2011

Vigilance Decrements in a Sustained Attention Task: Examination of a Mitigation Strategy

Guy A. French

Thomas R. Carretta

John K. Flach

VIGILANCE DECREMENTS IN A SUSTAINED ATTENTION TASK: EXAMINATION OF A MITIGATION STRATEGY

Guy A. French Thomas R. Carretta
Air Force Research Laboratory
Wright-Patterson AFB, OH 45433

John K. Flach
InfoSciTex
Wright-Patterson AFB, OH 45433

A study was conducted to examine the effectiveness of perceptual and cognitive intervention tasks on mitigating vigilance decrements commonly observed in sustained attention tasks. Sixteen participants were randomly assigned to one of two experimental intervention conditions (perceptual or cognitive). Half of the participants completed a 45-minute "No Intervention" Control trial first, followed by one of the Intervention trials, also 45 minutes. The other half completed one of the Intervention trials first, followed by the No intervention trial. Following each trial, participants completed the SSSQ and the NASA TLX. As expected, a general decrease in objective performance over time was observed. However, contrary to expectations and prior research (St. John & Risser, 2007), the experimental intervention tasks did not reduce decrements in target acquisition performance over time, nor did they reduce subjective workload. The authors discuss methodological differences between the St. John and Risser study and the current study that may have contributed to differences in the effectiveness of the vigilance mitigation interventions in the two studies.

The capability to capture video and/or still imagery through sensors on remotely piloted aircraft (RPA) and fixed sensors has increased dramatically in recent years. The use of RPAs to support military operations is becoming increasingly widespread. RPAs are particularly well suited for performing missions requiring persistence due to their unrivalled endurance and low visual, radar and acoustic signatures. Their ability to search for, loiter over, and track ground targets for extended periods of time make them highly valued assets. For many long endurance RPAs, the duration of missions is such that operator fatigue becomes a critical factor (Wilson, Russell, & Caldwell, 2006). Operator fatigue also plays a role in the ability of sensor operators/image analysts when the task is to monitor multiple video feeds from fixed sensors for suspicious activities and potential targets.

Operator fatigue has been addressed by applying work shifts, changing out the crew mid mission at scheduled intervals (typically 8 hours). However, vigilance decrements begin to affect human performance long before physical fatigue begins to affect operators. This is especially true for sensor operators and image analysts performing long missions. These missions require personnel to monitor sensor video imagery for long periods of time for targets and suspicious activity. The low rate of occurrence of such activities over long periods leads to performance decrements in terms of reduced target acquisition rates and longer response times. By building a better understanding of the factors that contribute to these vigilance decrements, we can identify and evaluate technologies designed to address and mitigate factors that contribute to vigilance decrements and augment human performance.

Perceptual vigilance (Davies & Parasuraman, 1982; Molloy & Parasuraman, 1996)) has been studied for several years, resulting in the identification and characterization of performance decrements in sustained attention tasks. Theorists have attempted to explain vigilance decrements as a function of arousal/motivation (Vroom, 1964; Yerkes & Dodson, 1908), workload/multiple resource theory (Wickens, 2002), and other factors. However few effective mitigation technologies have been implemented to address the issue (Schroeder, Touchstone, Stern, Stoliarov, & Thackray, 1994; St John & Risser, 2007).

St. John and Risser (2007) examined the utility of perceptual and cognitive interventions for mitigating the vigilance decrement. Their approach combined aspects of arousal and resource theories (Arousal-Resource model). They introduced an intervention in the form of a secondary task designed to draw upon resources separate from those required by the primary visual vigilance task. The interventions were an auditory alarm “ring tone” that required sensory perception only and two auditory cognitive tasks that required participants to mentally reorder strings of 3 or 4 spoken digits. They hypothesized that the two cognitive digit task interventions would arouse participants, replenish depleted resources, and re-engage them in the vigilance task. They also hypothesized that the cognitive tasks would be more effective than the simple alarm because they were more demanding and engaging. Participants performed a 45 minute laboratory vigilance task twice, once in a control condition without any intervention and once with one of the three interventions. In the intervention conditions, participants received the intervention whenever they missed a target. This intervention method served as a proxy for a closed-loop system in which operators would receive interventions whenever low attention was detected by psychophysiological measures, prior to an actual miss. All three interventions significantly reduced misses by approximately 30%. Participants who showed greater vigilance decrements in the baseline (no intervention) condition showed more improvement from all interventions. That is, more vulnerable participants benefited most. The cognitive interventions performed as well as, but no better than, the simple alarm. The cognitive tasks also interfered with target detection performance on occasions when the interventions occurred while a target was viewable. However, the alarm was rated as more frustrating and less appropriate than either of the cognitive intervention.

