USAARL-TECH-FR--2020-043

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

Sustaining Performance and Vigilance During Extended UAS Operations

Amanda Kelley, Amanda Hayes, Colby Mathews, & Kyle Bernhardt

Notice

Qualified Requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Fort Belvoir, Virginia 22060. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of Address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
15-09-2020

2. REPORT TYPE
Final Report

3. DATES COVERED *(From - To)*
July 2018 - September 2020

4. TITLE AND SUBTITLE
Sustaining Performance and Vigilance During Extended UAS Operations

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kelley, A., Hayes, A., Mathews, C., & Bernhardt, K.

5d. PROJECT NUMBER
2019-09

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER
USAARL-TECH-FR--2020-043

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
504 Scott Street
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)
USAMRDC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
Goldbelt Frontier, LLC, Oak Ridge Institute for Science and Education

14. ABSTRACT

Current scheduling of unmanned aerial system (UAS) operators often allows for 8 hour shifts, where operators are typically exposed to low event rate tasks, thus leading to the occurrence of underload. While a long, rich history of vigilance research exists, few studies have examined the threshold at which performance decrements occur in UAS operators in operational settings and the utility of strategies to mitigate vigilance and performance decrements. This study evaluated the performance thresholds relative to time-on-task during a 4 hour simulated UAS mission. Additionally, this study evaluated the effectiveness of countermeasure strategies (secondary task and ambient lighting) on sustaining performance and vigilance during simulated UAS missions. Finally, the study demonstrated patterns of psychophysiological indicators of operator states (comparing high and low workload). Findings suggest that performance begins to decline after 15 minutes on task and plateaus by 45 minutes on task for both workload conditions. There was insufficient evidence to support any of the countermeasures, yet patterns in EEG are consistent with past findings regarding shifts in perceived workload.

15. SUBJECT TERMS
Unmanned aerial systems, UAS, Vigilance, Sustained attention, Workload, Underload, FVL

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLAS	UNCLAS	UNCLAS	SAR	28

19a. NAME OF RESPONSIBLE PERSON
Loraine St. Onge, PhD

19b. TELEPHONE NUMBER *(Include area code)*
334-255-6906

This page is intentionally blank.

Acknowledgements

The authors would like to express their gratitude to the following people for their contributions to this project: Juan Colon-Cruz, Will Irvin, SPC Mackie, SPC Rhea, Kevin Baugher, Kevin O’Brien, and Jon Vogl for their countless hours in data collection and programming expertise.

This research was supported in part by an appointment to the Postgraduate Research Program at the U.S. Army Aeromedical Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Army Medical Research and Development Command.

This page is intentionally blank.

Table of Contents

Page

Table of Contents (continued)

List of Figures

Introduction

During a typical mission, unmanned aerial system (UAS) operators, outside of and often remote to the physical location of the aircraft, experience lengthy periods of low workload not unlike that of a manned aircraft pilot. However, the UAS operator's experience is unique in that the individual is removed from the environment of the aircraft, and thus, receives a limited amount of sensory input (e.g., the operator does not sense vibrations from the aircraft or perceive orientation). Therefore, during periods of underload and diminished sensory input, operators may become distracted from the task. Specific UAS accidents have been linked to the lack of sensory inputs as well as vigilance decrements (Tvaryanas, Thompson, & Constable, 2005; Whitlock, 2014). A 2014 investigation into more than 400 U.S. Military UAS mishaps and accidents that occurred between 2001 and the article's publication emphasizes "inattention" as one of the top contributors of these mishaps and also gives credit to the lack of sensory input experienced by the pilots of such aircraft (Whitlock, 2014). Loffi, Wallace, Jacob, and Dunlap (2016) approximate that 92 UAS mishaps and/or accidents occur per year based on a 2016 Federal Aviation Administration report (Federal Aviation Administration, 2016). This study focused on the impact of extended periods of underload on performance during a target detection task in a UAS simulator. Additionally, this study evaluated two potential mitigation strategies to sustain performance. Finally, this study explored the patterns of psychophysiological measurements as indices of attentional state. The execution of this study was impacted by COVID-19 related delays that limited the sample size, the precision of the statistical analyses, and ultimately, the depth of conclusions.

Background

Present UAS operations rely heavily on automated systems, such that the operator frequently monitors the overall flight, with few manual inputs (Cummings, Mastracchio, Thornburg, & Mkrtchyan, 2013; Wohleber et al., 2019). While several advantages are associated with highly automated systems, such as reductions in personnel needed to operate a flight and operator workload, highly automated systems can also lead to boredom and loss of vigilance. Additionally, reductions in workload can lead to what is known as underload, which occurs when an operator experiences minimal stimulation through low task loadings, resulting in fewer cognitive resources available for unexpected increases in task demands (Young, Brookhuis, Wickens, & Hancock, 2015). An operator in a state of underload becomes more susceptible to performance errors when sudden changes in task demands occur, particularly when the increases occur during the lowest point of engagement with the task at hand (Boyer, Cummings, Spence, & Solovey, 2015). Alternatively, an operator may experience a loss of vigilance if exposed to high workload characterized by high task demands. In this case, the loss of vigilance occurs due to a depletion of the cognitive resources available for the task (Epling, Russell, & Helton, 2016). Thus, the vigilance decrement can occur when underload sets in, with resources allocated elsewhere and unable to be reallocated when a sudden increase in demands. The vigilance decrement can also occur when allocated resources are depleted from sustained high task demands. Current scheduling of UAS operators often allows for 8 hour shifts, where operators are typically exposed to low event rate tasks that can lead to the occurrence of underload. A long history of vigilance research exists, much of which has shown vigilance decrements to appear within 30 minutes of time-on-task (Mackworth, 1948; Warm, Matthews, & Finomore Jr, 2017) and in cases of high task demand, can appear in as little as 5 minutes (Helton et al., 2007). Studies of vigilance decrements in UAS operators have employed task durations varying from 12 minutes (Helton et al., 2007) to 120 minutes (Senoussi et al., 2017) to 4 hours (Cummings et al., 2013). These studies highlight a number of factors related to the onset of vigilance decrements with workload (e.g., event rates) emerging consistently across studies. Performance decrements in underload settings have been shown to occur after approximately 30 minutes with time-on-task (Warm, Parasuraman, & Matthews, 2008), suggesting schedule or task adjustments should be made to reduce the likelihood of performance decrements. However, currently, few studies exist to demonstrate the threshold at which performance decrements occur in UAS operators in operational settings.

