The *Functional Resource Hypothesis* as a Basis for Understanding Cognitive Workload in Immediate Interactive Behavior

Jason Ralph, Wayne D. Gray, & Michael J. Schoelles
Rensselaer Polytechnic Institute
Troy, New York

Abstract

Understanding workload requires understanding the control of cognition at the 1/3 to 3s time span during which cognitive, perceptual, and motor operations become bound together into interactive routines. Interactive routines constitute unit tasks (3 to 30 s), and unit tasks constitute subtasks (30s to 3min). To reduce cognitive workload and overload, the *Functional Resource Hypothesis* maintains that an optimal allocation of interactive routines to task performance would be based on the functional resource of time not modality. Some of the implications of this hypothesis are investigated in an empirical study that varied memory load as well as the demands on the eyes, visual attention, auditory cognition, and motor operations. A microanalysis of the data revealed tradeoffs between groups in their pattern of resource allocation that were compatible with the Functional Resource Hypothesis.

Cognitive Workload and Cognitive Theory

Cognitive workload refers to the decrement in operator performance due to an overload of the human operator's cognitive, perceptual, and/or motor resources. For the human factors community, a goal as been to develop applied theories that are robust enough to guide the design of systems that reduce workload. Unfortunately, our current understanding of workload is very general and current measures of workload leave much to be desired in terms of predictive validity. The purpose of our current research program is to develop model-based measures of cognitive workload that take into account current theories of cognitive science and are built around understanding how task demands affect the allocation of cognitive, perceptual, and motor resources. We present an empirical study that varies the demands on physical resources such as the eyes and hands, as well as cognitive resources such as memory, visual attention, and auditory attention. We then present a microanalysis of our data that reveals tradeoffs between and within groups in their patterns of resource allocation.

In this section we present a brief review of the two major cognitive science based theories of cognitive workload and then present a sketch of our *Functional Resource Hypothesis*. In the next section we introduce our paradigm and methodology, and present some of our findings.

Cognitive Science Based Theories of Cognitive Workload

Cognitive Load Theory (John Sweller, 1988; John Sweller, van Merrienboer, & Paas, 1998) and *Multiple Resource Theory* (Wickens, 2002) are the dominant cognitive science based theories of workload in the fields of instructional design and human factors, respectively. Cognitive Load Theory focuses on working memory to identify factors in the task environment that make demands on this resource. A major focus of this work has been the workload on memory required to integrate multiple sources of information. For example, Leahy, Sweller, and Chandler (2003) use the example of integrating text explanations with graphical representations of data. The cognitive load necessary for this integration is greatly reduced by overlaying the text with the figure, eliminating the need for the user to glance back and forth to integrate the two pieces of information into a single concept. Several studies have demonstrated the increase in performance realized by exploiting this split-attention effect (Chandler & Sweller, 1991; J. Sweller, Chandler, Tierney, & Cooper, 1990; Tarmizi & Sweller, 1988).

Cognitive Load Theory further contends that separate modalities (e.g. visual, auditory) have independent channels of working memory capacity. Several studies have been interpreted as showing that people are able to store more information in working memory when using two modalities than when only using one modality (Leahy, et al., 2003). The modality effect may help to mitigate the split attention effect by presenting independent streams of information in different modalities (e.g., presenting the text descriptions of the figure aurally instead of overlaying them on the figure). Researchers report evidence that multi-modal displays and feedback lead to increases in performance and self-reports of task difficulty (Vitense, Jacko, & Emery, 2003). Cognitive Load Theory provides guidance in reducing memory load, but does not allow us to predict the amount of load elicited by a task. It also does not explain how memory load affects other resources such as attention and control.

Multiple Resource Theory (Wickens, 2002), which has been represented as a 2 x 2 x 2 cube of resource allocation, attempts to explain how separation of resources may reduce workload. The three dimensions of the cube represent modality (auditory and visual), stages (Perception/ cognition and responding) and codes (spatial and verbal). The suggestion is that two concurrent tasks, both of which require the use of the same block of the cube (e.g. the same modality or verbal information presented by 2 modalities), will result in an increase in cognitive load and a decrease in performance. Multi-tasking that does not share resources will not result in the same increase in load and consequently will not decrement

10.1518/107118109X12524441081668

performance. Unfortunately, although it does provide a way of explaining many experimental results, Multiple Resource Theory has struggled to provide accurate predictions of human performance in dual task experiments (Dixon, Wickens, & Chang, 2005; Horrey & Wickens, 2004).

