) was used in this modeling effort. EPIC is a unified theory of human perceptual, cognitive, and motor processing that provides a computational framework for modeling human information processing in simulated task environments. Based on fundamental human processing capabilities and limitations, the architecture is designed to accurately capture the details of human performance in predictive cognitive models.

Cognitive processing in EPIC is implemented as a production rule system that allows multiple rules to fire in parallel, and whose working memory maintains a declarative representation of the world. Perceptual-motor processing is implemented as a set of simulated sensory and motor processing peripherals, running in parallel with cognition, whose organization, performance, and constraints are derived from the human performance literature.

Models in EPIC are composed of three principal components: a reproduction of the interactive task environment; a set of perceptual features associated with the stimuli in the task environment; and a task performance strategy, implemented as a set of production rules. Models are run in the context of a task simulation, with the task strategy guiding various motor activities as well as the focus of perceptual attention (via eye movements), which in turn informs cognition, which further acts on the task simulation through the execution of motor processing.

Earlier, Related Modeling Work The EPIC framework was chosen a variety of reasons. It was used to model a nearly identical, earlier version of the dual task that was displayed on a single screen (Kieras, Ballas, & Meyer, 2001), and this presented an opportunity to elaborate on a body of existing work. However, EPIC was also chosen because its auditory processing is more complete that of other cognitive architectures, and because, in the core architecture, it accomplishes visual tasks by moving its simulated eyes, which corresponds tightly with our empirical observations of how people executed the dual task in the dual-screen configuration.

In the prior version of the dual task, audio cueing was not used to assist in real-time attention management between subtasks, but to reduce an "automation deficit" that occurred when participants resumed the tactical task after a period during which that subtask was completed "automatically" by the computer (Ballas, et al., 1999).

Similarly, the modeling work presented here examines a different set of issues than did the modeling work of Kieras et al. The prior modeling effort focused on audio-visual localization performance and on the problem of explaining observed patterns of response time sequences corresponding to the negative effects of the automation deficit. Also, in the prior work, switches of attention from tracking to tactical assessment in the model were motivated by the appearance of radar blips in peripheral vision. In the new version of the task presented here, the two visual displays are far apart, and so switches of attention are motivated either by auditory cues or volitional decisions to shift attention.

Architectural Extensions Modeling the complexity of the current dual task required two extensions to EPIC.

The first extension to EPIC was to add a new version of ocular motor movement that corresponds to both an eye *and head* movement. This was needed to model gaze shifts between the two task displays, which are separated by 90° of arc. Longer eye movements are generally accompanied by head movements (Corneil & Munoz, 1996), though the time course of these longer movements can be described by the same time course of smaller eye movements, and is linear for amplitudes from 5° to at least 90° (Becker, 1991). The time course of the newly-programmed eye-and-head movement corresponds to the linear relation given by Carpenter (1988).

The second modification to EPIC was to introduce, in effect, a "sense of timing." The architecture needed a way to maintain an internal sense of timing and priority of a subtask, which in this case was the timing associated with meta-cognitively prompted volitional shifts of attention to the tactical task in the absence of any new perceptual stimuli. (Recall that 56 self-motivated shifts were observed in the sound condition, and 174 in the no-sound condition.) This sense of timing was not needed when modeling earlier versions of the task because the two displays were adjacent and blips could be perceived peripherally.

The sense of timing was introduced via a mechanism in the production rule system. The timing command generates unprompted shifts of attention between tasks based on a generalized form a cumulative gamma distribution, which is characterized by McGill and Gibbon (1965) as useful for modeling multistage processes that are measured as a single reaction time. The distribution is characterized by two free parameters that specify its shape and scale. Manipulations of these parameters in the context of the model's comparative performance with and without audio cues are evaluated below in the discussion of the model's performance.

The Model

An EPIC model was constructed to simulate and predict how people perform the dual-display dual task. The model's organization largely follows the same hierarchical scheme developed by Kieras et al. (2001). A top-level executive

process controls the execution of three sub-processes, two of which carry out the subtasks of tracking and tactical assessment. The third sub-process performs a global monitoring role and updates working memory based the state of the radar display. The functions of these three sub-processes in the baseline model are described next, including how they have evolved from original model.