Individual differences (e.g., personality) also correlate with susceptibility to vigilance and performance decrements in underload conditions. A recent study that focused on individual differences in UAS operators and performance under conditions of varied task and workload found evidence of gender and personality differences (Lin, 2017). Specifically, men tended to perform better than women in high-demand conditions, but this gender effect was moderated by action video gaming experience. Additionally, this study found that those who scored high on conscientiousness and agreeableness (personality constructs) performed better on high-demand tasks (more resistant to overload effects and stress) but demonstrated a lower level of reliance on automated systems on these tasks. The length of a test session in this study was one hour and low workload conditions included six events per minute (versus 14 events per minute in the high workload conditions). Thus, it is unknown if these differences persist in longer missions with lower levels of task load and workload and the degree to which, if any, these differences can be utilized to optimize and maintain performance. Similarly, identification and development of strategies to mitigate negative effects of underload are being sought. One possible approach is to implement schedule or task adjustments to mitigate vigilance and performance decrements. Specifically, previous research has yielded promising results relative to secondary or dual-tasking (St. John & Risser, 2009), vigilance training (Daly et al., 2017), and use of a support system or control migration (Theißing & Schulte, 2016). Another possibility is to modify the operator environment or introduce a passive technology with known performance enhancing properties. For example, research has demonstrated that ambient blue light may serve such a function and may increase participants' motivation to complete simulated UAS monitoring tasks compared to red light and control conditions (Smith & Spiridon, 2018). To date, studies have been limited in terms of demonstrating effects that directly impact performance and currently available light sources in operator stations tend not to be used due to operator discomfort.

Finally, psychophysiological measurements as indicators of changes in workload and ultimately performance decrements is a rapidly developing area of research. To date, there have been promising results regarding the use of electroencephalogram (EEG), measuring brain activity, to predict an operator's state (Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014). Specifically, EEG may able to provide indications of the mental or physical engagement of the individual that correspond to changes in task demands such that alterations in the EEG waveform indicate changes in workload of the individual (Borghini et al., 2014; Fairclough, Venables, & Tattersall, 2005; Wilson, Fullenkamp, & Davis, 1994). Moreover, research has shown that cognitive state metrics of workload derived from EEG power spectral density (PSD) measurements correlate with overt performance decrements observed during high workload conditions (r = -.68) as well as subjective ratings of workload (r = .79; Berka et al. [2007]). However, additional research is required to gain a better understanding of psychophysiological changes that occur with underload, and the impact of different task loadings and environmental factors on those indices. Identifying the psychophysiological indicators of an operator in an underload state compared to periods of high workload is necessary to advancing the development and validation of real-time operator state monitoring. Ultimately, the goal of this monitoring is to use objective and non-invasive sensors to predict a compromised state of the operator and subsequently allow the system to adapt by initiating additional automation features (McColl, Heffner, Banbury, Charron, Arrabito, & Hou, 2017).

This study was designed to meet three objectives:

- Objective 1: To evaluate workload level as a predictor of performance and vigilance decrements relative to time-on-task while controlling for individual's boredom susceptibility during simulated UAS threat and change detection tasks.
- Objective 2: To compare the efficacy of vigilance decrement countermeasures (dual-task and ambient light) on sustaining performance in UAS operators and provide recommendations for operational use.
- Objective 3: To demonstrate patterns of psychophysiological indicators of operator state.

Methods

This study employed a mixed design including two within-subjects variables (workload and time-on-task) and one between-subjects variable (group). Workload (high and low) was manipulated by the event rate with 4 events/minute in the low condition and 19 events/minute in the high condition. Time-on-task was defined as 15-minute sequential blocks for the duration of the task (8 blocks total). Finally, group corresponded to the performance sustaining conditions experienced on day two of the experiment (control, dual-task, ambient light). Prior to execution, this study was reviewed and approved by the U.S. Army Medical Research and Development Command's Institutional Review Board.

Participants

Participants were healthy U.S. Army Soldiers (Active Duty, National Guard, Reserves) between the ages of 18 and 40 years (M_{age} = 32.14, SD = 5.98). Of the 22 participants, 21 were male (1 female). All participants reported normal hearing, vision (or corrected to normal vision), and cognitive function. Volunteers with a history of medical conditions affecting sleep or attention were ineligible for participation. Additionally, participants were required to sleep a minimum of 6 hours the night before participation. Participants were required to refrain from consumption of stimulants (including caffeine) and over-the-counter medications that may induce drowsiness for a minimum of 16 hours prior to each test session, alcohol and sedatives for 24 hours prior, and nicotine 8 hours prior to each testing session, which was assessed by self-report.

