

## Squeezing the Balloon: Analyzing the Unpredictable Effects of Cognitive Workload

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

Cognitive workload effects behavior like squeezing a balloon. If you squeeze at one place, it pops out at another, and it is hard to predict where it's going to pop out. 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 and led to surprising behavioral effects.

Cognitive workload, as the term suggests, refers to the amount of "work" required of the cognitive system. People tend to make more errors and perform slower on tasks when workload increases. Although we are able to measure workload, our current understanding of the factors that lead to increases in workload is vague and incomplete. We know, for instance, that tasks that require more of a given resource, such as memory, elicit increased workload (Gevins & Smith 2000; Gevins & Smith 2003). But to this point we do not know enough about resource allocation interactions to develop a predictive model of workload.

We are able to detect workload levels via self-report questionnaires such as the NASA TL/X after performance of a task, but for the human factors community, this information is too little too late. As a result, researchers in the field of augmented cognition have developed impressive real time categorization applications using physiological measures such as EEG, which allow us to measure workload during task performance (Wilson and Russell, 2003a; Wilson & Russell, 2003b) As impressive as they are, these applications do not inform us as to the cause of workload.

### Developing a Theory of Workload

It is clear that understanding resource allocation and use is an important pre-requisite to understanding workload. An influential approach to this is Christopher Wickens' (2002) Multiple Resource Theory (MRT).

MRT, which has been represented as a 2 x 2 x 2 cube of resource allocation, attempts to explain how the 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 performance. Unfortunately, although it does provide a way of explaining many experimental results, MRT has struggled to provide accurate predictions of human performance in dual task experiments (Dixon, Wickens, & Chang, 2005; Horrey & Wickens, 2004).

We believe that MRT's inability to consistently predict performance stems from a misunderstanding of how the brain allocates cognitive resources. Research in our lab focuses on interactive routines, consisting of perceptual, cognitive, and motor operations that are formed at the 1/3 of a second to three second timescale. The cognitive system chooses interactive routines to optimize functional resources such as time. In some cases these routines are interaction-intensive (e.g. require active visual perception and motor response). In other cases they are memory-intensive. We call this the Functional Resource Hypothesis, and it is the basis of our current research program. We believe that understanding interactive behavior at this timescale (1/3 of a second to 3-seconds) is the key to understanding and predicting workload effects at the 3 second to 30-second level of analysis that is the domain of most workload studies.

The creation and choice of interactive routines can be strongly affected by interface design elements. For instance, an interface that requires numerous eye movements can disrupt the creation of routines that involve a constant stream of visual information. Elimination of these saccades (by using auditory cues for instance) can lead to reduced workload and improved performance. Although it might be argued that the MRT cube predicts this outcome, it may be that the performance decrement is not caused by an overload of visual processing, but by an overload of visual movement. We refer to this alternative hypothesis as the *Busy Eyeball Hypothesis*. At issue is whether the decrement we have found is due to central or peripheral resources. The Busy Eyeball Hypothesis argues that the effect is peripheral and does not represent the depletion of more central processing resources.

The determination of whether central or peripheral resources are being depleted is of paramount importance.

Peripheral resource conflicts can be addressed by relatively simple alterations to the user interface. Central resource conflicts are usually inherent to the task and are harder to reduce in the interface design process. Thus an important first step is to determine which types of resource conflicts cause specific performance decrements.

### The Study

The *NavBack* paradigm was designed to collect detailed empirical data in a task at the approximate complexity of those used by Wickens in the studies discussed earlier. The *NavBack* task combines a tracking task with a working 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 keep the vehicle (in our case, an arrow) centered. Like the n-back memory task (Gevins & Smith, 2003; Jaeggi, Buschkuhl, 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 verbally or visually.