The Tracking Task The new model's tracking implementation remains an intentionally simple process that continuously follows the target with the eyes and manually pushes the cursor toward it. Consistent with participant behavior discussed by Ballas et al. (1999), the model suspends tracking when it turns its attention to the tactical assessment task.

The Monitoring Subtask The monitoring subtask updates working memory with status changes in the environment. Most of the responsibilities are carried over from the previous model, although its role in the allocation of attention between the tasks has changed. Formerly, this process ran in parallel with the tracking process and notified the dual-task executive of changes in the radar display to prompt a task switch. In the current model, however, with no peripheral access to the radar display, the monitoring process is used to trigger timing commands every time the tracking task resumes. These commands start EPIC's new "sense of timing" clock which, based on a gamma distribution, stochastically determines an appropriate time in the near future to notify the dual-task executive that it is time to switch to the tactical task.

As in the original model, the monitoring process also monitors the visual status of blips and notifies the dual-task executive as these events occur. In the new model, the monitoring process also classifies blips as hostile or neutral. An analog of this classification function was present in the single-screen model, but its parameters were different. Independently established free parameters for the time needed to inspect the behavior each of the three blip types were used in lieu of modeling eye movements for which there was no empirical data. An analysis of the participant response time data in Brock et al. (2004), however, suggests that, counter-intuitively, response times for red and blue blips were roughly equivalent to those for yellow blips. Thus, in the new model, all colors of blips are classified by the monitoring process using a single time parameter.

The Tactical Assessment Task The tactical assessment task is very complex and operates at the same level as the tracking and monitoring sub-processes. Unlike tracking and monitoring, though, tactical assessment is hierarchically organized as an executive sub-process. It controls the execution the three sub-sub-processes that select, classify, and respond to blips. Although the new model implements a number of changes here, only the blip selection sub-process differs substantially. It now follows a more straightforward search logic that is based on a careful reanalysis of the visual selection task.

A detailed description of the tactical subtask is beyond the scope of this paper. The task is very complex, though it is modeled with great detail and fidelity. Once the eyes arrive on the tactical radar display, a great many decisions are made and continue to be made throughout the subtask. Decisions pertain to which blips to put the eyes on, when and whether to classify blips as neutral or hostile, when to move to another blip based on the color of the currently-fixated blip and other blips, when to move the eyes based on changes in blip status during the task, the manual motor process of entering blip classifications, and even eye movements to the keypad.

The sub-processes include selection, classification, and response. Consistent with the architecture's constraints, though, much of the cognitive and motor processing for these sub-processes can overlap. The response sub-process first waits for the monitoring process to classify the selected blip as hostile or neutral, and then uses this information in working memory to select the appropriate key and carry out the keystroke. The response sub-process contains a simple probabilistic rule that causes the model to occasionally move its gaze to the keyboard while executing the keystroke. This contributes to both the gaze shifts and increased response times that are observed in the no-sound condition.

How the Model Responds to Audio Cues The new model performs the dual task on dual screens both with and without sound. To respond to audio cues, a rule was added to the top-level executive process to listen for audio cues. When a cue is detected, the tracking task is suspended and the gaze moves to tactical display. It was not necessary to add rules to the blip selection sub-process to associate the cues with their corresponding blip shapes. However, the rules in the sub-process that try to classify a black blip before returning to the tracking task led to unrealistically fast performance in the sound condition. Accordingly, this part of the selection strategy was removed for the sound condition. The implications are interesting, and discussed in the next section on modeling results.

Modeling results

Once the task analysis was implemented and other aspects of the model's structure were settled, its free parameters were derived and its performance strategy was adjusted for each condition. The time required for blips to be classified as either hostile or neutral was calculated by running a version of the model in which only audio cues prompted looks to events on the tactical assessment display, and the timing of the classification procedure was set to zero. The resulting response times represented the performance overhead associated with selecting and responding to blips. The times were subtracted from the empirical mean for the screen-centric sound condition. The difference (670 ms) was used as the time required to classify all blips.

The model's response time performance in the sound condition with this fitted classification time parameter, though, was unrealistically fast. In addition, it was spending too much time on the left screen classifying black blips. This consequence of the modeling suggests that, in the sound

condition, participants did not spend time trying to classify black blips as they did in the no-sound condition. This aspect of the tactical sub-strategy was thus removed for the sound condition.