Measures

Electroencephalogram (EEG). Data was collected using the B-Alert X-24 wireless wet electrode system with 20 channels corresponding to scalp locations according to the International 10-20 system (frontal channels: Fp1, Fp2, F7, F3, Fz, F4, F8; central channels: C3, Cz, C4, T3, T4; parietal and occipital channels: P3, POz, Pz, P4, T5, T6, O1, O2). PSD values were computed using the automated algorithms provided through the B-Alert Live Software (Advanced Brain Monitoring, Inc., Carlsbad, CA). Prior to computing PSD values, artifacts were identified and removed using the ABM algorithms for artifacts associated with electromyography (EMG), eye blinks, excursions, saturations, and spikes (B-Alert Live, 2009). The software performed Fast Fourier Transform (FFT) on the data and calculated the amplitudes of the sinusoidal components for designated frequency bins. Frequency domain variables were based on the PSD derived after application of a 50% overlapping window, and a FFT with and without application of a Kaiser window. Each window size was 1 epoch containing 1 second of data (256 decontaminated EEG samples). The software then provides PSD values ranging from 1 to 40 Hz for each EEG channel that are logged to obtain a Gaussian distribution. Selected 1 Hz bins were averaged, then logged to create the EEG bands to be used in analyses (theta: 4—8 Hz; alpha: 9—13 Hz; beta: 14—30 Hz). Frontal theta, alpha, and beta PSD values were examined from the following frontal channels: Fp1, Fp2, F7, F3, Fz, F4, and F8, with values averaged by workload condition and time-on-task blocks for use in data analyses.

Boredom Proneness Scale. The Boredom Proneness Scale (Farmer & Sundberg, 1986) is composed of 28 statements that participants rate on a scale from 1 (highly disagree) to 7 (highly agree). Reliability and validity of this instrument have been extensively studied and consistently shown to be strong (Vodanovich & Watt, 2016). The total score was calculated for each participant and entered in the data analyses primarily as a covariate.

NASA Task Load Index. The NASA Task Load Index (NASA-TLX, Hart and Staveland [1988]) is a questionnaire that measures subjective workload. The subject rates the previous task, in this case flight, on the following categories, using a 100-point scale: mental demand, physical demand, temporal demand, performance, effort, and frustration. The NASA-TLX then provides a total workload score and scores for the six subscales. The total score was used in the presented analyses.

Instantaneous Self-Assessment Workload Scale (ISW). This very brief scale was developed by Brennan (1992) as a subjective measure of mental workload for air traffic controllers. Participants provide a self-rating of current workload on a scale from 1 (low) to 5 (high). The scale was developed to be a minimally intrusive subjective measure of workload. Participants provided a rating every 2 minutes for the duration of each testing session. Typically, the ISW is completed in short intervals (2-5 minutes).

Feel Bright Light Visor. The Feel Bright Light (manufactured by Physician Engineered Products, Inc.) is an ambient light device that attaches to a visor using Velcro and weighs 2 ounces (Figure 1). The light is safe for use up to 30 minutes continuously; however, effects have been shown with as few as 5-minutes of exposure (unpublished data, personal communication with Dr. Lauren Fowler, University of South Carolina Medical School, 8 August 2019). The device "utilizes wavelengths of light along the visible blue-green spectrum that falls between 470-560 nanometers, peaking at 525, and is completely UV free" which is labeled as GreenSafe™ technology (www.feelbrightlight.com). The 12,000-lux setting was used in this study.

Figure 1. Feel Bright Light.

UAS simulation. The Universal Mission Simulator (UMS) is a full simulation training system for Gray Eagle and Shadow UAS flight crews. The UMS trains UAS flight crews to operate all aspects of the UAS systems including preflight, taxi, launch, flight route to mission area, target area exploitation, weapons employment, return to airfield, landing, taxi, and post flight. The UMS provides flight deck visuals and controls configurable to simulate currently available UAS systems (Figures 2 and 3). In this study, participants completed a UAS supervision task that consisted of identifying visually presented targets (change detection) as well as responding to threats (threat detection). Given limitations with the data recording capabilities of the simulator, a tablet with keyboard was incorporated in the operator space (Figure 4). The tablet was used to present the supervision task and record responses by means of an open-source experiment generator software program (Mathôt, Schreij, & Theeuwes, 2012). The task presented a map with symbols denoting friendly or foe targets (identical to those employed in actual UAS operations; Figure 5). Participants were instructed to monitor the map and respond with the left arrow key when a symbol change was detected and a right arrow key when a threat (foe) symbol appeared in the monitored space. As previously stated, task workload was manipulated by event rate with 4 events/minute in the low workload condition and 19 events/minute in the high workload condition. Performance was defined by response time (measured in milliseconds [ms] from onset of stimuli) and correct responses based on stimuli type (friendly or foe). In order to compare performance between workload conditions, the percent of correct responses was calculated.

Figure 2. In the foreground is the operator's station. The green boxes are the computer system used to create and drive the simulation.
(Source: https://www.army.mil/article/187989/army_simulator_provides_readiness_to_drone_ flight_crews)

Figure 3. UMS pilot and payload operator interface includes large panel mounted display, joystick, keyboard, system switches and buttons, and center multifunction displays. (Source:https://www.army.mil/article/187989/army_simulator_provides_readiness_to_drone_flight_crews)

Figure 4. USAARL UAS simulator with tablet integrated into space.

Figure 5. Example of map and symbols presented in task.