Functional Resource Hypothesis

Although a dual task experiment may provide information in multiple modalities, it is not clear that allocating different resources to different tasks provides the key to understanding cognitive workload. *Milliseconds matter* in the control of the integrated cognitive system (Gray & Boehm-Davis, 2000) and at the 1/3 to 3s level of analysis cognitive, perceptual, and action operators are bound together into elementary sets of interactive routines. Each of these interactive routines contains multiple resources and the tradeoff is not between a strict dichotomy of, say, vision or memory, but between blends of resources which may be more or less interaction-intensive versus more or less memory-intensive (Gray, Neth, & Schoelles, 2006; Gray, Sims, Fu, & Schoelles, 2006). Understanding optimization at this millisecond level of analysis can provide a theoretical foundation for understanding workload at longer time intervals, such as the level of 10s of seconds or minutes discussed in Wickens' and Sweller's work. A focus on this millisecond level optimization may lead to more predictive workload models.

Prior work has led us to formulate the *soft constraints hypothesis* (Gray & Fu, 2004; Gray, Sims, et al., 2006), one aspect of which is that the cognitive system allocates cognitive, perceptual, and motor resources as if to conserve the *functional resource of time*. The *cognitive impartiality principle* follows from the soft constraints hypothesis:

> The central [cognitive] controller makes no functional distinction between knowledge in-the-head versus in-the-world or the means of acquiring that information (such as eye movement, mouse movement and click, or retrieval from memory) (Gray & Veksler, 2005)

(see also, Clark, 2008, chapter 6).

Functional Resource Hypothesis for Cognitive Workload maintains that the functional resource being conserved is time and that understanding how the designed environment influences workload requires an understanding of how the cognitive controller selects one set of interactive routines rather than another. The focus of the Functional Resource Hypothesis is on immediate interactive behavior ranging from 1/3 of a second to 3 seconds. An auxiliary assumption is that local optimization at this level of analysis may not yield global optimization at the unit task (3 to 30s) or subtask levels of analysis (30s to 3min).

The Study

The *NavBack* paradigm was designed to collect detailed empirical data in a task at the approximate complexity of those used by Sweller and Wickens in the studies discussed earlier. The NavBack task combines a tracking task with a memory task. Like other tracking tasks (Ballas, Heitmeyer, & Perez, 1992; Martin-Emerson & Wickens, 1992; Salvucci, Taatgen, & Kushleyeva, 2006) our jitter task requires that the operator attend to the task to keep the vehicle (in our case, an arrow) centered. Like the n-back memory task (Gevins & Smith, 2003; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; McEvoy, Smith, & Gevins, 1998) we make demands on memory and cognitive control by requiring the operator to maintain a list of items in memory and to update that list every few seconds. In our case, the list maintained is a list of instructions for what to do at the next three intersections; for example, turn left, turn right, or go forward. After each intersection the operator has to delete the just completed instruction and add a new instruction to the end of the mental list. Hence, in the NavBack paradigm the two related tasks of "jitter" and "turning" make demands on memory, attention, and action. Across conditions our subjects are either presented their turn instructions auditorily or visually. Our analyses of these data will be interpreted as showing a rich interleaving of mental resources that are sensitive to the time constraints of the task measured in milliseconds. As we have argued, it is here where our microanalysis of behavior must look for a theory of workload that is sensitive to the functional resource of time.

Methods

Subjects

68 undergraduate students of Rensselaer Polytechnic Institute (mean age = 19) volunteered to participate in this study for course credit. Twenty-three subjects were randomly assigned to one of three conditions: *Transient Visual*, *Auditory*, or *Persistent Visual*. One subject from the Persistent Visual condition was removed from the analysis due to extremely poor performance on the turning task leaving 22 subjects in that condition.

