Development of a Neuroergonomic Application to Evaluate Arousal

Daniel Gartenberg, Ryan McGarry, Dustin Pfannenstiel, Dean Cisler,
Tyler Shaw, Raja Parasuraman
George Mason University
Fairfax, VI 22031, USA

ABSTRACT

We developed and tested a neuroergonomic smartphone application called Mind Metrics that can be used to evaluate vigilance and working memory under naturalistic conditions. The application met a requirement to the field of neuroergonomics because the cognitive tasks were made for a smartphone platform, allowing the ability to make predictions about the neural processes that impact human performance during naturalistic work related activities (i.e. ubiquitous computing). However, if naturalistic tasks are to be developed that are sensitive to cognitive processes, these tasks must be tested and evaluated for validity by comparing performance to data obtained in controlled laboratory environments. In this study, we developed tasks that measure working memory and vigilance, two processes that are well known to affect human performance at work. We then tested participants on these tasks using both a smartphone and a desktop computer platform. The tasks we used to measure vigilance included a vigilance task called the Psychomotor Vigilance Task (PVT) and a vigilance task called the Spatial Discrimination Vigilance Task (SDVT). To measure working memory, we used a Color N-back Task (CNB) and a Spatial N-back Task (SNB). Using a mixed group design, participants were assigned to a desktop or smartphone condition and completed all four tasks. As predicted, there was a vigilance decrement for both the PVT and the SDVT, which was demonstrated by an overall slowing of responses as the tasks progressed. This decrement occurred for both the smartphone and the desktop tasks. Another interesting finding related to improvement over time for the N-back tasks-- participants performed faster on the n-back tasks as the task progressed. This indicates that task learning is an important factor to consider when developing neuroergonomic tasks aimed at detecting cognitive functioning in the wild. In previous research it was found that increased resource demands exacerbate the vigilance decrement. These findings suggest that learning can play a role in attenuating the vigilance decrement effect in tasks with high resource demands. If vigilance tasks developed on the smartphone can be administered in naturalistic environments this platform will provide a method of easy-to-obtain samples of repeated task performance, thereby reducing the impact of learning effects that can mask the vigilance decrement. The possible implications of this research are a more sensitive measure of the vigilance decrement for detecting vigilance in the wild.

Keywords: Neuroergonomics, Brain Arousal, Vigilance, Sleep, Health Care

INTRODUCTION

Through the merger of neuroscience, human factors psychology, and engineering, neuroergonomics aims to optimize mental functioning during cognitive and physical work (Parasuraman, 2003). The human brain's arousal system exerts an important influence on performance in work environments for both simple and complex tasks (Balkin, Rupp, Picchioni, et al., 2008). Lower arousal is associated with increased rate of accidents (Stutts, Wilkins, Osberg, et al., 2003). There is therefore a need to develop a neuroergonomic application that can provide the worker and co-worker with feedback on the worker's current level of alertness (Rizzo, Robinson, & Vicki, 2007). Yet there are obstacles to obtaining data in real-world settings because these settings require the use of different measurement tools. Additionally, the worker is less willing to devote a large amount of daily time to participate than participants in conventional laboratory experiments. At the same time there is the potential to obtain repeated measures over a long period of time when utilizing naturalistic data collection techniques.

Smartphones can be used as research tools to easily collect data in naturalistic environments; however, it is unclear how findings generalize across smartphone and desktop platforms and whether the former provide similar data to that obtained under controlled laboratory conditions. The iPhone, in addition to operating under different processing speeds, utilizes a touch screen interface and software that samples at different rates than desktop software that is typically used in laboratory experiments. Therefore, tasks requiring precise timing measurements, such as simple reaction time tests, may not produce the same results across platforms.

When developing a neuroergonomic application for the detection of variations in alertness in the workplace it is important to choose a test that is both sensitive and has minimal time costs. There are many laboratory-based tests and questionnaires that have been used to assess arousal (Matthews, Davies, Westerman, et al., 2000). While laboratory-based tasks, such as the psychomotor vigilance task (PVT) (Dinges and Powell, 1985), are usually more sensitive than questionnaires at detecting alertness changes (Van Dongen, Maislin, Millington, et al., 2003), laboratory-based tasks frequently require long periods of at least 10 minutes to administer (Dinges and Powell, 1985).