Procedures

While past research suggests that a one-hour task duration is sufficient to induce vigilance decrements, this is atypical to UAS operations and thus the mission length was set at 4 hours. Two hours presented with the low workload condition and 2 hours with the high workload condition separated by a 15-minute break. Past research of UAS operator boredom has employed tasks of similar durations (Cummings et al., 2013; Thompson, Lopez, Hickey, DaLuz, Caldwell, & Tvaryanas, 2006). Participants completed two test sessions on separate days.

Session 1. Participants completed informed consent and then completed a medical screening form to determine eligibility. Next, participants completed a set of surveys including demographics and the Boredom Proneness Scale. Participants were then trained on how to complete the NASA-TLX, by being provided definitions and explanations of each subscale, as well as example ratings. Following completion of the questionnaires and training, participants were fitted with the psychophysiological recording devices. A 15-minute EEG baseline was collected while the participant completed three unscored tasks (three-choice vigilance task, eyes open task, and eyes closed task) on a laptop computer. Participants were then introduced to the UAS simulator and task, receiving a 10-minute training session. Research team members verbally confirmed understanding with the participants before proceeding. Participants then began the 4-hour simulated UAS mission. Two hours of the mission were at a high workload level and two hours at a low workload level separated by a 15-minute break. The order of the presentation of workload level was counterbalanced. Participants received an auditory prompt to provide a subjective rating of workload every 2 minutes (ISW). At the end of each two-hour session, participants completed the NASA-TLX. Recording devices were then removed and participants released for the day.

Session 2. Session 2 was very similar to session 1 in that participants confirmed eligibility and were fitted with recording devices upon arrival. Participants were then informed of which group they were randomly assigned to: control, dual-task, or ambient blue-green light. For those in the dual-task group, a brief introduction to the unrelated secondary task (the continuous manual tracking of a UAS payload) was presented. Participants were instructed to continuously and manually control the UAS payload to focus on a specified target. This task was modeled after Adams, 1961 although it has been modified for use in a UAS simulator. The manual tracking of a UAS payload is realistic in terms of UAS surveillance missions. The ambient light group experienced 5-minutes of light exposure using the Feel Bright Light Visor (Physician Engineered Products, Inc.) in a room with adequate lighting. Following this initial exposure, participants received a re-exposure following the first two-hour session during the 15-minute break. During the re-exposure, the task was paused, participants faced away from the monitors, and received 5-minutes of exposure. All participants then completed a 10-minute re-familiarization session with the simulated UAS mission and tasks. The remainder of the session mimicked that of session 1.

Quality control and statistical analyses

All data were inspected for impossible values and technical errors prior to analyses. All performance variables were calculated for each 15-minute segment of the test session by workload condition. Each variable includes responses to stimuli presented within that segment, thus, independent of each other. Percent of correct responses was calculated in order to compare accuracy across high and low workload conditions. The raw numbers of hits and misses could not be directly compared given that the number of stimuli presented in the high workload condition far exceeded that in the low workload condition. In order to evaluate the relationship between workload level (high, low), time-on-task (15, 30, 45, 60, 75, 90, 105, and 120 minutes), and boredom susceptibility, repeated-measures ANCOVAs were run on mean response time and percent correct responses from the first test session. Boredom Proneness scores were included as a covariate. Analyses for each performance variable were run independently given a high-degree of multicollinearity (Pearson's *r* correlation coefficients > 0.75) between performance outcomes. The second objective was to evaluate the utility of two countermeasures to maintain performance over the 120-minute test sessions. To do so, 3 (Group: dual-task, ambient light, control) X 8 (time-on-task) mixed model ANCOVAs were run on mean response time and percent correct responses during the second test session. Analyses were run separately for each workload level given that workload was not an independent variable of interest for this objective. Boredom Proneness scores were included as a covariate. Correlational analyses (Pearson's *r*) were used to evaluate the concurrent validity of the EEG (total frontal alpha, total frontal beta, and total front theta) as indices of workload (relative to the TLX scores). Also, paired-samples *t*-tests were used to evaluate whether brain wave activity varied by workload condition.

Results

In order to determine whether the workload manipulation was effective in terms of participants' perceptions of workload, paired-samples *t*-tests were run using the total NASA TLX scores in the high and low workload conditions for both test sessions. For the baseline session, total TLX scores were significantly lower in the low workload condition (*M* = 39.79, *SD* = 20.98) than the high (*M* = 46.62, *SD* = 14.84), *t*(21) = 2.38, *p* = 0.02. The same was true in the experimental session (low, *M* = 39.57, *SD* = 17.61; high, *M* = 47.79, *SD* = 17.91; *t*(19) = 3.48, *p* = 0.002). For the experimental session, two participants' data were incomplete, and thus, not included in the analysis.

Problematic data

Responses on the ISW did not sufficiently vary, and therefore, precluded analyses. Specifically, participants tended to give the same response for the entire administration session. It is unclear if this was truly an indication of a lack of change in perceived workload throughout the task or the responses were due to a lack of understanding of how to respond or insufficient attention paid to this aspect of the test sessions. Thus, ISW responses were excluded from the following analyses.

Objective 1

Accuracy. The full model was not significant for percent correct responses ($p > 0.05$). However, the observed power was less than optimal (power < 0.80) and thus two independent ANCOVAs (one per workload condition) were run with one within-subjects variable (time; 8 levels) and one covariate (Boredom Proneness). The model was not significant for the high workload condition ($p > 0.05$) but there was a significant effect of time for the low workload condition whilst controlling for Boredom Proneness ($F(7, 140) = 2.07$, $p = 0.05$) (Figure 6). Post-hoc paired comparisons revealed that the mean percent correct was significantly greatest in the first interval ($p < 0.01$). The mean percent correct in the second interval was greater than that from the remaining six intervals ($p < 0.01$). As can be seen in Figure 6, the mean percent correct was approximately 92% in the first interval, 88% in the second interval, and, roughly, 80% for the remainder of the task. By comparison, the overall mean percent correct in the high workload condition was approximately 91% ($SE = 1.4$).