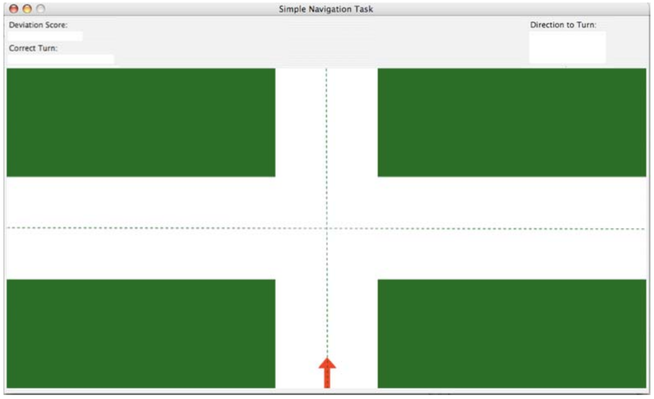

Figure 1: Screen shot of the NavBack task showing city blocks (in green), the arrow "vehicle" (in red), the roads (white), the center line (dotted), and an intersection.

Apparatus and Materials

The experiment was run on an Apple Mac Mini computer (running Mac-OS 10.4) at a 1024x768 screen resolution. Eye fixation data were collected using an LC Technologies tracker at a 60 Hz sampling rate. A chinrest was used to stabilize head movements and ensure a fixed viewing distance of 60 cm. We were unable to calibrate 12 subjects with the eye-tracker. These subjects are included in the analysis of performance data, but not in the analysis of eye data.

The NavBack software is a custom application implemented in Lispworks 5.1. NavBack, (see Figure 1), is a simple navigation task combining two related, but independent tasks of cursor control (jitter) and navigation (turning).

Design

The jitter task was constant across each of the three conditions. All conditions received a new direction (the turn to take 3 intersections in the future), 1s after they left the last intersection. The *Transient Visual* group received turn directions in the box at the top right corner of the display. The direction remained on the screen for 2s. The *Auditory* group received instructions via a computer-generated voice. In common with the Transient Visual condition, the *Persistent Visual* group received instructions in the direction box, but turn instructions for all of the next three intersections remained on the screen at all times. The Persistent Visual condition was included as a control condition as the persistent nature of the turn instructions eliminated the need to perform the working memory task of updating and maintaining the list of three turn instructions.

Procedure

Each subject completed a 2-min practice session to familiarize them with the demands of the task, followed by two 20-min sets of episode cycles. The two sets were separated by a required five-minute break.

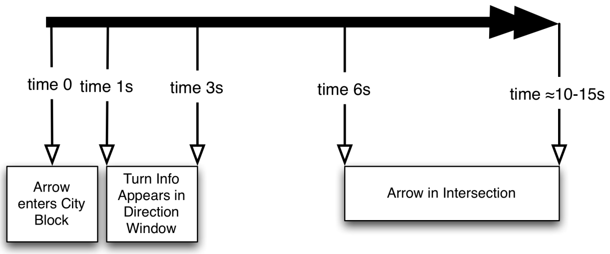

Figure 2: Episode Cycle of the NavBack Task for the Transient Visual condition.

Each 20-min set consisted of a continuous series of *episode cycles* (see Figure 2 for the episode cycle for the Transient Visual condition). For analysis purposes each cycle began when the arrow left an intersection and entered the next *city block* (city blocks are the green areas in Figure 1). One second after the arrow entered the new city block a new turn instruction appeared in the Direction box on the upper right of the screen (see Figure 1) for the two Visual conditions or was presented auditorily for the Auditory condition. At 6s the arrow entered the next intersection.

Time in the intersection varied depending on when the subject made her turn. Minimizing the jitter score required the subject to turn at the exact center of the intersection at 8,000 ms. However, once they were in the intersection, subjects could turn at any time. The animation for the turn added 1,500 to the intersection time.

On each cycle subjects had to do two related tasks: the jitter task and the memory-update task. These two tasks are discussed next.

Jitter Task. Subjects were to keep the arrow in the center of the road (on the dotted line in Figure 1) as the arrow “jittered” from side to side based on a pseudo-random function. Subjects corrected the arrow’s horizontal position by pressing the ‘a’ and ‘d’ keys on a standard keyboard. Their goal was to keep the arrow as close to the center of the lane as possible. The computer logged the absolute value of the number of pixels deviated from the center every 200 ms. Feedback on the average jitter score between the most recent two intersections was constantly present at the top left of the screen (see Figure 1).