### Previous Research Findings

Previous research using the *Navback* paradigm (Ralph, Gray, & Schoelles, 2009) suggested that predictable timing of visual instructions led to anticipatory eye-movements. These just-in-time saccades allowed subjects in the visual condition to minimize time spent not looking at the arrow, allowing them to minimize the workload effects of the busy eyeball hypothesis. In the current study, stimulus presentation was made unpredictable in order to test whether performance decrements were due to visual processing constraints, or visual movement requirements. Our data analysis was conducted at the millisecond level to show how eye movements elicit increased cognitive demands as an example of how design elements can affect cognitive workload.

### Methods

#### Subjects

45 undergraduate students of Rensselaer Polytechnic Institute (mean age = 19) volunteered to participate in this study for course credit. Fifteen subjects were randomly assigned to one of three conditions: *Transient Visual*, *Auditory*, or *Persistent Visual*.

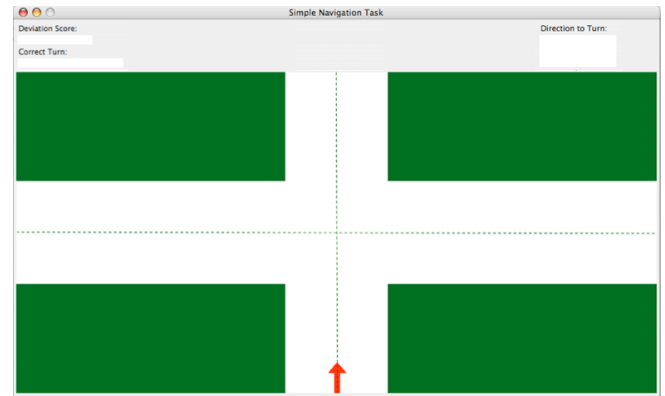


Figure 1: Screen shot of the *NavBack* task showing city blocks (in green), the arrow “vehicle” (in red), the roads (white), the center line (dotted), and an intersection.

### 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.

### 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) during each “city block”. 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 1s. 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 8 5-minute experimental blocks.

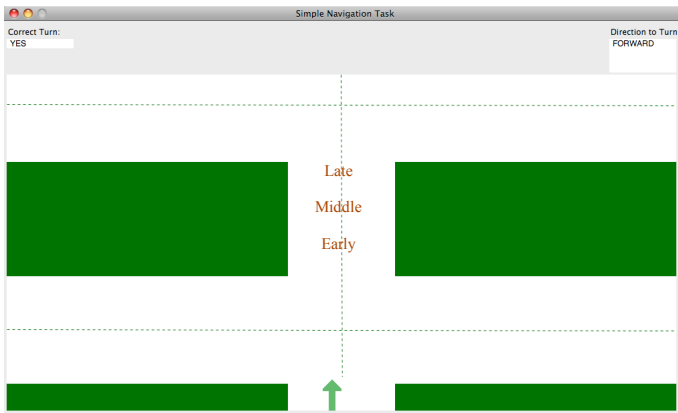


Figure 2: Direction presentation times of the NavBack Task for the Transient Visual condition. Directions were presented at one of three randomly chosen times.

Each 5-minute block consisted of a continuous series of *episode cycles*. Each cycle began when the arrow left an intersection and entered the next *city block* (city blocks are the green areas in Figure 1 and 2). At one of three randomly chosen times a new instruction appeared in the direction box on the upper right of the screen (see Figure 1 and 2) for the two Visual conditions or was presented auditorily for the Auditory condition. Instructions either appeared 1 second, 3.4 seconds, or 5 seconds after the beginning of the episode cycle. 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. Once they were in the intersection, subjects could turn at any time. The animation for the turn added 1,500 ms 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. In order to provide visual feedback, the arrow color changed from green to blue to red as the arrow moved away from the center of the lane. The arrow’s color changed from green to blue when the arrow deviated 5 pixels from the center, and turned red when it deviated 15 pixels from the center.

The arrow’s position at the beginning of each city block was determined by the timing of the previous turn. If the previous turn (left or right) was initiated at the exact vertical center of the intersection, then the arrow began the next city block in the center of the lane. If the turn was initiated early or late, it began the next city block deviated from the center by an amount proportional to the distance from the turn point to the center of the intersection.