The current study examined the utility of St John and Risser’s (2007) Arousal-Resource approach for mitigating the effects of vigilance decrements in a task where a single operator was required to monitor sensor feeds from two fixed remote sources.

Methods

Participants

Participants were 16 civilian and military employees (13 males, 3 females) stationed at Wright-Patterson AFB, OH. They ranged in age from 21 to 44, with a mean of 27.75 years. All participants reported being in good health, with normal visual acuity after correction and no problems with peripheral vision or color blindness. Most reported some prior simulator experience (56%) and video game experience (81%). All participants were volunteers and no compensation was provided for their participation.

Measures

Several types of data were collected. These included objective measures of perceptual vigilance ability and target acquisition performance (hits, misses, false alarms), demographic/background data, and subjective measures of mood/stress and workload.

Objective Experimental Task Performance. Objective measures of the experimental task performance were hits, misses, and false alarms.

Perceptual Vigilance Task (PVT). In this task (Temple et al, 2000) participants monitored the presentation of 8- by 6-mm light grey capital letters consisting of ‘**O**’, ‘**D**’, and a backwards ‘**D**’ centered on a video display screen. The letters were constructed in 24-point type using an AvantGarde font and were exposed for 40 milliseconds against a visual mask that consisted of unfilled circles on a white background. The participants task was to use the mouse to indicate when the target letter ‘**O**’ was presented. Responses were scored as hits, misses, and false alarms.

Biographical Questionnaire. This questionnaire collected information in order to characterize the sample and assist in interpretation of participants’ performance on the target detection task. Items elicited information about participants’ sex, age, general health, wellbeing, previous experience with simulator-type environments, previous experience with video games, and whether they had vision correctable to 20/20 acuity and normal peripheral and color vision.

Short Stress State Questionnaire (SSSQ). The SSSQ (Helton, 2004) is an abbreviated version of the Dundee Stress State Questionnaire (DSSQ; Matthews et at., 2002). It consists of four subscales: mood state, motivation, thinking style, and thinking content. Scores for three factors are derived from the subscales: task engagement, distress, and worry. Additionally, one component of the SSSQ assesses participants' perceptions of their physical and mental workload. One version of the SSSQ is administered before task performance and the other is administered after task performance.

NASA Task Load Index (NASA TLX). The NASA TLX (Hart & Staveland, 1988) is a subjective workload assessment measure that allows users to evaluate their interactions with various human-machine systems. A paper-and-pencil version which is part of the SSSQ was used to assess operator workload during the experiment. Participants rated their subjective workload on a scale that ranged from 0 (low) to 100 (High) for each of 6 subscales: Mental, Physical, Temporal, Effort, Performance, and Frustration An overall workload score based on a weighted average of ratings on the 6 subscales also was computed.

Procedures

Participants remained seated throughout the experiment, except during breaks. The study began with a pre-briefing regarding research objectives, procedures and informed consent. This was followed by administration of the Demographic Data Questionnaire. Next, participants completed an abbreviated version of the Dundee Stress State Questionnaire (DSSQ) called the Short Stress State Questionnaire (SSSQ) to establish a baseline of their stress level. Next, the 12-minute Perceptual Vigilance Task (PVT) was administered to estimate participants' perceptual vigilance ability. Scores from the baseline PVT were correlated with objective measures of performance on the experimental task to examine relations between the tasks. Following the short vigilance task, participants completed the SSSQ to assess changes from their baseline (pre-vigilance task) level. They also completed the NASA TLX to assess subjective workload. There was a 10-minute break following completion of the PVT and questionnaires.

Next, participants were randomly assigned to one of the two experimental intervention conditions (perceptual or cognitive). Half of them completed the "No Intervention" trial first, followed by one of the Intervention trials. The other half completed one of the Intervention trials first, followed by the No intervention trial. Following each target acquisition block, participants completed the SSSQ and the NASA TLX.

Those doing the No Intervention trial first completed a 3-minute practice session during which the target was 3 times more frequent than during the actual experiment. Task performance feedback was provided (i.e., hits, misses, and false alarms) during the practice trial. After a short break, the 45-minute experimental trial followed during which no performance feedback was given. Following the experimental trial, participants completed the post-trial SSSQ and NASA TLX to assess their subjective stress and workload, then took a 10-minute break before beginning the second experimental session. The procedures were similar for the Intervention conditions. For these conditions, participants began with a 3-minute practice session during which the targets occurred 3 times more frequently than in the experimental session. During the practice session, participants received feedback on target acquisition performance (i.e., hits, misses, and false alarms) and the intervention occurred 6 times. After a short break, participants completed a 45-minute experimental trial during which no performance feedback was given. The post-test SSSQ and NASA TLX were completed following the experimental session to assess participants' subjective stress (mood state) and workload.