The shape and scale parameters of the gamma distribution used to simulate self-prompted shifts of attention in the baseline condition were determined manually. A gamma probability density curve was fit to the frequency histogram of observed right-to-left gaze transition latencies, which were taken to represent the duration of dwell times on the tracking task. The values of the shape and scale parameters of this fitted distribution were respectively 2.5 and 0.95. The tail of this distribution was noticeably steeper than the tail of the empirical data. An explanation for this discrepancy might be that there was a greater degree of variability in the experimenter's recording process for long latencies. At this point, the fitted model for the no-sound condition was considered complete.

Appropriate gamma distribution parameters were determined to motivate the 56 additional self-motivated shifts of attention in the sound condition. The gamma function's shape parameter is commonly interpreted as the *n*th occurrence of some event. Taking this to be descriptive of an internal process that determines when a self-prompted look to the tactical decision task should be carried out, it can be reasoned that the same process is likely to apply in both the no-sound and sound conditions, only at different rates. Therefore, the shape parameter should be held constant, and only the scale parameter varied across conditions. Using this reasoning to make the final fit, it was quickly determined that widening the scale parameter to 2.5 in a run of the model in the sound condition resulted in an average of 120 looks to the tactical assessment task. This difference in the scale parameter effectively measures the increase in meta-cognitive volitional processing necessary when sound is removed from the task environment.

Figure 3 compares the fitted model's performance in the no-sound and sound conditions. Each of the performance measures shown for the model is the mean of six randomly seeded runs, each of which was driven by a different tactical task scenario. The model's close fit with the mean number of gaze shifts in each condition is a direct consequence of the stochastic approach used to simulate self-prompted looks.

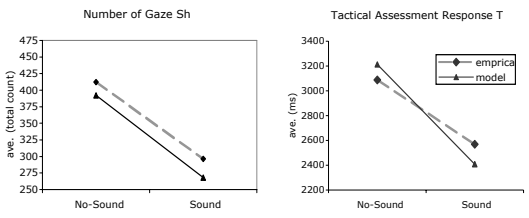

Figure 3: The observed and predicted response times and total number of gaze shifts (based on head turns) for the no-sound (baseline) and sound conditions.

Table 2 shows mean counts of the model's gaze shifts among the two displays and the keyboard, along with the mean empirical counts from Table 1. The most important of these shifts are the counts of right-to-left looks, which capture the model's allocation of attention to the tactical assessment task. As can be seen, this measure of the model's performance is quite close to the empirical data.

Table 2: Predicted/observed counts of attentional shifts (based on head turns) in the no-sound and sound conditions. Location key: right = tracking and left = tactical.

No-Sound - Mean Count of Attentional Shifts

Shift from	Right	Left	Keybd	
Right to	0	174/174	0/7	174/180
Left to	168/170	0	25/27	193/197
Keybd to	7/10	18/23	0	25/34
	175/180	192/197	25/34	392/411

Sound - Mean Count of Attentional Shifts

Shift from	Right	Left	Keybd	
Right to	0	120/121	0/6	120/127
Left to	93/116	0	28/23	121/139
Keybd to	27/11	0/19	0	27/29
	120/127	120/139	28/29	268/295

Discussion

Two particularly interesting aspects of this model include (a) the difference blip assessment strategies necessary between the no-sound and sound conditions and (b) the model's emergent behavior of looks away from the keypad.

There are several reasons why participants might assess black blips much less frequently in the sound condition. Looks in the sound condition are in part driven by prompts in the task environment and, as a result, are generally more efficient. If a participant turns away from a blip that is about to change color in this condition, he or she is alerted to this fact. In the no sound condition, however, the cost of turning back to the tracking task just before a blip changes color is much greater because the response time for that blip is more likely to be poor. As a result, participants have incentive to dwell on the left screen longer when they feel a blip is getting close to changing color.