The computer logged the absolute value of the number of pixels deviated from the center every 200 ms. We refer to this value as “jitter score” for the remainder of this paper.

### 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 5-min set, subjects were presented with the initial three turns that were to be made in the first, second, and third intersection. After the set began, subjects were presented with a new turn direction at one of the three randomly chosen times described earlier. The direction presentation time was randomly chosen for each city block.

Success in our two memory-update conditions (Transient Visual and Auditory) essentially required Ss 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 (see Figure 1 and 2).

## Results

### Turn Accuracy

A mixed 3 condition (between) by 8 block (within) ANOVA showed a main effect of condition on turn accuracy  $F(2,45) = 10.82$   $p < .001$ . Bonferroni post hoc tests showed that transient visual subjects ( $M=85.70\%$   $SD=9.67\%$ ) made a lower percentage of correct turns than the auditory ( $M=95.06\%$   $SD=4.75\%$ ) and persistent visual ( $M=95.67\%$ ,  $SD=4.17\%$ ) conditions. Both of these comparisons were significant at the  $p < .001$  level. Despite this result, it is difficult to determine if errors made by the transient visual group were made because of difficulty remembering directions, or if these subjects did not look to the direction box in time to acquire the directions

### Jitter Task

Figure 4 is a graph of jitter scores by condition and time (block). An ANOVA was performed on jitter scores by condition and found a significant effect of condition  $F(2,45) = 5.63$   $p < .01$ . Bonferroni post hoc analyses determined that the auditory condition ( $M=5.36$   $pixels$   $SD=0.64$ ) significantly outperformed both visual conditions (transient visual  $M=7.31$   $pixels$   $SD=2.45$ ; persistent visual  $M=7.44$   $pixels$   $SD=2.12$ ) on this metric. Although this is a significant result, we wanted to look deeper, and examine the specific time periods within a city block that are the key to understanding the effect.

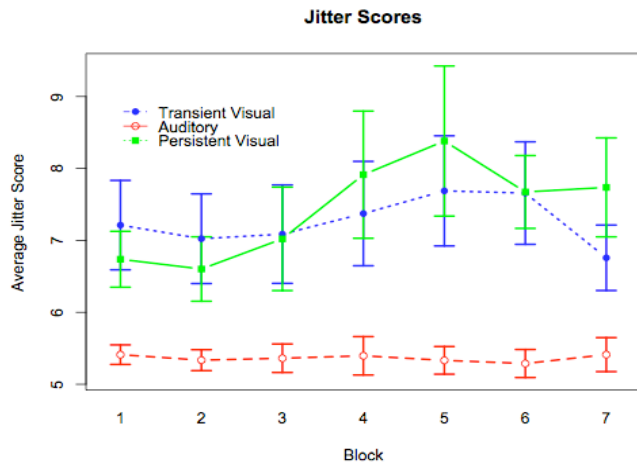


Figure 3: Mean jitter scores by 5-minute block for each condition

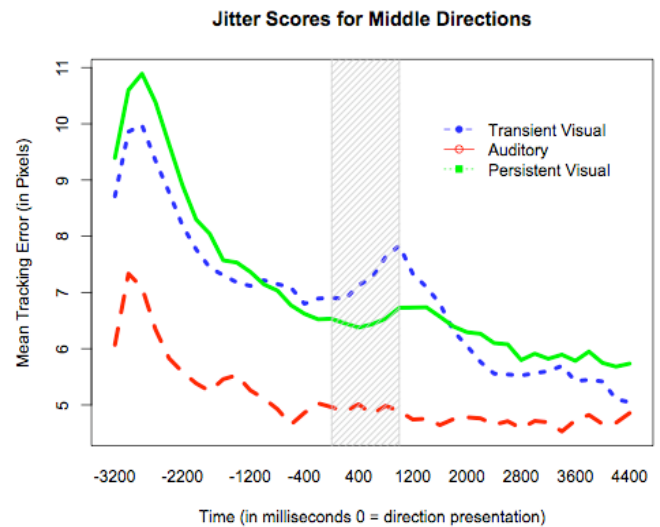


Figure 4: Jitter scores in an average “city block” (stripped gray region indicates the 1000 ms duration of the turn instruction). Notice that only the Transient Visual condition is affected by the presentation of the turn instruction.