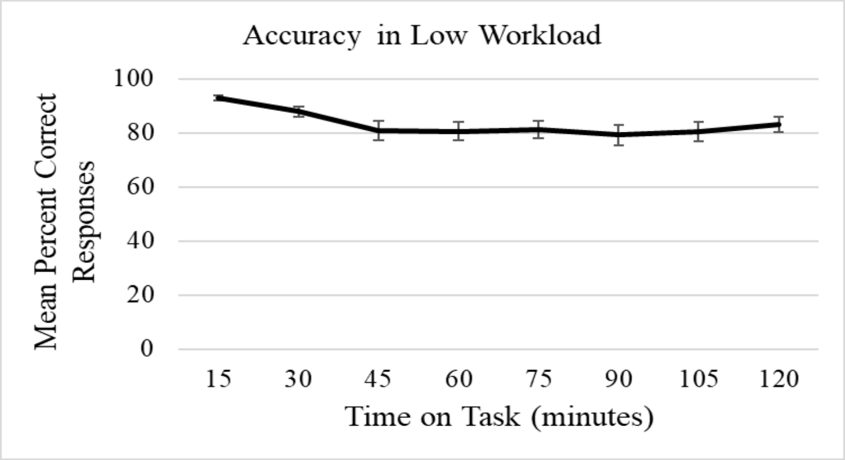

Figure 6. Accuracy expressed by percent correct responses by 15-minute intervals in the low workload condition. Error bars represent standard error of the mean.

Speed. Response speed was defined by mean response time (ms). The results of the 8 (time-on-task) X 2 (workload) ANCOVA showed a main effect of time (*F*(7, 140) = 3.94, *p* = 0.01) and a main effect of workload (*F*(7, 140) = 3.94, *p* = 0.01) (Figure 7). Specifically, responses were significantly slower in the low workload condition (*M* = 1051.29 ms, *SE* = 21.96) than in the high workload condition (*M* = 1448.53 ms, *SE* = 61.36). With respect to time-on-task, responses were significantly fastest in the first 15 minutes of the task (*M* = 1035.11 ms, *SE* = 30.92), slowed in the second interval, 30 minutes, (*M* = 1213.05 ms, *SE* = 44.88), and began to plateau by 45 minutes (*M* = 1282.69 ms, *SE* = 41.76).

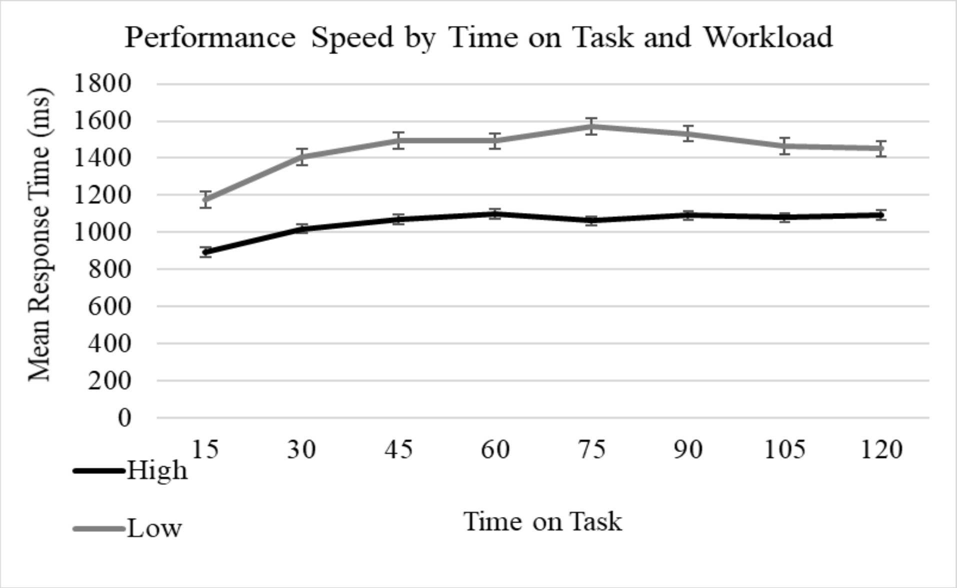

Figure 7. Speed expressed as mean response time (ms) by 15-minute intervals and by workload condition. Error bars represent standard error of the mean.

Objective 2

Accuracy. The results of the 3 X 8 ANCOVA with mean percent correct in the low workload condition as the outcome variable showed a significant interaction between time-on-task and countermeasure group, *F* (14,119) = 1.98, $p = 0.025$ (Figure 8). Post-hoc paired-samples *t*-tests were conducted and a Bonferoni correction was applied ($\alpha = 0.05/24 = 0.002$). No comparisons met the corrected threshold. However, there was only a significant main effect of time-on-task in the high workload condition, *F* (7,126) = 2.11, $p = 0.047$ (Figure 9). Specifically, accuracy was significantly greatest, overall, in the first segment (first 15 minutes of task) compared to that in 3rd (minutes 30 to 45) to 7th segments (minutes 90 to 105) of the session (paired-samples t-tests; $p < 0.05$).