The ideal measure of alertness is therefore a task-based measure that can be administered for a short period of time. Based on the resource theory of vigilance (Parasuraman, 1985; Warm, Parasuraman, Matthews, 2008), a vigilance task that requires more resources will be more sensitive to a decrement in performance than one that requires less cognitive resources. Since resources are depleted during low arousal states (Dinges, Orne, Whitehouse, et al., 1987), a high resource demanding vigilance task will be more sensitive to variations in worker arousal. The longitudinal and repeated data collected on the smartphone can be used to address the learning effects that are likely to occur in these more resource demanding tasks.

VIGILANCE

Vigilance tasks typically involve the detection of signals over a long period of time, that are intermittent, unpredictable, and infrequent. As vigilance tasks progress, performance steadily declines, and there is a marked steep decline at about 20 minutes (Davies & Parasuraman, 1982; Parasuraman, 1986; Boksem, Meijman, Lorist, 2005, Lim, Wi, Wang, et al., 2010), although some evidence suggests that the decrement function is complete after 5 minutes in especially demanding circumstances (e.g. Helton, Dember, Warm, & Matthews, 2000). A conceptual framework for understanding the vigilance decrement performance is provided by resource theory (Parasuraman & Davies, 1984; Parasuraman, Warm, & Dember, 1987; Warm & Dember, 1998; Warm, Parasuraman & Matthews, 2008) within which the decrement results from the depletion of information-processing assets that cannot be replenished during continuous task performance.

Vigilance tasks require discriminations that either involve holding a representation in working memory (successive vigilance task) and comparing that representation with the current image, or the information needed to make the discrimination is presented on the screen, and no or very little working memory is required (simultaneous vigilance tasks) (Parasuraman, 1979). Given that successive tasks require working memory and simultaneous tasks only require comparative judgments, performance usually degrades more steeply and quickly in the more demanding successive task condition (e.g. Caggiano & Parasuraman, 2004)

The PVT is the gold standard task used by sleep researchers to measure the arousal system and it has been found to be sensitive to all the components of sleep (Dinges, et al., 1987; Van Dongen, et al., 2001). The PVT is a simultaneous vigilance task that requires responding to a visual stimulus (Dinges, et al., 1985). It typically lasts for 10-minutes, but can also be effective in measuring the components of sleep after as little as 1 minute (Gartenberg & Parasuraman, 2011).

Recently, Shaw, Warm, Finomore, Tripp, Matthews, Ernest, & Parasuraman (2009) used transcranial doppler (TCD) to determine that the declining performance characteristic of vigilance tasks was due to a lack of resources and not from changes in systemic activity. Lim et al. (2010) found that when participants were sleep deprived, a 20-minute time-on-task could detect the vigilance decrement using the PVT, yet the task was sensitive to sleep deprivation in as little as 2 minutes. This suggests that the PVT is resource demanding, but that its sensitivity to detecting fatigue does not rely on the vigilance decrement function. In support of this, tasks that do not involve the decrement, such as brief cognitive tasks that require speed of cognitive throughput, working memory, and other aspects of attention have been found sensitive to sleep deprivation (Mallis, et al., 2008).

Based on the resource theory of vigilance, tasks that require more resources, such as successive tasks that tax working memory, will be more sensitive to detecting the vigilance decrement than the traditional PVT (Shaw, Warm, Finomore, Tripp, Matthews, Ernest, & Parasuraman, 2009). Perhaps they will also be more sensitive to detecting grogginess. Yet, when developing a neuroergonomic task that requires more resources, it is important to consider possible learning effects because learning can attenuate the task's sensitivity to detecting grogginess.

We developed and tested a simultaneous vigilance task, a successive vigilance task, and two N-back tasks on an iPhone platform and a Desktop platform in order to investigate potential issues involved in administering the Mind Metrics application in naturalistic environments.

MIND METRICS SMARTPHONE APP

We developed an neuroergonomic smartphone application that detects alertness level and tested the application on both an iPhone (with 320px by 480 px dimensions) and a desktop computer. The application included three types of tasks: vigilance tasks, memory tasks, and combined vigilance and memory tasks. After completing a task the participant received real-time feedback on their performance. User performance was saved to a viewable table where it could be exported.

