Development and Evaluation of a Neuroergonomic Smart Phone Application to Assess Vigilance and Arousal

Daniel Gartenberg, Ryan McGarry, Dean Cisler, Tyler Shaw, Raja Parasuraman

George Mason University
Fairfax, VI 22031, USA

ABSTRACT

We developed and tested a neuroergonomic smart phone application called Mind Metrics, where the goal was to evaluate vigilance and working memory capacity under naturalistic conditions. Naturalistic data collection meets a requirement of neuroergonomics practice because such data can make predictions about human performance during work related activities. Yet naturalistic data need to be validated against data obtained in controlled laboratory environments. Accordingly, we tested participants on the same cognitive tasks using both a smart phone and a desktop computer. Tasks included a psychomotor vigilance task (PVT), spatial discrimination vigilance task (SDVT), and two working memory tasks, a color n-back task (CNB), and spatial n-back task (SNB). Vigilance decrements were detected for both the simultaneous vigilance task (PVT) and the successive vigilance task (SDVT). Both devices were sensitive to the detection of the vigilance decrement. The results show that the naturalistic platform of the smart phone is sensitive to detecting vigilance changes in the workplace. Testing of the application revealed that task learning was an important factor to consider when detecting the vigilance decrement. Performance on the n-back tasks improved over time, despite the finding that participants perceived the working memory tasks as more difficult than the other tasks. While previous research has shown that increased resource demands exacerbates the vigilance decrement, this suggests that task learning moderates this effect. The smart phone is offered as a tool that can be used to address these learning effects.
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 (Slutts, Wilkins, Osberg, et al., 2003) that incur high costs and casualties (cite). 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.

Smart phones can be used as research tools to easily collect data in naturalistic environments; however, it is unclear how findings generalize across smart phone 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 like E-Prime that is typically used in laboratory experiments. Therefore, tasks sensitive to precise timing measurements, such as simple or choice 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 (cite).

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 the vigilance decrement than one that requires less cognitive resources. Since resources are depleted during low arousal states (cite), a vigilance task that is more sensitive to the vigilance decrement will also be more sensitive to variations in worker arousal.

A neuroergonomic application that measures arousal cannot simply collect
data; it must also analyze the results and display feedback designed to improve mental and physical functioning. Since naturalistic data collection on the smart phone enables longitudinal and repeated data collection, such feedback can be tailored to the individual. Additionally, the longitudinal and repeated data collected on the smart phone can be used to address the learning effects that can occur during performance on cognitive tasks.

**VIGILANCE**

Vigilance tasks typically involve the detection of signals over a long period of time. The signals are intermittent, unpredictable, and infrequent. The discrimination either involves 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 working memory is required (simultaneous vigilance tasks) (Parasuraman, 1979). 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). The steep decline is known as the vigilance decrement. The vigilance decrement is characterized by a right hemispheric system involved in functional control and that is independent of modality (Shaw, Warm, Finomore, Tripp, Matthews, Ernest, Parasuraman, 2009), where this lateralization is moderated by task difficulty (find cite Parasuraman Neuropsychologia 2010).

The PVT is the gold standard task used by sleep researchers to measure the arousal system (cite). Dinges et al. (1985) developed the PVT after modifying Wilkinson’s (1970) auditory vigilance task. The PVT is a simultaneous vigilance task that requires responding to a visual stimulus. It typically lasts for 10-minutes, but can also be effective in measuring the components of sleep after 5-minutes (cite), or even 3 minutes (cite Dinges paper, 2011). The PVT was found to be sensitive to all the components of sleep (Dinges, Orne, Whitehouse, et al., 1987; Van Dongen, et al., 2001). Neuroimaging and subjective ratings indicate that vigilance tasks can be characterized by high workload where cognitive resources are depleted in a time-on-task driven manner (Warm, Parasuraman, Mathews, 2008). fMRI studies on the neural basis of the vigilance decrement show that the vigilance decrement activate a right fronto-parietal attentional network that lateralized to the basal ganglia and sensorimotor cortices (Lim, Wi, Wang, et al., 2010). This activation pattern is typically found after about 20 minutes.