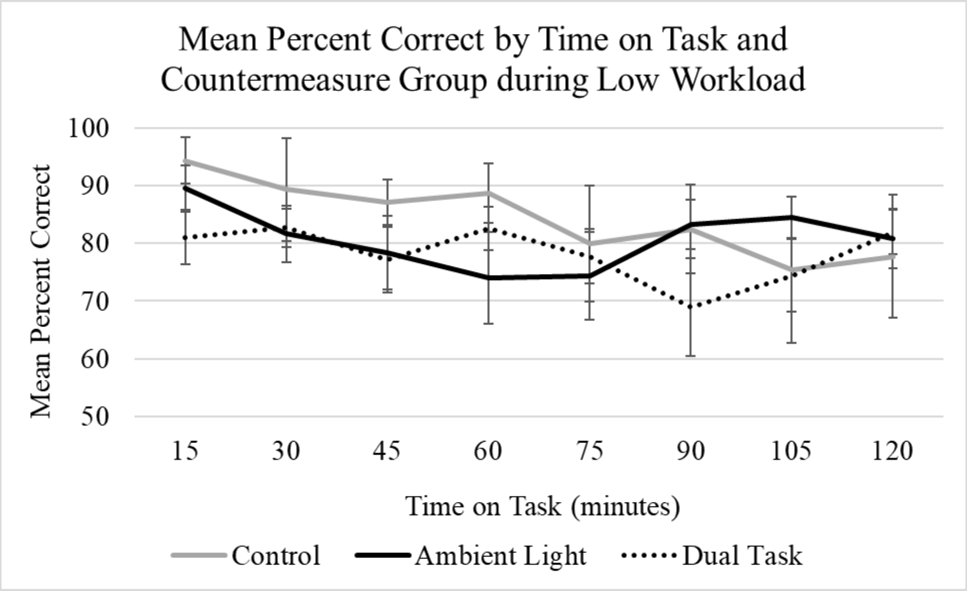

Figure 8. Accuracy expressed as mean percent correct responses by 15-minute intervals and countermeasure group in the low workload condition of the experimental test session. Error bars represent standard error of the mean.

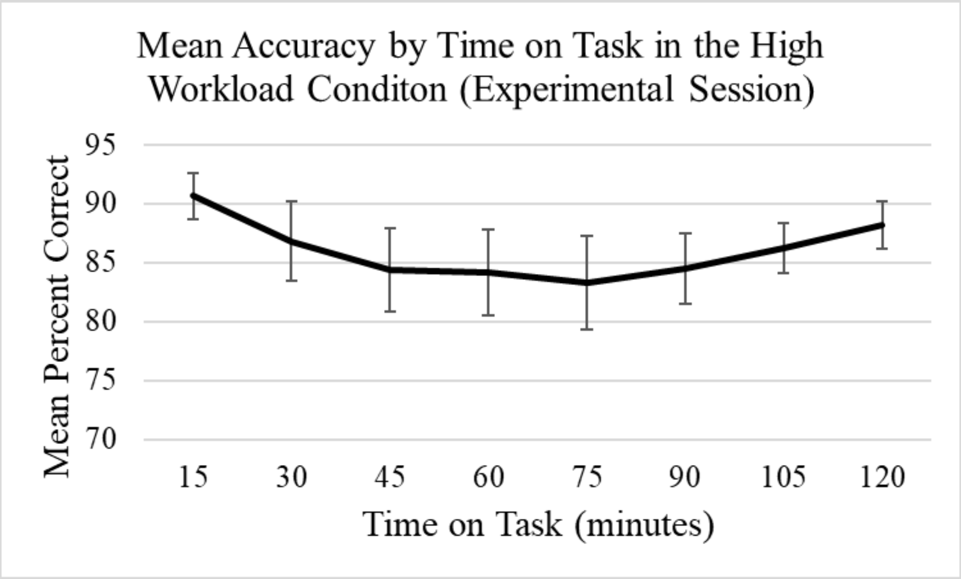

Figure 9. Accuracy expressed as mean percent correct responses by 15-minute intervals in the high workload condition of the experimental test session. Error bars represent standard error of the mean.

Speed. The results of the 3 X 8 ANCOVAs with mean response time as the outcome variable did not support any significant effects in either the low or high workload conditions.

Objective 3

Results of the paired-samples *t*-tests did not support any differences in total frontal alpha and beta activity between workload conditions. However, frontal theta activity was significantly higher in the high workload condition (*M* = 0.855, *SE* = 0.004) than the low workload condition (*M* = 0.849, *SE* = 0.005; *t*(19) = -3.04, *p* = 0.007). Correlational analyses revealed one significant relationship between total frontal alpha and NASA TLX scores in the high workload condition (*r*(20) = 0.447, *p* = 0.048). The relationship suggests that frontal alpha increased as perceived workload increased.

Discussion

This study aimed to meet three objectives: 1) to evaluate performance decrements during an extended UAS simulated operational task relative to time-on-task; 2) to evaluate the utility of two potential countermeasures for implementation in periods of underload (dual-task, ambient light); and 3) to demonstrate whether patterns in EEG correspond to workload levels. Due to COVID-19 related delays, this study was executed on a shortened time line resulting in a smaller sample size than originally estimated, which, in turn, limited the depth of the findings. Despite this challenge, the study yielded significant findings.