The experimental task was to monitor simulated video feeds from 2 remote fixed sensors positioned to monitor traffic intersections in an urban setting and to designate targets/suspicious behaviors as they were detected. Participants were instructed to designate targets and respond to interventions, depending on the experimental condition. Figures 1 illustrates the displayed imagery as viewed by study participants. Each screen (left and right) displayed information for one of the two remote sensors.

Figure 1. Displayed imagery as viewed by study participants.

Analyses

In order to examine trends over time, the Perceptual Vigilance Task (PVT) and Experimental Task (ET) each were divided into 3 equal time intervals and objective performance (hits, false alarms) scores were calculated for each interval. Repeated measures analyses of variance were used to examine trends in performance over the 3 intervals. Correlational analyses examined the relations between the PVT and ET measures. Pre-PVT SSSQ scores served as a baseline by which to evaluate changes in subjective mood state following each of the Experimental Tasks (Control and Intervention conditions) using related samples t-tests. NASA TLX means were examined to compare subjective workload levels for the Control and Intervention conditions of the Experimental Task using related samples t-tests. All analyses used a .05 Type I error rate.

Results

Perceptual Vigilance Task

As expected, over time there was a trend toward a lower target acquisition (hit) rate (67.75%, 61.84%, and 62.28%) and a higher number of false alarms (7.50, 7.81, and 7.93) on the Perceptual Vigilance Task (PVT). However, neither of these trends was statistically significant. Further, contrary to expectations, performance on the PVT was not related to performance on the experimental task. The strongest relations were between the number of false alarms on the PVT and the number of false alarms on the control ($r = .287$, $p = .14$) and intervention ($r = .384$, $p = .07$) tasks. Examination of the pre- and post-PVT SSSQ scales revealed an increase on Distress (6.38 vs. 14.31, $t(15) = 14.17$, $p < .001$), but no difference for Engagement (24.13 vs. 22.63, $t(15) = 1.33$, ns) or Worry (6.38 vs. 5.75, ($t(15) = 0.74$, ns). Mean post-PVT workload scores were elevated for Mental (85.6), Temporal (71.9), and Effort (80.6) scales relative to the scale midpoint of 50. Mean Overall workload was 62.5.

Experimental Task

Results on the ET were mixed. Comparisons between the Control condition and each of the intervention conditions were not statistically significant for either the number of hits or false alarms. There was a significant overall effect for time interval for both hits ($F(2, 30) = 8.33$, $p < .001$) and false alarms ($F(2,30) = 3.82$, $p < .05$). However, neither trend showed a consistent decline in performance as was expected. The number of hits declined from the first to the second time interval, but recovered somewhat in the third interval. See Figure 1. The number of false alarms decreased from the first to the second interval (i.e., performance improved), but increased in the third interval.

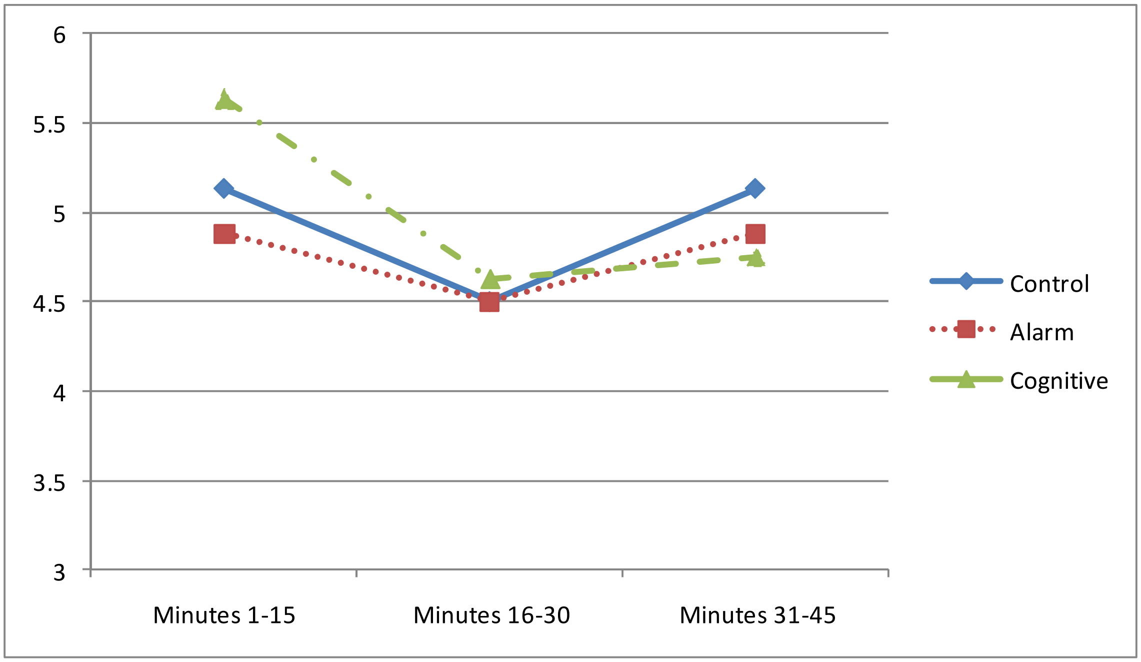

Figure 2. Number of Hits on the Experimental Task by Intervention Condition and Time Interval.