[1] Vigilance, Working Memory, and Combined Tasks

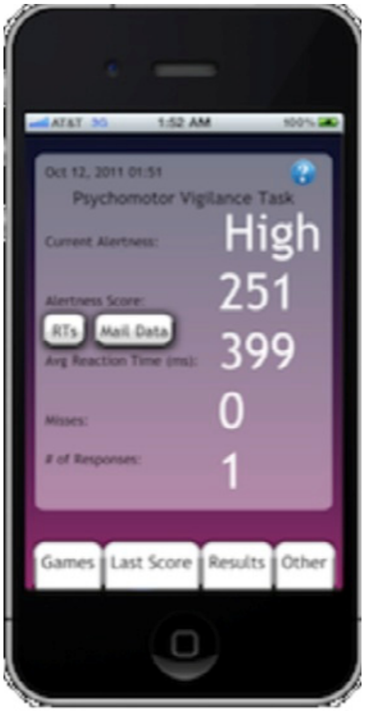
The PVT is a simultaneous vigilance task that requires participants to respond when a sun appears in the center of the screen. There is a stimulus onset window of 10,000 milliseconds, for which the sun randomly appears for 1,000 milliseconds. A total of 60 trials were run in the 10-minute version of this task.

The 10-minute spatial discrimination vigilance task (SDVT) is a successive vigilance task that involved discerning the distance between two stimuli. The stimuli consisted of a stationary cloud positioned in the center of the screen and a moon that appeared at one of two distances from the cloud (either 110 pixels or 130 pixels). The moon was presented close to the cloud 80% of the time (noise) and far from the cloud 20% of the time (signal). A response was only required when the moon was far from the cloud. Each trial lasted 4,300 milliseconds. The cloud remained present for the entire duration of each trial. The trials began with 1,800 milliseconds of inter-trial interval where only the cloud was presented. This was followed by the presentation of the moon stimulus, which lasted for 300 milliseconds. After the presentation of the stimulus the participant was given 1,900 milliseconds to respond. Feedback was then presented on whether the answer was correct or incorrect. The feedback lasted for 200 milliseconds.

The 10-minute color n-back (CNB) and spatial n-back (SNB) were 2-back working memory tasks. Both tasks required participants to determine if a set of lightning bolts was the same or different from a set of lightning bolts that occurred two trials previously. Each trial lasted for 4,300 milliseconds. Each trial began with 2,100 milliseconds interval where the cloud and the lightning bolts were present and the participant had the opportunity to respond. The feedback was then presented where the answer was either or incorrect. This feedback lasted for 200 milliseconds. There was an inter-trial interval of 1900 milliseconds where only the clouds were presented. The participant either indicated that the current stimulus was the same as the lightning pair that appeared two trials before or different from the lightning pair that appeared two trials before.

The CNB and SNB differed based on the modality of the stimuli. The CNB involved holding a color representation in working memory, while the SNB required holding a spatial representation in working memory. In the CNB two lightning bolts were presented that were different colors. In the SNB two lightning bolts were presented that were oriented in different spatial locations.

The combined vigilance and working memory task had an identical stimulus onset as the CNB and SNB, with the exception that the SDVT was administered in conjunction with the working memory tasks. This task was designed to be a task with a high task load.



Oct 12, 2011 01:51
Psychomotor Vigilance Task
Current Alertness: High
Alertness Score: 251
Avg Reaction Time (ms): 399
Misses: 0
# of Responses: 1

Figure 1. Mind Metrics feedback screen of the PVT.

[2] Feedback

Figure 1 illustrates the feedback that participants receive after completion of the task. Participants get information on the number of trials completed, their average accuracy, reaction time, and a score that combines accuracy and reaction time. Based on other scores that the participant received on the task, feedback is provided on alertness. This provides users with a real time measure of their alertness.

METHOD

[1] PARTICIPANTS

48 George Mason University students voluntarily participated for course credit. Each participant had normal or corrected vision. The sample consisted of 26 men and 22 women. The average age of participants was 20.25 years with a standard deviation of 3.41 years. Two participant's data were eliminated in the iPhone condition due to loss of Internet connection during the experiment.

[2] Design and Procedure

The experiment was a mixed design with device between groups and task within groups. The experiment was conducted using a desktop computer running E-Prime software and a 3rd generation iPod touch using an application running the iOS 4.0 SDK. The desktop and the iPod touch were programmed with four tasks where all participants experienced each task: a Psychomotor Vigilance Task (PVT), a Spatial Discrimination Task (SDVT), a Color N-Back (CNB), and a Spatial N-Back (SNB). Participants were assigned to platform conditions using Latin-squared randomized. The desktop-platform tasks and the iPod-platform tasks were identical, with the exception that in the desktop-based task participants responded with a button-press.