Memory-Update Task: Concurrent with the Jitter Task, subjects were required to keep three turns in memory (e.g., left, right, forward). At the beginning of a 20-min set, subjects were presented with the initial three turns that were to be made in the first, second, and third intersection that the arrow entered. After the set began, subjects were presented with a new turn direction 1s after they entered a new city block. Those in our two memory conditions (Transient Visual and Auditory) were required to mentally delete the instruction for the just completed turn and to append a new instruction at the end of their mental list. Subjects in the Persistent Visual condition could glance at the Direction box to get the new instruction anytime after 1s in the new city block and before they decided to turn in the next intersection. Feedback on the correctness of the most recent turn was available at the top left of the screen, below the jitter feedback (see Figure 1).

Results

Performance data

For all analyses reported in this section, we divided the two 20-min sets into 8 equal mini-sets of 5-min each.

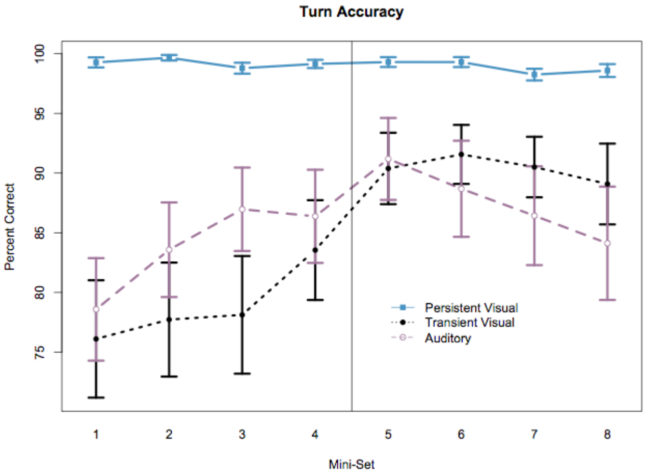

Figure 3: Turn accuracy by mini-set. (Error bars represent the standard error.) The vertical line denotes the break between 20-min experimental sessions

Turn Accuracy. A mixed, 3 condition (between) by 8 mini-set (within) analysis of variance (ANOVA) was conducted on turn accuracy. The main effect of condition was significant, *F (2, 65) = 8.67, MSE = 11482, p < .0001.* The main effect of mini-set [*F (7, 455) = 7.49, MSE = 683, p < .0001*] was also significant as was the interaction of condition by mini-set [*F (14, 455) = 3.46, MSE = 316, p < .0001*].

As expected, the Persistent Visual condition was at ceiling with minimum variance compared to the Transient Visual and Auditory conditions (see Figure 3). Therefore, we conducted a 2nd ANOVA to compare Transient Visual with Auditory on this measure. This ANOVA showed no effect of condition (*p* = 0.81) but a significant effect of mini-set [*F (7, 308) = 7.78, MSE = 1036, p < .0001*] and a significant interaction of condition (Transient Visual versus Auditory) by mini-set [*F (7,, 308) = 2.06, MSE = 274, p = 0.047*].

The Transient Visual condition starts out lower but improves faster than the Auditory condition. Interestingly, the Transient Visual maintains its improvement throughout the last part of the study whereas the Auditory condition shows a decline.

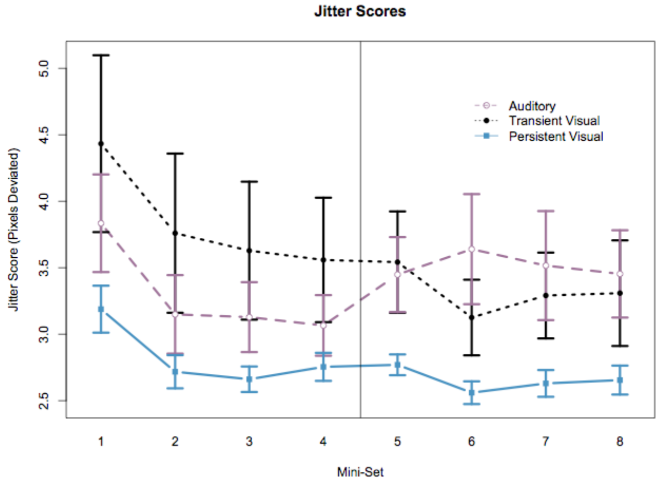

Figure 4: Jitter scores by 5-min mini-set (lower is better). Error bars show the standard error. The gray vertical line denotes the break between the two 20-min sets.

