Physiological Indicators of Workload in a Remotely Piloted Aircraft Simulation

Michael Hoepf

Matt Middendorf

Samantha Epling

Scott Galster

PHYSIOLOGICAL INDICATORS OF WORKLOAD IN A REMOTELY PILOTED AIRCRAFT SIMULATION

Michael Hoepf
Oak Ridge Institute for Science and Education
Matt Middendorf
Middendorf Scientific Services
Samantha Epling
Ball Aerospace & Technologies Corp.
Scott Galster
Air Force Research Laboratory
Dayton, Ohio

Toward preventing performance decrements associated with mental overload in remotely piloted aircraft (RPA) operations, the current research investigated the feasibility of using physiological measures to assess cognitive workload. Two RPA operators were interviewed to identify factors that impact workload in target tracking missions. Performance, subjective workload, cortical, cardiac and eye data were collected. One cardiac and several eye measures were sensitive to changes in workload as evidenced by performance and subjective workload data. Potential future applications of this research include closed loop systems that employ advanced augmentation strategies, such as adaptive automation. Thus, by identifying physiological measures well suited for monitoring workload a realistic simulation, this research advances the literature toward real-time workload mitigation in RPA field operations.

U.S. armed forces are increasingly using remotely piloted aircraft (RPA) to accomplish missions in hostile environments because of their standoff capability in areas that are difficult to access or otherwise considered too hazardous for manned aircraft or personnel on the ground (U.S. Department of Defense, 2011). It has been documented that the military intends to increase the number of RPA in service while simultaneously reducing the number of operators (Dixon, Wickens, & Chang, 2004). One proposal to accomplish this is to allow operators to control multiple aircraft simultaneously (Rose, Arnold, & Howse, 2013). However, piloting one aircraft remotely is a complex task, and operating additional aircraft could increase task demands sharply. This is potentially problematic because cognitive overload can negatively impact performance (Young & Stanton, 2002). One solution to offset this risk is to monitor operator workload in real-time and provide augmentation before performance decrements occur. Physiological measures, which have been shown to reflect changes in cognitive workload in various environments (e.g., Wilson & Russell, 2007), are well suited for this goal.

Before physiological measures can be used to monitor workload in RPA field operations, additional research is needed using realistic task environments. This is because each category of physiological measures (e.g., cortical, cardiac, and eye-based) is sensitive to workload in different situations. Hankins and Wilson (1998), for instance, found that cortical measures were sensitive to workload during mental calculation, cardiac measures were related to workload during flight segments heavily dependent on instrument use, and eye activity was associated with workload during visually demanding flight segments.

To address the research needs outlined above, an experiment was designed using a high-fidelity RPA simulation. In this study workload was experimentally manipulated and several physiological measures were collected while participants performed a target tracking task. Two RPA subject matter experts (SMEs) were interviewed to identify factors that can affect workload in this task. Two factors that were identified were weather (clear vs. hazy) and the route the target would follow (city vs. country). One of the SMEs engaged in a test run of the experiment to test and adjust the implementation of the factors. In addition to these two factors, a third factor was implemented which manipulated the number of targets being tracked (1 vs. 2).

In this study the physiological measures included the electroencephalogram (EEG), the electrocardiogram (ECG) and several eye measures. Researchers have demonstrated that EEG can be used in real-time to assess mental workload in aviation environments (e.g., Wilson & Russell, 2007). ECG was used to obtain heart rate (HR) and heart rate variability (HRV). In both laboratory and field settings, researchers typically observe HR increases and HRV decreases in high workload situations (Wilson, 1992). Eye measures included blink rate, blink duration, and

Distribution A: Approved for public release; distribution unlimited.
88ABW Cleared 03/02/2015; 88ABW-2015-0755.

pupil diameter. Generally, during increased cognitive workload, blink rate and duration decrease (Fogarty & Stern, 1989), and pupil diameter increases (Beatty, 1982).

Methods

Participants

Six people who were either students at a Midwestern university or recent graduates participated in the experiment. They were paid $15 per hour for their participation. Three participants were female and three were male. Age ranged from 19-28 years, with a mean of 22.3. They were screened for motor, perceptual, cognitive, heart, and neurological conditions, as well as hearing impairments. They were fluent in English, right-handed, had normal or corrected-to-normal eyesight with no color blindness, and provided written informed consent in accordance with human research ethics guidelines prior to the start of the experiment. All study procedures were reviewed and approved by the Air Force Research Laboratory Institutional Review Board.