With respect to time-on-task, the results of this study suggest that performance, specifically accuracy, during the low workload condition diminished by an average of 12% by 30-minutes whereas accuracy was consistent across the two-hour duration in the high workload condition. Accuracy began to diminish by 15-minutes on task, reached its lowest point by 30-minutes, and then plateaued for the remainder of the task in the low workload condition. Response times did differ by workload condition with quicker responses in the high than low workload condition. The pattern of response time overall (collapsed over workload condition) was similar to that seen for accuracy in the low workload condition such that times slowed by 15-minutes on task and reached their peak by 45-minutes on task followed by a plateau. These findings are consistent with those previously reported in the literature (Warm, Parasuraman, & Matthews, 2008) and further validate these findings by controlling for individual boredom proneness. In the context of this study, within 30-minutes of task onset, accuracy dropped to an average of 80% with a low event rate. From an operational perspective, event rates in real-world settings are likely inconsistent over time and arguably often lower than those presented in this study. To fully understand the variability in performance and accuracy during such conditions, additional study is needed, however, these findings suggest that performance is compromised within a relatively short window of time following task onset.

While the results of this study did not support utility of either countermeasure evaluated, the pattern of performance accuracy over the duration of the task appears to differ between groups. Again, this pattern is not statistically supported and but may justify further evaluation of these strategies. Specifically, it appears that the performance decline is linear in nature for the control group such that accuracy steadily declines over time. On the other hand, the ambient light group presents a curvilinear pattern such that accuracy declines in the first hour of the task and then steadily increases over the second hour of the task. The dual-task group, however, appears to steadily decline similar to the control group but then exhibits a boost in accuracy during the last 15-minutes on task. Without further exploration, the true nature of these relationships are unknown, however, further evaluation of both countermeasures is warranted.

Finally, the patterns in frontal alpha and beta activity measured by EEG did not differ between workload conditions. However, the total frontal theta activity was greater in the high workload than low workload condition (workload conditions defined by event rate). This result is contradictory to that published in the literature (Feltman, Bernhardt, & Kelley, 2020) and may be driven by an unknown moderating variable. However, when evaluating the relationship between reported workload level, specifically NASA-TLX scores, and brain wave activity, the results do align with that reported previously. In particular, a recent study showed that during a visually demanding task, frontal alpha activity was positively related to workload such that frontal alpha activity increased as workload increased (Feltman et al., 2020). The current study replicated this finding of a moderate, positive relationship during a visually demanding task. Taken together, this further supports the exploration of EEG and specifically frontal alpha activity with respect to operator state monitoring.

The findings of this study are to be interpreted with caution in light of its limitations. First, as previously stated, the sample size for this study was compromised by COVID-19 delays. This is has the most substantial impact on the between-groups manipulation such that two groups consisted of five participants and one group had ten participants. Technical errors yielded unusable data for two participants' day 2 session. Not only was the sample size diminished but uneven between the groups. Given this limitation, at best, the data are inconclusive with respect to this objective. Another limitation is that participants were not specifically UAS operators. Granted, the task was designed such that anyone could complete it with minimal instructions, however, we do not know if and how performance may be different in the sub-population explicitly. Finally, a number of individual differences may moderate any such effects of underload that were not explicitly measured in this study such as intelligence.

The findings presented here do support further research to understand how performance can be maintained during low workload tasks. Given that performance was compromised by 30-minutes on task, a possible approach to mitigate this is to introduce task switching at set intervals. Also, further evaluation of the countermeasures included here is necessary to either rule them out as options or determine how best to employ them. Finally, these data will be further evaluated in a subsequent study with a higher degree of precision using predictive modeling to validate whether EEG recordings show utility in terms of performance prediction. This modeling approach is ongoing.

Conclusions

Overall, the findings of this study are consistent with past research on performance decrements in conditions of underload such that deficits were evident by 15-minutes and reached a minimum by 30-minutes on task. In low workload, accuracy and response time were degraded whereas, in high workload, only response time increased as a function of time-on-task. Whether, and if so to what degree, these decrements can be mitigated by dual-tasking or use of a short-duration ambient light exposure cannot be concluded given the limitations of this study. However, the observed patterns of brain wave activity relative to perceived workload replicate past research and further support the evaluation of EEG as a component of operator state monitoring.

References

Adams, J. A. (1961). Human tracking behavior. *Psychological Bulletin, 58,* 55–79. DOI: 10.1037/h0041559

B-Alert Live. (2009). [Computer software]. Carlsbad, CA: Advanced Brain Monitoring, Inc.

Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., . . . & Craven, P.L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine, 78*(5), B231-B244.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews, 44*, 58-75.

Boyer, M., Cummings, M.L., Spence, L.B., & Solovey, E.T. (2015). Investigating mental workload changes in a long duration supervisory control task. *Interacting with Computers, 27*(5), 512-520. DOI: 10.1093/iwc/iwv012

Brennan, S.D. (1992). *An experimental report on rating scale descriptor sets for the instantaneous self-assessment (ISA) recorder* (DRA Technical Memorandum [CAD5] 92017). DRA Maritime Command and Control Division: Portsmouth, VA.

Cummings, M.L., Mastracchio, C., Thornburg, K.M., & Mkrtchyan, A. (2013). Boredom and distraction in multiple unmanned vehicle supervisory control. *Interacting with Computers, 25*(1), 34-47. DOI:10.1093/iwc/iws011

Daly, T., Murphy, J., Anglin, K., Szalma, J., Acree, M., Landsberg, C., & Bowens, L. (2017). Moving vigilance out of the laboratory: Dynamic scenarios for UAS operator vigilance training. In: Schmorrow D., Fidopiastis C. (Eds.) *Augmented Cognition. Enhancing Cognition and Behavior in Complex Human Environments. AC 2017. Lecture Notes in Computer Science, vol 10285*. New York: Springer.