Jitter Scores. A 3 condition (between) by 8 mini-set (within) analysis of variance (ANOVA) was conducted on jitter scores. The main effect of condition was not significant (p = 0.13). However, there was a significant effect of mini-set $F(7,455) = 5.95\ MSE=3.87\ p < .0001$ which interacted with condition, $F(14, 455) = 1.81, MSE = 1.17, p = 0.04$ (see Figure 4, note that for jitter scores, lower is better than higher).

It is interesting and maybe important to note that the performance of the Auditory group on jitter follows the same pattern as for turn accuracy. Auditory does best during the first half of the study and performs worse during the second half. In contrast the Transient Visual condition shows a steady improvement throughout the first 5 mini-sets and appears to plateau over the last three mini-sets.

Eye Data

Several effects of condition and mini-block on eye activity were found. As increases in fixations toward the direction box necessarily correlate with decreases in fixations on the arrow, we will focus our analysis on the direction box and limit our analysis to the two visual conditions.

A 2 by 98 mixed ANOVA was conducted where the between factor was Transient Visual versus Persistent Visual and the within factor was 50 ms time bins ranging from 100-5000 ms after the arrow entered a new city block. (Note that for the Transient Visual condition the turn direction was available between 1000 to 3000 ms. To better highlight this period Figure 5 only shows data over the first 3900 ms.)

The main effect of condition (transient versus persistent) was not significant (p = 0.95). However both the main effect of time bin [$F(97, 3394) = 63.87, MSE = 0.315, p < .0001$] and interaction of condition by time bin [$F(97, 3394) = 15.44, MSE = 0.076, p < .0001$] were significant.

Figure 5 shows that the Transient Visual condition tended to look at the direction box soon after the new direction appeared. In contrast, the Persistent Visual condition, which could find the direction at any time, looked less during this interval, their peaks occurred later, and they looked more during the period leading up to the intersection (when the instruction would have vanished for the transient group).

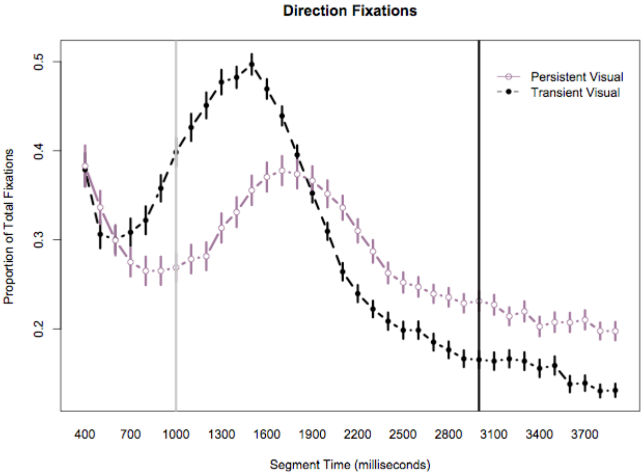

Figure 5: Dwells (and standard errors) over the first 3900 ms of the episode cycle to the Direction box as a function of experimental condition. The new direction appears for both the Persistent Visual and Transient Visual (gray line) 1s after the arrow enters the new city block and vanishes for the Transient Visual (but not the Persistent Visual) 2s later (at the black line).

Discussion

The NavBack task yields an interesting pattern of performance differences related to turn instruction modality. On two different measures, jitter and turn accuracy, the Auditory group starts better than the Transient Visual group but by the end of the study is performing no better but maybe worse than that group.