Figure 4 displays the average jitter score as a function of time within a city block for all trials in which the direction was presented at the middle (3.4 seconds) presentation time. There are 2 significant time periods to consider. The first is the time period immediately surrounding the presentation of the direction (shown as the 1-s wide, gray-striped region of the plot), and the second is the beginning of the city block (shown as the leftmost data points in figure 4).

As the busy eyeball hypothesis would have predicted, the presentation of the direction has an effect only on the jitter scores of the transient visual condition (dotted blue line in Figure 4). We focused on the middle direction trials and performed an ANOVA to determine if jitter scores varied before, during, and after the presentation of the direction for the transient visual condition. Jitter scores surrounding the direction presentation were split into three 2-second time intervals. The first interval began 2 seconds before the direction presentation and ended 200 ms before the direction was presented (1.4 – 3.2 seconds). The second began at the direction presentation time (3.4 – 5.2 seconds) and the third began 2 s after the direction was presented (5.4 - 7.2 seconds).

We used 2-second intervals in order to include the full effect of direction presentation in the second time interval. As Figure 4 shows, it took approximately 2 seconds for transient visual subjects to recover from the effect of the direction presentation. There was a significant effect of interval on transient visual jitter scores  $F(2,138) = 3.83 p < .05$ . Bonferroni post-hoc tests showed that jitter scores recorded after the direction presentation ( $M = 5.83$  pixels  $SD = 1.94$ ) were significantly lower than those recorded before ( $M = 7.02$  pixels  $SD = 2.14$  pixels) and during ( $M = 6.61$  pixels  $SD = 2.19$  pixels) direction presentation. These comparisons were significant at the  $p < .05$  level. Thus, in addition to affecting jitter performance as the direction was presented, the need to monitor the direction box until it appeared helps to account for the difference between transient visual and auditory performance.

This result is backed up by eye data, which showed a high percentage of eye fixations on the direction box for the transient visual condition until the direction appeared. Figure 5 displays the average proportion of total fixations on the direction box for transient visual subjects for each of the three direction presentation times (early, middle, and late). Saccades to the direction box persisted and increased for each presentation time until the direction appeared. Also, as figure 5 demonstrates, the large spike in eye movements to the direction box occurs just after the direction appears. In a previous study in which the direction presentation time was more predictable (Ralph, Gray & Schoelles 2009), these fixation spikes began before the direction appeared. The lack of these predictive saccades, as well as sustained failed saccades (movements to the direction box at the wrong time) suggests that subjects in the current study were not able to successfully predict the timing of the experiment.

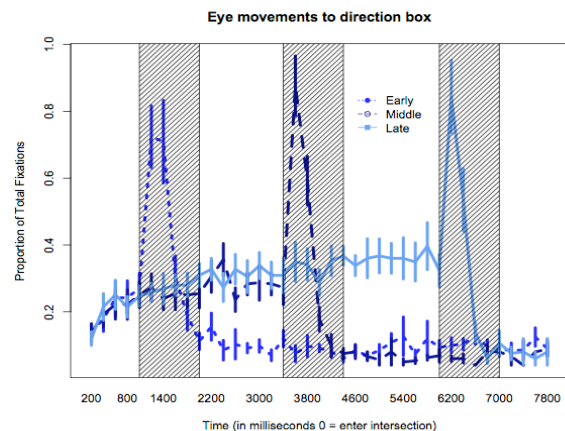


Figure 5: Mean proportion of eye fixations within a city block for transient visual subjects. Each color represents trials with common direction presentation times. Shaded areas represent time periods where the direction was available for each presentation time

The effect of direction presentation on jitter performance was a predictable result, since eye movements away from the arrow (e.g. toward the direction box) necessarily affect jitter performance especially when they occur multiple times throughout the city block.