Epling, S.L., Russell, P.N., & Helton, W.S. (2016). A new semantic vigilance task: Vigilance decrement, workload, and sensitivity to dual-task costs. *Experimental Brain Research, 234*(1), 133-139. DOI: 10.1007/s00221-015-4444-0

Fairclough, S.H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology, 56*(2), 171-184. DOI: 10.1016/j.ijpsycho.2004.11.003

Farmer, R., & Sundberg, N.D. (1986). Boredom proneness: The development and correlates of a new scale. *Journal of Personality Assessment, 50*, 4–17. DOI: 10.1207/s15327752jpa5001_2

Federal Aviation Administration. (2016). Pilots' role in collision avoidance [Advisory Circular, AC 90-48D].

Retrieved from http://www.faa.gov/ documentLibrary/media/Advisory_Circular/AC_90-48D.pdf.

Feltman, K.A., Bernhardt, K.A., & Kelley, A.M. (2020). Measuring the domain specificity of workload using EEG: Auditory and visual domains in rotary-wing simulated flight. *Human Factors*, 0018720820928626.

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.), *Advances in Psychology Vol. 52* (pp. 139-183). Amsterdam, Netherlands: North-Holland.

Helton, W.S., Hollander, T.D., Warm, J.S., Tripp, L.D., Parsons, K., Matthews, G., . . . Hancock, P.A. (2007). The abbreviated vigilance task and cerebral hemodynamics. *Journal of Clinical and Experimental Neuropsychology, 29*(5), 545-552.
DOI: 10.1080/13803390600814757

Lin, J. (2017). *The Impact of Automation and Stress on Human Performance in UAV Operation*. (Doctoral dissertation), University of Central Florida, Orlando, FL.
Retrieved from http://purl.fcla.edu/fcla/etd/CFE0006951

Loffi, J.M., Wallace, R.J., Jacob, J.D., & Dunlap, J.C. (2016). Sensing the threat: Pilot visual detection of small unmanned aircraft systems in meteorological conditions. *International Journal of Aviation, Aeronautics, and Aerospace, 3*(3).

Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 1*(1), 6-21.
DOI: 10.1080/17470214808416738

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324

McColl, D., Heffner, K., Banbury, S., Charron, M., Arrabito, R., & Hou, M. (2017). Auditory pathway: Intelligent adaptive automation for a UAS ground control station. *Engineering Psychology and Cognitive Ergonomics: Performance, Emotion, and Situation Awareness*, 329-342.

Senoussi, M., Verdière, K.J., Bovo, A., Ponzoni Carvalho Chanel, C., Dehais, F., & Roy, R.N. (2017). Pre-stimulus antero-posterior EEG connectivity predicts performance in a UAV monitoring task. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics Banff, AB, Canada 5-8 October 2017*.
Retrieved from: http://oatao.univ-toulouse.fr/19280/

Smith, S. & Spiridon, E. (2018). Influence of ambient light and feedback on motivation to carry out a task: implications for operation of unmanned aircraft. *International Journal of Unmanned Systems Engineering*, 7(1), 12-23.

St. John, M., & Risser, M.R. (2009). Sustaining vigilance by activating a secondary task when

inattention is detected. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53*(3), 155-159.
DOI: 10.1177/154193120905300304

Theißing, N., & Schulte, A. (2016). Designing a Support System to Mitigate Pilot Error While Minimizing Out-of-the-Loop-Effects. *Engineering Psychology and Cognitive Ergonomics*, 439-451.

Thompson, W.T., Lopez, N., Hickey, P., DaLuz, C., Caldwell, J.L., & Tvaryanas, A.P. (2006). *Effects of shift work and sustained operations: Operator performance in remotely piloted aircraft (OP-REPAIR)* (Report No. HSW-PE-BR-TR-2006-OOO1). Brooks AFB, TX: United States Air Force.
Retrieved from: http://www.dtic.mil/dtic/tr/fulltext/u2/a443145.pdf

Tvaryanas, A.P., Thompson, W.T., & Constable, S.H. (2005). *US military unmanned aerial vehicle mishaps: Assessment of the role of human factors using human factors analysis and classification system (HFACS)* (Report No. HSW-PE-BR-TR-2005-0001). Brooks AFB, TX: United States Air Force.
Retrieved from: http://www.dtic.mil/dtic/tr/fulltext/u2/a435063.pdf

Vodanovich, S.J., & Watt, J.D. (2016). Self-report measures of boredom: An updated review of the literature. *The Journal of Psychology, 150*(2), 196-228.

Warm, J.S., Matthews, G., & Finomore Jr, V.S. (2017). Vigilance, workload, and stress. In *Performance under stress* (pp. 131-158). Cleveland, OH: CRC Press.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433-441.
DOI: 10.1518/001872008X312152

Whitlock, C. (2014). When drones fall from the sky. *The Washington Post*, 20 June 2014. Retrieved from: https://www.washingtonpost.com/sf/investigative/2014/06/20/when-drones-fall-from-the-sky/?utm_term=.ff1efdff8381

Wilson, G.F., Fullenkamp, P., & Davis, I. (1994). Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation, space, and environmental medicine*.

Wohleber, R.W., Matthews, G., Lin, J., Szalma, J.L., Calhoun, G.L., Funke, G.J., Chiu, C.P., & Ruff, H.A. (2019). Vigilance and automation dependence in operation of multiple unmanned aerial systems (UAS): A simulation study. *Human Factors, 61*(3), 488-505.

Young, M.S., Brookhuis, K.A., Wickens, C.D., & Hancock, P.A. (2015). State of science: Mental workload in ergonomics. *Ergonomics, 58*(1), 1-17.
DOI: 10.1080/00140139.2014.956151

U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama

Army Futures Command
U.S. Army Medical Research and Development Command