Just as noteworthy is the fact that the best performance on both tasks is produced by the Persistent Visual condition. Although the turn accuracy performance might be expected, their superiority on the jitter score shows that they were able to interleave glances at the direction box with performance on the jitter task. As the Auditory condition provided little incentive for subjects to take their eyes off the moving arrow, it is noteworthy that the jitter performance of the Auditory group seems to have been strongly affected by the demands of the working memory task. It is not clear that either Cognitive Load Theory or Multiple Resource Theory would have made that prediction.

The comparison of eye dwells on the direction box during the first 3900 ms of each cycle reveals important differences in information seeking between the Transient and Persistent Visual conditions. The Transient group starts looking at the Direction box sooner than the Persistent group and peaks earlier. It is clearly the case that it would be more costly to the Transient group to look late and risk not seeing the new instruction than it is to look early and, thereby, keep the eyes away from the arrow (and jitter tasks) longer. However, despite this pattern of performance, by the end of the study the Transient Visual group does slightly better on the jitter task than does the Auditory group. (Please note that we have examined many other aspects of these data but space does not permit a fuller report.)

Summary and Conclusions

Cognitive Load Theory and Multiple Resource Theory attempt to separate and isolate resource conflicts. Although not incompatible with these results, our work with the soft constraints hypothesis (Gray & Fu, 2004; Gray, Sims, et al., 2006) has demonstrated that the brain conserves the functional resource of time at the expense of resources such as memory, attention, perception, and action. Hence, the focus of the Functional Resource Hypothesis of cognitive workload is not on the use of different pools of resources per se, but on finding sets of interactive routines that minimize processing time for the resources that are required for the task at hand.

We recognize that while the results presented in this short paper are intriguing, they provide only weak evidence for the Functional Resource Hypothesis. Further studies and more analyses are underway. However, the strongest support for the Functional Resource Hypothesis must await the development a computational cognitive model that adapts to our three experimental conditions to make the same resource tradeoffs as our human operators did.

Acknowledgements

Research reported in this paper has been supported, in part, by grant N000140710033 to Wayne D. Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer. Melissa Gallagher assisted in data collection for this experiment.

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* (pp. 127–134). New York: ACM Press.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293-332.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press.

Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: A workload analysis. *Human Factors, 47*(3), 479-487.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human--computer interaction. *Theoretical Issues in Ergonomics Science, 4*(1-2), 113-131.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds Matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied, 6*(4), 322–335.

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*(3), 359-382.

Gray, W. D., Neth, H., & Schoelles, M. J. (2006). The functional task environment. In A. F. Kramer, D. A. Wiegman & A. Kirlik (Eds.), *Attention: From theory to practice* (pp. 100-118). New York: Oxford University Press.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*(3), 461-482.

Gray, W. D., & Veksler, V. D. (2005). The acquisition and asymmetric transfer of interactive routines. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *27th Annual Meeting of the Cognitive Science Society, CogSci2005* (pp. 809-814). Austin, TX: Cognitive Science Society.

Horrey, W. J., & Wickens, C. D. (2004). *Cell phones and driving performance: A meta-analysis*. Paper presented at the Human Factors and Ergonomics Society 48th Annual Meeting.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829-6833.

Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia. *Applied Cognitive Psychology, 17*(4), 401-418.

Martin-Emerson, R., & Wickens, C. D. (1992). The vertical visual field and implications for the head-up display *Proceedings of the 36th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1408–1412). Santa Monica, CA: Human Factors and Ergonomics Society.

McEvoy, L. K., Smith, M. E., & Gevins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: effects of memory load and task practice. *Cereb. Cortex, 8*(7), 563-574.

Salvucci, D. D., Taatgen, N. A., & Kushleyeva, Y. (2006). Learning when to switch tasks in a dynamic multitasking environment. In D. Fu, F. Del Missier & A. Stocco (Eds.), *ICCM-2006: The Seventh International Conference on Cognitive Modeling* (pp. 268-273). Trieste, Italy.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology-General, 119*(2), 176-192.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem-solving. *Journal of Educational Psychology, 80*(4), 424-436.

Vitense, H. S., Jacko, J. A., & Emery, V. K. (2003). Multimodal feedback: an assessment of performance and mental workload. *Ergonomics, 46*(1-3), 68-87.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science, 2*(3), 159-177.