Each task began with instructions. The instructions were followed by a 1-minute practice session. The practice continued until criteria were met, where the criteria for each task differed due to different likelihoods of responding correctly. The PVT has very little practice effect the PVT due to the simplicity of the task, making it not necessary to train to criteria for this task. For the SDVT, if the participant never gave a response, they were able to obtain a score of 80%. Therefore, the criterion for the SDVT was set to 90%. For the N-Back tasks participants had a 50% chance of responding correctly. The criteria for the N-Back tasks were thus set to 80% correct.

After criterion was met, participants were reminded to respond as quickly and accurately as possible and to respond using their index finger with the finger hovering above the response interface. The 10-minute task then began.

Upon completion of each task, the perceived workload of the task was measured by administering the NASA-TLX (Hart & Staveland, 1988). Once the NASA-TLX was finished, the participant was given a short 5-minute break.

[3] Measures

Reaction time and accuracy were used as measured. As is customary with analysis of vigilance tasks, each 10-minute task was divided into 5 blocks of 2-minutes each. This enabled for the detection of changes in performance as the task progressed.

RESULTS AND DISCUSSION

The vigilance decrement was calculated by measuring reaction time in 2-minute block intervals, as the task progressed. For the PVT, this meant that there were 12 trials per interval. For the SDVT, CNB, and SNB there were 24 trials per interval.

For the PVT, a mixed ANOVA was run with block as a within groups factor and device type as a between groups factor. There was a main effect of block on reaction time, where participants performed worse as the task progressed, $F(4, 172) = 9.93$, $p < .05$. There was a main effect of device where participants were slower on the PVT when using an iPhone ($M = 489.76$ ms, $SD = 79.29$ ms) than when using a desktop ($M = 348.67$, $SD = 46.75$), $F(1, 43) = 54.49$, $p < .05$ (see Figure 2). There was no interaction between block and device condition, $F(4, 172) = 9.93$, $p < .05$. The finding that the overall reaction time was slower for the iPhone condition than the desktop condition suggests that the iPhone platform samples reaction times at a slower rate than a desktop computer running E-Prime software. However, the main effect of interval time and the lack of an interaction suggested that the simultaneous vigilance tasks were sensitive to the vigilance decrement despite these differences in the sample rates between devices.

For the SDVT, there was only reaction time data collected for the desktop condition. A simple within groups ANOVA was conducted on the 2-minute block intervals. As expected, over time participants performed worse on the task, $F(4, 84) = 2.73$, $p < .05$ (see Figure 2). This suggested that, successive vigilance tasks were sensitive to the vigilance decrement after a 10 minute task period.

It was important to determine if the two types of n-back tasks were equated on difficulty because these tasks are used for the combined vigilance and working memory task. Reaction time data was collected for the desktop condition and accuracy data was collected for both the desktop and iPhone condition. There was no difference in reaction time between the CNB ($M = 489.76$ ms, $SD = 79.29$ ms) and SNB ($M = 489.76$ ms, $SD = 79.29$ ms), $t(22) = 0.62$, $p = .54$. There was no difference in accuracy between the CNB ($M = 83.41\%$, $SD = 11.53\%$) and SNB ($M = 83.02\%$, $SD = 10.04\%$), $t(41) = 0.34$, $p = .74$. Since these tasks were equated for difficulty, this suggests that any differences in performance for the combined vigilance and working memory task should be related to the modalities of the n-backs (i.e. spatial vs color).

A simple within groups ANOVA was conducted on block. For the CNB there was no difference in how participants performed over time, $F(4, 88) = 0.70$, $p = .59$ (see Figure 2), but for the SNB participants responded more quickly over time, $F(4, 88) = 3.67$, $p < .05$ (see Figure 2). This suggested that overall difficult was similar between the two N-Back tasks. While there was a general trend of faster response times, this effect was only found for the SNB.

A global measure of perceived workload was determined using the NASA-TLX and a mixed ANOVA with task as the within groups factor and device as the between groups factor. There was no difference in perceived workload between the iPhone conditions and the Desktop conditions, $F(1, 45) = 0.79$, $p = .38$. Perceived workload differed based on task condition, $F(1, 45) = 70.41$, $p < .05$. Post-hoc tests were conducted using the Benjamini Hochberg correction method. All groups were significantly different from all other groups ($p < .05$), with the exception of the SNB and the CNB ($p = .33$). Participants rated the CNB (*M* = 66.94, *SD* = 16.23) and SNB (*M* = 65.21, *SD* = 15.85) as more difficult than the SDVT (*M* = 57.70, *SD* = 19.70) and the PVT (*M* = 42.82, *SD* = 24.55). There was no interaction between task condition and device, *F*(1, 45) = 1.94, $p = .17$. This suggested that the participant perceived the n-backs to be the most difficult, followed by the SDVT, and the PVT.