The 20-minute time-on-task required to detect the vigilance decrement using the PVT indicates that the traditional 10-minute PVT task is too short to measure the steep decline in performance that is characterized as the vigilance decrement. This suggests that the sensitivity of the PVT to the components of sleep does not rely on the decrement. In support of this, performance on brief cognitive tasks that require speed of cognitive throughput, working memory, and other aspects of attention have been found to be sensitive to sleep deprivation (Mallis, et al., 2008), and these tasks do not necessarily involve the vigilance decrement.
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 the vigilance decrement (Caggiano & Parasuraman, 2004). Tasks that are more sensitive to the decrement may be more sensitive to detecting the components of an individual’s arousal system. A requirement of a neuroergonomic arousal detection application is that it is sensitive to the components of arousal with minimal time on task. However, when developing a neuroergonomic arousal detection task that requires more resources, it is important to consider possible learning effects because learning can result in improved performance, as opposed to a decrement.

We developed and tested a simultaneous vigilance task, successive vigilance task, and two working memory tasks in order to determine how sensitive these tasks were to the vigilance decrement. Memory tasks have high resource demands, yet demonstrating learning effects (cite). A neuroergonomic arousal detection system can overcome learning effects due to the ability to repeatedly administer the task, enabling for the development of a task that requires more cognitive resources, and is therefore more sensitive to the components of sleep after a shorter administration of the task.

NEUROERGONOMIC SMART PHONE APP

Technology to detect drowsiness in real-world environments must be unobtrusive to the user and able to calculate drowsiness in real-time (Mallis, et al., 2008). We developed a smart phone application that met this requirement and tested the application on both an iPhone and a desktop computer. The application includes three types of tasks: vigilance tasks, memory tasks, and combined vigilance and memory tasks. After completing a task the participant receives real-time feedback on their alertness. Their alertness for every time the task was completed is then saved to a table, where it can be exported. Detailed depictions of the application can be found at www.proactivelife.org.

[1] Vigilance 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 (Figure 1).
The SDVT is a successive vigilance task that involved discerning the distance between two stimuli. The stimuli consisted of a stationary cloud that is positioned in the center of the screen and a moon that appears at one of two distances from the cloud (either 110 pixels or 130 pixels). As is characteristic of vigilance tasks, the SDVT involved responding to a stimulus irregularly and infrequently. 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 1,800 milliseconds to respond. Feedback was then presented on whether the answer was correct (i.e. +1 flashed on the screen) or incorrect (i.e. 0 flashed on the screen). This feedback lasted for 400 milliseconds. There were a total of 140 trials in the 10-minute task.

Figure 2. SDVT.


The CNB and 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 1,800 milliseconds inter-trial interval where only the cloud was presented. This was followed by the presentation of the moon stimulus for 2,100 milliseconds. A response was required during every presentation of the stimulus, where 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.
Feedback was then presented on whether the answer was correct (i.e. +1 flashed on the screen) or incorrect (i.e. 0 flashed on the screen). This feedback lasted for 400 milliseconds. There were a total of 140 trials in the 10 minute task.

The CNB and SNB differed based on 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.

[3] Combined Vigilance and Working Memory Tasks

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 modeled based on the study of Caggiano and Parasuraman (2004) and was designed to be a task with a high resource load. Provided that resource theory is correct and that there is minimal learning in this task, this task may be more sensitive to the components of sleep than the PVT.


Figure 4 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. If the participant’s score is in the bottom 33rd percentile, alertness is low. If it is above the 33rd percentile, but below the 66th percentile, it is deemed medium. If it is above the 66th percentile it is high. This provides users with a real time measure of their own unique level of alertness. Additionally, this measure can then be exported to co-workers for evaluation.
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. 24 participants were assigned to the smart phone condition and 24 participants were assigned to the desktop condition. Two participant’s data were eliminated in the iPhone condition due to loss of Internet connection during the experiment.


The experiment lasted about 1.5 hours. 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). The order for which the tasks were presented and which device the participant was assigned to use were Latin-squared randomized.