More surprising was the clear difference between conditions of jitter scores at the beginning of a new city block. As Figure demonstrates, auditory subjects tended to start city blocks with the arrow closer to the center of the lane (i.e., lower jitter scores on the left-side of Figure 4). An ANOVA performed on the jitter values at the start of each "city block" resulted in a significant effect  $F(2,45)=5.09$   $p=.01$ . Since the initial arrow position is determined by how close the previous turn was made relative to the exact vertical center of the intersection, this suggests that auditory participants were better at timing their turns.

## Discussion

Two interesting results emerged from this study. The first was the effect of direction presentation and anticipation on the jitter scores of the transient visual condition. This performance degradation, which MRT would have interpreted as a cognitive visual processing resource conflict, appears in reality to be an instrumental resource conflict (e.g. the busy eyeball hypothesis).

The second interesting result is the better accuracy in turn timing by the auditory subjects (resulting in lower initial jitter scores) compared to both visual conditions. Perhaps the lower workload created by audio directions allowed these subjects to spend more time preparing for the turn process. They may have been able to retrieve the correct turn from memory ahead of time, and use eye movements to the approaching intersection to better time the turn. Transient visual subjects on the other hand may have had to spend so much cognitive activity on preparing for and anticipating the direction instruction that they neglected to properly prepare for the upcoming intersection.

Another explanation is that the ability to maintain the list of instructions is affected by the modality in which the instructions were presented. Perhaps the memory task requires less cognitive effort (rehearsal, updating) for auditory stimuli than visual stimuli. There is no way to determine this comparison in the current study as it lacked a control "no memory" comparison condition for auditory presentations.

Ongoing research seeks to answer some of the questions posed by this research. Presenting the direction inside the arrow instead of in the direction box will determine the relative effect of visual processing vs eye movements. Including 1-Back conditions (where subjects turn the direction of the most recent instruction) will allow us to determine the effect that memory use may have on jitter performance, turn timing and turn accuracy.

## Summary and Conclusions

Cognitive workload is a simplistic term that does not recognize the complexity of human behavior. MRT's simplistic representation of resource allocation is not sufficient to model the complexity of interactive behavior. Complex

resource interactions cause difficult to predict behavior in even relatively simple tasks. For example, we know of no theory that would have predicted that instruction modality would lead to better timed turning in the auditory condition.

Research on workload has focused on squeezing the balloon at one end. It is time we focused on predicting where it pops out. In prior work, Gray et al (2006) showed that delaying information access by as little as 200 ms shifted data acquisition and performance strategies from being "interaction intensive" to being more "memory intensive." In the current study we have demonstrated that changes in workload can affect behavior in profound ways. Understanding the relationship between resource allocation and strategy allocation is the key to understanding, modeling, and predicting cognitive workload.

## Acknowledgements

Research reported in this paper has been supported, in part, by grant N000140710033 to Wayne Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer. Brian Donlan 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.
- 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.
- Gevens, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex*, 10, 829-839.
- Gevens, 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., 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.
- 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.
- 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., & Gevens, A. (1998). Dynamic cortical networks of verbal and spatial working memory: effects of memory load and task practice. *Cereb. Cortex*, 8(7), 563-574.
- Ralph, J., Gray, W., & Schoelles, M. (2009) The functional resource hypothesis as a basis for understanding cognitive workload in interactive behavior. *Proceedings of the 53<sup>rd</sup> Annual Meeting of the Human Factors and Ergonomics Society* Santa Monica, CA: Human Factors and Ergonomics Society
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, 2(3), 159-177.
- Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psycho physiological features in an air traffic control task, *Human Factors*, 45,381-385
- Wilson, G.F. & Russell, C.A. (2003). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors*, 45, 635-645