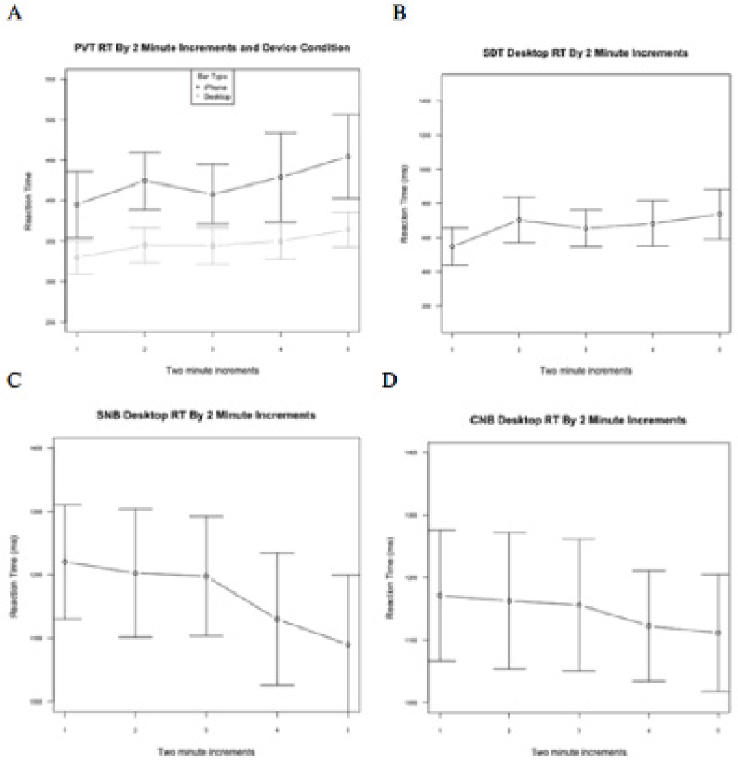


Figure 2. Graph of RT broken into 2 minute blocks. A = PVT, B = SDVT, C = SNB, D = CNB.

GENERAL DISCUSSION

We developed a neuroergonomic smartphone application called Mind Metrics, which provides people with real-time measures of their arousal state. The application includes a simultaneous vigilance task, successive vigilance task, working memory tasks, and combined vigilance and working memory tasks. Users of the application can set the duration and difficulty of the tasks, get feedback based on their own unique individual performance, and save and export their data.

In this experiment, the PVT, SDVT, CNB, and SNB were tested on both desktop and iPhone devices. Both devices detected the vigilance decrement, but the iPhone registered slower reaction times than the desktop. Possible explanations for this include slowing of response due to the touch interface and the iPhone software more slowly registering touch presses than a desktop computer. Nonetheless, the iPhone was used to measure the vigilance decrement, which will help foster naturalistic data collection, a requirement of neuroergonomics.

Another requirement of neuroergonomics measurement tools are non-invasiveness to a worker's daily routine. A vigilance task that incurs more resource demands on the user improves the time sensitivity of these tasks to measure the vigilance decrement (Warm, & Dember, 1998). As a result, increasing resource demands may be more sensitive to changes in an individual's arousal system.

We discovered that memory tasks such as the N-back, have higher perceived resource demands than the vigilance tasks administered in this study. However, no decrement in performance was found in the n-back tasks. The reason for this may be that these tasks do not have the characteristics of a vigilance task and that learning plays a larger role in these tasks. Administering tasks repeatedly could reduce these learning effects.

The neuroergonomic smartphone application developed in this paper was still in its testing phase. A limitation of this study was that the tasks were not administered repeatedly. In future research, the application will be applied to naturalistic environments and administered repeatedly. This will enable for real-time detection of the worker's arousal system, which can be used to prevent accidents and the negative consequences of accidents.

REFERENCES

Baehr, E., Revelle, W., & Eastman, C. (2000). Individual differences in the phase and amplitude of the human circadian temperature rhythm: with an emphasis on morningness-eveningness. *Journal of Sleep Research, 9,* 117-127.

Balkin, T., Rupp, T., Picchioni, D., & Wesensten, N. (2008). Sleep Loss and Sleepiness. *Chest, 134,* 653-660.

Boksem, M. A., Meijman, T. F., & Lorist, M.M. (2005). Effects of mental fatigue on attention: an ERP study. *Brain Res. Cogn. Brain Res. 25*, 107-116.