The desktop-based tasks and the iPod-touch tasks were identical, with the exception that in the desktop-based tasks participants responded using a button-press box, while in the iPod touch tasks participants were required to respond by using the touch-screen interface. Device was manipulated between groups, where some participants were assigned to perform the tasks on a Desktop computer and others were assigned to perform the tasks on a 3G iPod touch. All participants began the experiment by filling out a preliminary questionnaire that included the Epworth Sleepiness Scale, sleep amount from the previous night, use of stimulants such as caffeine, and familiarity with using touch screen and iPhone technologies. Participants then began one of the four tasks. Each task began with instructions on how to complete the task.

The instructions were followed by a 1-minute practice session that was performed until the participant reached criteria. Criteria for each task differed because each task had a different likelihood of responding correctly. There is almost no practice effect the PVT (cite) 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 criteria for the SDVT was set to 90%. For the N-Back tasks participants had a 50% chance of responding correctly. Criteria for the N-Backs was 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. Once the NASA-TLX was finished, the participant was given a short 5-minute break before moving on to the next task.


The participant’s reaction time and accuracy were measured. As is customary with detecting the vigilance decrement, each 10-minute task was divided into 5 blocks, 2-minutes each. This enabled for the detection of changes in reaction time and accuracy as the task progressed.

RESULTS AND DISCUSSION

Sleepiness, caffeine use, cigarette use, and hours slept were not different across the two device conditions. For ratings on the Epsworth sleepiness scale, participants in the iPhone condition ($M = 7.78, SD = 2.50$) did not rate their sleepiness as different from participants assigned to the Desktop condition ($M = 7.70, SD = 3.30$), $t(44) = -0.10, p = .92$. For amount of caffeinated beverages, participants in the iPhone condition ($M = 0.13, SD = .35$) did not consume a different amount than participants in the Desktop condition ($M = 0.57, SD = 1.20$), $t(44) = 0.11, p = .11$. For amount of cigarettes smoked, participants in the iPhone condition ($M = 0.00, SD = 0.00$) did not smoke a different amount than participants in the Desktop condition ($M = 0.43, SD = 0.21$), $t(44) = 0.98, p = .33$. The amounts of sleep were also similar between participants in the iPhone ($M = 5.87, SD = 2.58$) and Desktop conditions ($M = 5.21, SD = 2.91$), $t(44) = 0.87, p = .43$.

The amounts of sleep were also similar between participants in the iPhone and Desktop conditions ($M = 5.87, SD = 2.58$), $t(44) = 0.87, p = .43$. 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 ($M = 348.67$ ms, $SD = 46.75$), $F(1, 43) = 54.49, p < .05$ (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 time.

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 \((M = 489.76 \text{ ms}, SD = 79.29 \text{ ms})\) and SNB \((M = 489.76 \text{ ms}, SD = 79.29 \text{ ms})\), \( t(22) = 0.62, p = .54 \). 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, \( F(4, 88) = 3.67, p < .05 \) (see Figure 2). This suggested that overall difficulty 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.

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's 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.
GENERAL DISCUSSION

We developed a neuroergonomic smart phone 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 simultaneous vigilance task, the successive vigilance task, the spatial working memory task, and the color working memory task were tested on both desktop and iPhone devices. The iPhone registered slower reaction times than the desktop; however, for both vigilance tasks, a vigilance decrement was detected on both devices. This suggests that the iPhone can be 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-invasive, or incur minimal time costs on the worker’s daily routine. A vigilance task that incurs more resource demands on the user improves the sensitive of these tasks to measure the vigilance decrement (Warm, & Dember, 1998). As a result, tasks with increased resource demands may also be more sensitive to detecting changes in an individual’s arousal system, thereby reducing the time required to administer the task.

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 affect these learning effects. Additionally, comparing overall performance between tasks may be enough to detect changes in the arousal system, preventing the need to detect a vigilance decrement. The smart phone is the ideal tool to enable for repeated administration, thereby promoting the development of a vigilance task that is more sensitive to detecting changes in the human arousal.

The neuroergonomic smart phone application developed in this paper was still in its testing phase. In future research, the application will be applied to naturalistic environments. 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