The model's emergent pattern of looks *away* from the keyboard is quite interesting. Looks *to* the keyboard were modeled for fidelity. However, the implementation led to an unforeseen result: The model reveals how looks *away* from the keyboard interact with the blip selection sub-process. When color-coded blips remain on the left screen, the model always returns to the tactical assessment task from the keyboard. In all other circumstances, the model returns to the tracking task. No attempt was made to motivate these moves. It is particularly interesting that in the no-sound condition, the proportion of gaze shifts away from the keyboard in either direction matches observed data. This strengthens the likelihood that the blip selection sub-process used in this condition is close to what participants actually used. It also follows that the corresponding disparity in the sound

condition suggests that the blip selection sub-process is more subtle.

Eye-tracking data will enable us to directly examine aspects of the sub-strategies that participants use in the two conditions. For instance, it will show how often, and in which conditions, participants spend time looking at black blips; whether time spent on a black blip directly benefits its corresponding response time; whether or not assessments that are interrupted effect response times; whether subjects spend time on the left screen after a gaze shift from the keyboard; and whether the left screen is only a brief stop for the eyes on their way back to the tracking task.

Conclusion

The long-term motivation for the modeling effort presented in this paper is to analyze and predict the costs and benefits of using 3D audio in the information displays of operational settings such as those the Navy expects to deploy in the next ten to fifteen years. Although many aspects of the model's implementation, its performance strategies, and the process of deriving appropriate values for its free parameters may not appear to comment directly on this goal, several important aspects of generally overlooked issues in the simulation of human performance are addressed here. A computational model of the performance benefits associated with an uncluttered auditory information design is presented. The model addresses the problem of usefully characterizing the parameters of self-regulated allocation of attention. The model predicts the effect of system level strategies for ameliorating effort when concurrent demands are involved.

Multi-task operational settings can be notoriously more complex than the dual task modeled here, but designers of supervisory control systems absolutely need to know the baseline requirements for the allocation of attention before they can design and implement effective attention management solutions. In particular, the ground truth for modeling inter-task performance depends on knowing the demands of process combinations of unassisted access rates to information that must be acted upon, acceptable levels of error, and requirements for initiative and physical effort that can be quantified.

Acknowledgments

This work was supported by the Office of Naval Research under work request N0001406WX20042.

References

Ballas, J., Heitmeyer, C., and Perez, M. (1992). Evaluating two aspects of direct manipulation in advanced cockpits. In *CHI'92 Conference Proceedings: ACM Conference on Human Factors in Computing Systems.* Monterey, CA.

Ballas, J., Kieras, D., Meyer, D., Brock, D., and Stroup, J. (1999). Cueing of display objects by 3-D audio to reduce automation deficit. In *Proceedings of the 4th Annual Symposium and Exhibition on Situational Awareness in the Tactical Air Environment*. Patuxent River, MD, 1999.

Becker, W. (1991). Saccades. In (Carpenter, R. H. S., ed.) *Eye movements*. Boca Raton, FL: CRC Press.

Begault, D. (1994). *3-D sound for virtual reality and multimedia.* Chestnut Hill, MA: AP Professional.

Brock, D., Ballas, J., Stroup, J., and McClimens, B. (2004). The design of mixed-use, virtual auditory displays: Recent findings with a dual-task paradigm. *Proceedings of the 10th International Conference on Auditory Display*. Sydney, Australia.

Brock, D., Stroup, J. and Ballas, J. (2002a). Effects of 3D auditory cueing on dual task performance in a simulated multiscreen watchstation environment. In *Proceedings of the Human Factors and Ergonomics Society 46th annual Meeting*. Baltimore, MD.

Brock, D., Stroup, J. and Ballas, J. (2002b). Using an auditory display to manage attention in a dual task, multiscreen environment. In *Proceedings of the 8th International Conference on Auditory Display*, Kyoto, Japan.

Carpenter, R. (1988). *Movements of the eyes.* London: Pion.

Corneil, B., and Munoz, D. (1996). The influence of auditory and visual distractors on human orienting gaze shifts. *Journal of Neuroscience*. 16(24), 8193-8207.

Kieras, D., Ballas, J., and Meyer, D. (2001). *Computational models for the effects of localized sound cuing in a complex dual task.* TR-01/ONR-EPIC-13, University of Michigan, Ann Arbor, MI.

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

Osga, G. (2000). 21st Century Workstations: Active partners in accomplishing task goals. In *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, San Diego, CA.