Examination of the SSSQ scores revealed mixed results. Using the pre-PVT scores as a baseline, SSSQ Engagement decreased after both the ET Control condition (24.13 vs. 20.25, $t(15) = 2.90$, $p < .01$) and the Intervention condition (24.13 vs. 20.38, $t(15) = 4.34$, $p < .001$). Worry increased following the Control condition (6.38 vs. 8.63, $t(15) = -2.12$, $p < .05$), but not the Intervention condition (6.38 vs. 7.63, $t(15) = -0.99$, ns.). There were no significant changes in Distress scores following the experimental tasks. Comparisons between the post-ET Control and Intervention conditions showed no significant effects.

Comparisons between the NASA TLX subjective workload scores for the post-PVT and each of the post-ET conditions revealed that subjective workload for most of the scales and for Overall workload was higher following the PVT (62.50) than after either of the ET tasks (Control: 62.50 vs. 34.27, $t(15) = 6.36$, $p < .001$; Intervention: 62.50 vs. 32.39, $t\ (15) = 7.39$, $p < .001$). The Control and Intervention tasks did not differ on Overall workload (34.27 vs. 31.92, $t(15) = 0.53$, ns).

Discussion

As expected, there appeared to be a vigilance decrement across the 45 minute task, with the larger decline from the first 15 minute period to the second 15 minute period. What was somewhat surprising was the lack of a difference between the baseline and intervention conditions. St. John and Risser (2007) reported fairly stable performance for each of the three intervention conditions. Methodological differences between the two studies may have contributed to differences in results. To begin, there was a major difference in the primary tasks in the two studies. In St. John and Risser, participants performed a sensor monitoring task that involved detecting a single critical signal repeatedly over 45 minutes. The event exposure duration was 400 ms and the event rate was 1 per 2 seconds with a critical signal (target) rate of 3 per minute. The critical signal was a truck icon that was slightly larger than the non-critical signal. All signals occurred in one of 6 fixed screen locations.

The primary task in the current study was a step closer to a common sensor operators task. Rather than a single critical signal occurring at fixed locations, there were 11 suspicious behaviors used as critical signals that could occur anywhere in the two sensor streams. Further, the non-critical signals were “daily life” activities in an urban scene continuously presented. Thus, the non-critical signals overlapped with the targets. Finally, the critical signal rate in the current study was 1 per 2.5 minutes across the two scenes with a median exposure time of 11.0 seconds. Despite these differences in event rate, exposure time, number of displays monitored, and observer uncertainty about spatial location of the critical signal, a vigilance decrement was observed. The difference between the studies occurred in the effectiveness of the interventions.

One possible explanation for the difference between the findings of the two studies is the intervention schedule. St. John and Risser implemented their interventions with simulated psychophysiological monitoring where the intervention was triggered by a missed target. In the present study, a simpler approach was taken by implementing a constrained randomized schedule. Such an approach, if effective, would eliminate the need for physiological monitoring to achieve performance benefits. Alas, the current study indicated that such hopes were unjustified. Whether the reason for different results lies in the differences in the nature of the experimental tasks, the intervention schedule, or some interaction of the two, it appears at this point that physiological monitoring may be necessary to achieve the performance benefits of a system based mitigation strategy.

References

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

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

Helton, W. S. (2004). Validation of a short stress state questionnaire. *Proceedings of the 48th Meeting of the Human Factors and Ergonomics Society.*

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., Grier, R., &Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion, 2*, 315-340.

Molloy R, & Parasuraman, R. (1996). Monitoring an automated system for a single failure: vigilance and task complexity effects. *Human Factors, 38*, ,311-322.

Schroeder, D. J., Touchstone, R. M., Stern, J, A., Stoliarov, N., & Thackray, R. (1994). *Maintaining vigilance on a simulated ATC monitoring task across repeated sessions*, DOT/FAA/AM-94/6. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration, Office of Aviation Medicine.

St. John, M., & Risser, M. R. (2007). Mitigating the vigilance decrement: Perceptual versus cognitive interventions. *Proceedings of the 4th International Conference on Augmented Cognition*, Baltimore, MD.

Temple, J. G., Warm. J. S., Dember, W. N., Jones, K. S., LeGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors*, *42*, 183-194

Vroom, V. H. (1964). *Work and motivation*. NY: Wiley.

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 150-177.

Yerkes, E. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482.