The Task

Primary task. The primary task was developed using the RPA software platform "Vigilant Spirit." This software was produced by the Air Force Research Laboratory System Control Interfaces Branch (RHCI). The goal of the primary task was to track either one or two high value targets (HVTs) depending on the condition. Participants were instructed to keep the RPA sensor positioned over the HVTs, which they accomplished by clicking in the video feed with the mouse, causing the video feed to center on where they had clicked. The sensor slaved tracking feature would then automatically update the aircraft position to fly a loiter circle around this center point, thereby eliminating the need for manual aircraft navigation. Participants actively tracked the HVT(s) for the duration (4.5 minutes) of each trial.

Secondary task. A secondary task was presented concurrently with the primary task. The task consisted of answering cognitively challenging questions. There were three math questions and one mental rotation question per trial. Questions were presented verbally over a headset and transcriptions were displayed. Participants were instructed to press and hold the spacebar while they responded verbally.

Apparatus and Measures

Performance assessment. Performance was assessed using a composite scoring algorithm, which was based on components from both the primary and secondary task. For each trial, the maximum possible score was 1,000 points (800 primary and 200 secondary). To obtain points from the primary task, participants were required to keep the HVT(s) in their video feed(s). Maximum points were accumulated when using the highest two levels of zoom and half as many were accumulated at lower levels of zoom. For the secondary task, there were four questions per trial, each worth a maximum of 50 points. In order to obtain all points, participants had to respond correctly within 10 seconds. After 10 seconds, the participants would lose 1 point per second for the next 10 seconds, and then 2 points per second for the following 10 seconds. After 30 seconds, no points were given. Answering incorrectly resulted in a 5 point penalty.

Subjective workload. Subjective workload was assessed using the National Aeronautics and Space Administration-Task Load Index (NASA-TLX), a multidimensional measure that assesses perceived workload (Hart & Staveland, 1988). Workload was determined by averaging across the six sub-scales (mental demand, physical demand, temporal demand, performance, effort, and frustration). This average has been found to be psychometrically equivalent to the weighted sub-scale averaging suggested by the NASA-TLX authors (Nygren, 1991). Empirically, the weighted averages have not been found to be superior to the simple average of the sub-scales (Christ et al., 1993; Hendy, Hamilton, & Landry, 1993).

Physiological data acquisition and processing. The physiological data collected in this study included the EEG, ECG, vertical EOG (VEOG), and pupil diameter. The EEG, ECG, and VEOG signals were sampled using a Cleveland Medical Devices BioRadio 150. The EEG and VEOG were sampled at 480 Hz, and the ECG signal was sampled at 960 Hz. All signals connected to the BioRadio 150 were subjected to a hardware high pass filter with a break frequency of 0.5 Hz. The sampled data were transmitted wirelessly to a computer for processing and recording.

The EEG data were acquired using electrodes placed directly on the scalp and secured in place with an Electro-Cap manufactured by Electro-Cap International, Inc. EEG was measured at seven sites on the scalp in accordance with the international 10/20 system (Jasper, 1958). The seven sites were the F7, F8, T3, T4, Fz, Pz, and O2. The right and left mastoids were used as the reference and ground for the EEG signals. All initial electrode impedances were measured to be at or below 5 kΩ. The frequency bands (i.e., pass bands) used in the EEG signal processing were delta (1-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), gamma 1 (31-40 Hz), gamma 2 (41-57 Hz) and gamma 3 (63-100 Hz). A two second time domain window was used to process the raw EEG data. The raw data in the two second window was filtered using a 4^{th} order Butterworth band pass filter. A Hanning window was applied to the filtered data and power spectral analysis was performed. The resulting power in the pass band was then averaged. These steps were repeated for each frequency band and electrode site. The two second time domain windows had a 50% overlap, thus yielding one measure of average power every second. This signal processing approach yielded 49 EEG measures per second (7 sites with 7 bands per site).