Bonnet, M. H. (1991). The Effect of Varying Prophylactic Naps on Performance, Alertness and Mood throughout a 52-Hour Continuous Operation. *Sleep, 14*, 4, 307-315.

Caggiano, D., & Parasuraman, R. (2004). The role of memory representation in the vigilance decrement. *Psychonomic Bulletin and Review, 11,* 5, 932-937.

Davies, D. & Parasuraman, R. (1982).*The Psychology of Vigilance.* London:Academic Press.

Dinges, D. F., Orne, M. T., Whitehouse, W. G., & Orne E. C. (1987). Temporal placement of a nap for alertness: contributions of circadian phase and prior wakefulness. *Sleep, 10*, 313–329.

Dinges, D. F., & Powell, J.W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav Res Meth Instr Comp, 17*, 652–655.

Gartenberg, D. & Parasuraman, R. (2010). Understanding Brain Arousal and Sleep Quality Using a Neuroergonomic Smart Phone Application. In Marek, T., Karwowski, W., &

Rice, V. (Eds.), Advances in Understanding Human Performance, *3rd International Conference on Applied Human Factors and Ergonomics* (pp. 210-220).

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload. (pp. 139-183). Oxford, UK: North-Holland.

Helton, W., Dember, W., Warm, J., & Matthews, G. (2000). Optimism, pessimism, and false failure feedback: Effects on vigilance performance. *Current Psychology, 18*, 311-325.

Lim, J., Wi, W., Wang, J., Detre, J. A., Dinges, D. F., & Rao, H. (2010). Imaging brain fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect. *Neuroimaging, 49*, 3426-3435.

Nuechterlein, K., Parasuraman, R., & Jiang, Q. (1983). Visual sustained attention: Image degradation produces rapid sensitivity decrement over time. *Science, 220*, 327-329.

Mackworth, J. F. (1968). Vigilance, arousal, and habituation. *Psychol. Rev., 75*, 308-322.

Mallis, M., Banks, S., & Dinges, D. Sleep and circadian control of neurobehavioral functions. Ed. Parasuraman, R., & Rizzo, M. *Neuroergonomics: The Brain at Work.* New York: Oxford Univsersity Press, 2007.

Matthews, G., Davies, D. Westerman, S. J., & Stammers, R. B. (2000). *Human performance: cognition, stress, and individual differences*. Philadelphia: Taylor and Francis.

Maquet, P. (2001). The role of sleep in learning and memory. *Science, 294*, 1048–1052.

Mori, C., Bootzin, R., Buysse, D., Edinger, J., Espie, C., & Lichstein, K. (2006). Psychological and Behavioral Treatment of Insomnia: Update of the Recent Evidence (1998-2004). *Sleep, 29,* 11.

Parasuraman, R. (1985). Sustained attention: A multifactorial approach. In M. I. Posner & O. S. Marin (Eds.) *Attention and Performance XI.* (pp. 493-511). Hillsdale, New Jersey: Erlbaum Associates.

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance. Vol. 2: Cognitive processes and performance* (pp. 43.1-43.39). New York: Wiley.

Parasuraman, R., (2003). Neuroergonomics: research and practice. *Theor. Issues. Ergon. Sci. 4*, 5-20.

Rizzo, M., Robinson, S., & Neale, V. The Brain in the Wild: Tracking Human Behavior in Naturalistic Settings. Ed. Parasuraman, R., & Rizzo, M. *Neuroergonomics: The Brain at Work.* New York: Oxford University Press, 2007.

Shaw, T. H., Warm, J. S., Finomore, V., Tripp, L., Matthews, G., Ernest, W., & Parasuraman, R. (2009). Effects of sensory modality on cerebral blood flow velocity during vigilance. *Neuroscience Letters*, *461*, 207-211.

Stutts, J. C., Wilkins, J. W., Osberg, S. J., & Vaughn, B. V. (2003). Driver risk factors for sleep-related crashes. *Accident Analysis and Prevention, 35,* 321-331.

Van Dongen, H. P., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep, 26,* 117-126.

Warm, J.S., Parasuraman, R., & Mathews, G. (2008). Vigilance requires hard mental work and is stressful. *Hum. Factors, 50*, 433-441.

Warm, J. S., & Dember, W. N. (1998). Tests of vigilance taxonomy. In R. R. Hoffman, M. R. Sherrick, & J. S. Warm (Eds.), Viewing psychology as a whole (pp. 87-112). Washington, DC: American Psychological Association.