The ECG data were acquired using two electrodes placed on the sternum and xiphoid process, and the VEOG data were acquired using two electrodes placed above and below the left eye. The initial electrode impedances for the VEOG and ECG were measured to be at or below 20 kΩ. The left mastoid was used as the ground for the ECG and VEOG signal. Interbeat intervals (IBIs) were extracted from the ECG data. The IBIs were used to calculate heart rate and heart rate variability. Blink rate and duration were extracted from the VEOG data using a blink detection algorithm (see Epling, Middendorf, Hoepf, & Galster, this volume). Pupil diameter data were acquired using the Smart Eye Pro 5.9 system with four cameras sampling data at 60Hz.

Competition. A social (no monetary compensation) competition was implemented to maintain motivation and prevent task disengagement (Sherif, Harvey, White, Wood, & Sherif, 1961). Top session average scores were posted on a whiteboard.

Procedure

Participants were brought into the laboratory for one day of training and four days of data collection. For training, participants first viewed a PowerPoint presentation containing a description of the task and measures, and then completed part-task training for the primary and secondary tasks. Training concluded with the completion of eight practice trials. On data collections days, participants were equipped with the physiological measurement devices and then completed eight experimental trials per day, for a total of 32 trials.

Experimental Design

The current investigation utilized a 2 x 2 x 2 full factorial design. There were three manipulations intended to impact workload, each containing two levels (easy and hard). The first manipulation was weather, which included clear (easy) and hazy (hard) conditions. The clear condition was free of clouds and visibility was unobstructed. A layer of fog was present in the hazy condition, which reduced visibility. The second manipulation was route difficulty, which referred to the roads the HVTs traveled. For the country (easy) routes, HVTs simply traveled back and forth along a long straight road, the view of the HVT was generally unobstructed. Conversely, for the city (hard) routes, the HVTs took many turns and sometimes became occluded by buildings. The third manipulation was the number of HVTs, which was either one (easy) or two (hard). In single HVT conditions, participants needed to track only one HVT, requiring only one RPA. Conversely, in the two HVT conditions, participants were required to utilize two RPA to track two HVTs simultaneously. Latin squares were used for counterbalancing such that each condition preceded every other condition an equal number of times, thereby combatting order effects.

Results

Performance. The performance, subjective workload, and physiological data were statistically evaluated using a three-way (weather, HVT, route) repeated-measures ANOVA. Performance in hazy conditions (M = 785.0, SE = 25.6) was not significantly different than the performance in clear conditions (M = 776.2, SE = 23.4). Performance was higher in conditions with country routes (M = 814.5, SE = 19.2) than in conditions with city routes (M = 746.7, SE = 31.6), $F(1, 5) = 10.18$, $p < .05$, and higher in one HVT conditions (M = 873.6, SE = 24.1) than two HVT conditions (M = 687.6, SE = 25.4), $F(1,5) = 220.30$, $p < .001$.

Subjective workload. Subjective workload in hazy conditions ($M = 43.5$, $SE = 4.3$) was not significantly different than clear conditions ($M = 43.3$, $SE = 5.0$). Subjective workload was higher in city conditions ($M = 47.6$, $SE = 5.3$) than country conditions ($M = 39.1$, $SE = 4.1$), $F(1, 5) = 18.52$, $p < .01$, and higher in two HVTs conditions ($M = 54.6$, $SE = 6.1$) than one HVT conditions ($M = 32.1$, $SE = 4.2$), $F(1, 5) = 18.97$, $p < .01$.

Cortical measures. The EEG measures (power at each site and frequency band) were analyzed for each manipulation, but for conciseness only the significant ($p < .05$) results are reported and the means, standard errors, and *F* values are not included. In regards to the weather manipulation, there was less power in hazy conditions than clear conditions at the O2 site in the alpha band. For the route manipulation, there was less power in city conditions than in country conditions at 7 sites, including F7, Fz, F8, T3, T4, and Pz in the delta band, and F7 in the theta band. For the HVT manipulation, there was more power for two HVT conditions than one HVT conditions at 19 sites, including F7, F8, T3, and O2 in the delta band, F7, F8, T3, T4, Pz, and O2 in the theta band, F7, F8, T3, Pz, and O2 in the alpha band, and T4 in the beta, and all three gamma bands. These effects may not be due to neural activity in the brain, but rather artifacts from eye activity (see discussion section).

Cardiac measures. HR was not significantly impacted by any of the experimental manipulations. HRV in hazy conditions ($M = 0.0539$, $SE = 0.0050$) was not significantly different than HRV in clear conditions ($M = 0.0533$, $SE = 0.0046$). HRV was significantly lower in city conditions ($M = 0.0530$, $SE = 0.0049$) than in country conditions ($M = 0.0542$, $SE = 0.0047$), $F(1,5) = 7.44$, $p < .05$, and significantly lower in two HVT conditions ($M = 0.0517$, $SE = 0.0046$) than one HVT conditions ($M = 0.0555$, $SE = 0.0050$), $F(1,5) = 19.46$, $p < .01$.

Eye measures. The weather manipulation did not significantly impact blink rate or duration. Blink rate was lower in city conditions ($M = 18.34$ bpm, $SE = 4.88$) than in country conditions ($M = 19.59$ bpm, $SE = 5.23$), $F(1,5) = 8.23$, $p < .05$. Blink rate was also lower in two the HVT conditions ($M = 16.28$ bpm, $SE = 4.50$) than in the one HVT conditions ($M = 21.65$ bpm, $SE = 5.87$), but this difference was not statistically significant $F(1,5) = 3.98$, $p = .10$. Blink duration was significantly shorter in city conditions ($M = 0.1041$s, $SE = 0.0042$) than in country conditions ($M = 0.1064$s, $SE = 0.0043$), $F(1,5) = 16.77$, $p < .01$, and shorter in two HVT conditions ($M = 0.1005$s, $SE = 0.0047$) than one HVT conditions ($M = 0.1099$s, $SE = 0.0041$), $F(1,5) = 13.81$, $p < .05$.

Interestingly, pupil diameter was significantly larger during clear conditions ($M = 4.06$mm, $SE = 0.68$) than hazy conditions ($M = 3.87$mm, $SE = 0.59$), $F(1,5) = 14.85$, $p < .05$. To investigate this finding, a Minolta Chroma-Meter CS-100 was used to assess the luminance of the two conditions. Results suggested that the pupil light reflex was most likely responsible for this difference, as hazy conditions were brighter than clear conditions. Pupil diameter was larger during city routes ($M = 4.01$mm, $SE = 0.63$) than during country routes ($M = 3.92$mm, $SE = 0.66$), although this difference was not significant, $F(1,5) = 3.38$, $p = .125$, and larger during two HVT conditions ($M = 4.09$mm, $SE = 0.62$) than during one HVT conditions ($M = 3.84$mm, $SE = 0.65$), $F(1,5) = 44.33$, $p < .01$.

Discussion

To meet increasing demand for RPA operations, future systems are envisioned in which single operators control multiple aircraft (Dixon et al., 2004). Such systems would allow an efficient use of resources during low workload operations. However, a concern is that workload could become excessive due to increased mental demand from managing multiple aircraft, possibly leading to performance decrements and mission failure. One solution to address excessive workload from controlling multiple vehicles, as well as existing challenges RPA operators experience, is to monitor operator state in real-time so that mental overload can be identified and handled appropriately. That is, accurate workload assessment would allow the implementation of augmentation strategies *before* performance decrements occur. By examining the feasibly of using physiological measures to monitor workload, this project advances the literature toward real-time workload mitigation in field operations.

In regard to the experimental manipulations, the results indicated that the weather manipulation did not impact workload. The HVTs were the only motorcycles in the simulation (other traffic consisted of cars, trucks, vans etc.), so it could be that the haze did not sufficiently obscure the visual cues necessary to track them. The route manipulation did effectively impact workload, so a strong point of this research is that this factor (identified in SME interviews) has physiological correlates that can be used for workload assessment. Similarly, results showed that tracking two targets had a strong impact on performance, workload, and physiological measures. The control of multiple semi-autonomous air vehicles is not a current capability, and so the present research is valuable in that it provides a preview of what may be expected if such a capability is implemented in future RPA workstations.

Physiological results revealed that one cardiac and several eye measures were sensitive to the same workload manipulations as the performance and subjective workload measures. Specifically, HRV, blink duration, blink rate, and pupil diameter were generally sensitive to the route and HVT manipulations. The haze manipulation did not impact performance or subjective workload, and the physiological measures generally did not reflect a change in workload either. These physiological findings are consistent with previous research (e.g., Beatty, 1982; Fogarty & Stern, 1989; Wilson, 1992), suggesting that the results did not occur by chance, and that these physiological measures may be well suited for real-time workload monitoring in RPA field operations.

Despite these promising results, other physiological measures did not function as expected. HR, for instance, did not demonstrate sensitivity to any workload manipulation. Additionally, the EEG data was difficult to interpret, as results were not always in the expected direction. Alpha power, for instance, increased at several sites in two HVT conditions, which were shown to be the more difficult conditions as evidenced by the performance and subjective workload data. This is in contrast to the classic concept that alpha activity is an idling rhythm of humans at rest, which becomes desynchronized during cognitive processes (Pfurtscheller & Lopes da Silva, 1999). One possible explanation is that EOG artifacts (see Fatourechi, Bashashati, Ward, & Birch, 2007) were present in the EEG data due to the additional saccades associated with two target tracking conditions. Although an effort was made to remove these artifacts, some residual effects remained in the data. Specifically, the band pass filter used to process the EEG data continues to ring for up to one second after the artifact occurs. This indicates a need for an improvement to the EEG signal processing software for future research.

Admittedly, a limitation of the current study is the small sample size, such that these findings could be specific to the sample. In addition, several measures (i.e., blink rate, pupil diameter) sometimes failed to yield significant differences, despite trending in the expected direction. A larger sample size would help to either confirm or deny the utility of these physiological measures.

Conclusion

The current research investigated the feasibility of using physiological measures to monitor workload in a high fidelity RPA tracking task. One cardiac (HRV) and several eye measures (blink rate, blink duration, and pupil diameter) demonstrated sensitivity to changes in workload, and thereby appear well suited to real-time workload monitoring. In the future, these measures could be used in physiologically driven adaptive automation (see Scerbo, 1996) systems to prevent performance decrements. Given the promising nature of these results, future research is encouraged on the topic.

Acknowledgements

The authors would like to thank Chelsey Credlebaugh and Jonathan Mead for their assistance in data reduction and preparation of this manuscript, as well as Chuck Goodyear for his help with statistical analysis. We would also like to thank Kevin Durkee, Noah DePriest, and Mark Squire for their technical support. This research was supported in part by an appointment to the Student Research Participation Program at the U.S. Air Force Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and USAFRL. The views expressed in this report are solely those of the authors and do not necessarily reflect the views of the employers or granting organizations.

References

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91,* 276-292.

Christ, R. E., Hill., S. G., Ayers, J. C., Iavecchia, H. M., Zaklad, A. L., & Bittner, A. (1993). *Application and validation of workload assessment techniques* (Technical Report 974). Alexandria, VA: U.S. Army Research Institute for the Behavioral Sciences.

Dixon, S. R., Wickens, C. D., & Chang, D. (2004). Unmanned aerial vehicle flight control: False alarms versus misses. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 152-156). New Orleans, LA: Human Factors and Ergonomics Society.

Epling, S., Middendorf, M., Hoepf, M., & Galster, S. (this volume). The electrooculogram and a new blink detection algorithm. In *Proceedings of the 18th International Symposium on Aviation Psychology*. Dayton, OH: Wright State University.

Fatourechi, M., Bashashati, A., Ward, R. K., & Birch, G. E. (2007). EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology, 118*(3), 480-494.

Fogarty, C., & Stern, J. (1989). Eye movements and blinks: Their relationship to higher cognitive processes. *International Journal of Psychophysiology, 8,* 35-42.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine, 69*, 360-367.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA- TLX (Task Load Index): Results of experimental and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp.138-183). Amsterdam: North-Holland Press.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35,* 579-601.

Jasper, H. (1958). Report of the committee on methods of clinical examination. *Electroencephalography and Clinical Neurophysiology, 10*, 370-375.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33,* 17-31.

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology, 110*, 1842-1857.

Rose, M. R., Arnold, R. D., & Howse, W. R. (2013). Unmanned aircraft systems selection practices: Current research and future directions. *Military Psychology, 25*(5), 413-427.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications*. Mahwah, NJ: Erlbaum.

Sherif, M., Harvey, O. J., White, B. J., Wood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The robber's cave experiment.* Normal, OK: Institute of Group Studies.

U.S. Department of Defense. (2011). *Unmanned systems integrated roadmap: FY2011-2036* (Reference No. 11-S-3613). Washington, DC: Department of Defense.

Wilson, G. F., (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology*, *34*(2-3), 163-178.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human Factors, 49*(6), 1005-1018.

Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, *3*(2), 178-194.