Brain Biomarkers Based Assessment of Cognitive Workload in Pilots under Various Task Demands*

Rodolphe J. Gentili, Jeremy C. Rietschel, Kyle J. Jaquess, Li-Chuan Lo, CAPT Michael Prevost, Matt W. Miller, Jessica M. Mohler, Hyuk Oh, Ying Ying Tan, Bradley D. Hatfield.

***Abstract*— Cognitive workload is an important element of cognitive-motor performance such as that exhibited during the piloting of an aircraft. Namely, an increase in task demands on the pilot can elevate cognitive information processing and, thus, the risk of human error. As such, there is a need to develop methods that reliably assess mental workload in pilots within operational settings. The present study contributes to this research goal by identifying physiological and brain biomarkers of cognitive workload and attentional reserve during a simulated aircraft piloting task under three progressive levels of challenge. A newly developed experimental method was employed by which electroencephalography (EEG) was acquired via a dry (i.e., gel-free sensors) system using few scalp sites. Self-reported responses to surveys and piloting performance indicators were analyzed. The findings revealed that as the challenge (task demands) increased, the perceived mental load increased, attentional reserve was attenuated, and task performance decreased. Such an increase in task demands was also reflected by changes in heart rate variability (HRV), as well as in the amplitude of the P300 component of event-related potentials to auditory probes, and in the spectral power of specific EEG frequency bands. This work provides a first step towards a long-term goal to develop a composite system of biomarkers for real-time cognitive workload assessment and state assessment of pilots in operational settings.**

I. Introduction

Human cognitive-motor performance is highly dependent on the efficiency of allocation of brain resources during demanding tasks such as the piloting of aircraft, which is multifaceted in nature. The increase of task demands on a pilot can result in an increase of mental workload and a corresponding reduction of attentional reserve. If such attentional reserve is depleted below a certain threshold, the cognitive processing of incoming information can be delayed or even impeded, leading to a high likelihood of human error [1,2]. As such, many studies have examined mental workload in the aviation domain [2-5]. Specifically, the goal of new fields such as operational neuroscience and augmented cognition is to identify robust physiological markers to assess/monitor changes in cognitive workload. This would allow for the development of adaptive human-machine interfaces, novel methods for selecting/training pilots, and any applications involving human-machine interactions [6].

In order to quantify the cognitive workload in pilots, previous studies have examined various psychophysiological signals (e.g., HRV, EEG, eye movements, and pupillometry) [7]. Overall, these studies used EEG systems that are gel-based, which can be problematic in operational settings due to time-intensive preparatory routines. Also, typical montages include multiple electrodes; however, the availability of brain markers of cognitive workload using a limited number of electrodes would reduce the preparatory burden, computational cost of the EEG, and subsequently be better suited for real-time processing in operational contexts. A new method using EEG-derived ERPs was used here to assess cognitive workload by evaluating attentional reserve [8].

Therefore, the proposed study provides a unique contribution to the existing literature by: i) examining physiological (i.e., HRV) and brain biomarkers (i.e., spectral content, ERPs) of cognitive workload and attentional reserve in pilots employing few dry (i.e., gel free) EEG electrodes during a series of tasks characterized by various flight demands, and ii) assessing if the new method developed by Miller et al. [8] can be used with dry EEG sensors in more ecologically valid tasks (flight simulator) to detect cognitive workload with resolution beyond the binary low/high demand. This study is an initial step in identifying and selecting multiple biomarkers to form a composite metric sensitive to the dynamic construct of cognitive workload and attentional reserve while combining data mining (i.e., feature selection) and classification (i.e., machine learning) techniques. Our long-term goal is to develop an approach using single/few trials for combining multiple markers characterized by different time scales and allowing for real-time assessment of the cognitive workload/attentional reserve via the use of portable systems (e.g., dry EEG system in concert with other biomarkers) in ecologically valid settings.

II. Materials and Method

A. Task

Thirty-eight healthy young participants (ages ranged from 19 to 23 years) who were enrolled in the United States Naval Academy (USNA) Powered Flight Program and received

*Research supported by the Lockheed Martin Corporation (project # 13051318), USA.

R. J. Gentili is with the Department of Kinesiology, School of Public Health, Maryland Robotics Center, NACS Program, University of Maryland, College Park, MD 20742 USA (corresponding author, phone: 301-405-2490; fax: 301-405-2458; e-mail: rodolphe@umd.edu).

J.C. Rietschel is with the Veterans Health Administration, Maryland Exercise and Robotics Center of Excellence, Baltimore, MD 21201, USA (e-mail: jcrietschel@yahoo.com).

K. J. Jaquess, LC Lo, H. Oh and YY Tan are with the Department of Kinesiology, School of Public Health, NACS Program, University of Maryland, College Park, MD 20742 USA (e-mail: kjaquess@umd.edu; llo@umd.edu; hyukoh@umd.edu; ying2@umd.edu).

CAPT M. Prevost, J.M. Mohler are with the United States Naval Academy (USNA), Annapolis, MD 21402 USA (e-mail: prevost@usna.edu; mohler@usna.edu).

M. M. Miller is with the Department of Kinesiology, Auburn University, Auburn, AL 36849 USA (e-mail: mwm0024@auburn.edu).

B. D. Hatfield is with the Department of Kinesiology, School of Public Health, NACS Program, University of Maryland, University of Maryland, College Park, MD 20742 USA (e-mail: bhatfiel@umd.edu).

training in basic flight instruments enrolled in this study. They performed a visuo-motor task in a flight simulator (Prepar3D® v1.4, T-6A Texan II SP2 USN aircraft, Lockheed Martin Corporation™) at the USNA in three scenarios each having a different task demand. Each scenario was composed of a 1-minute familiarization period followed by a 10-minute flight challenge during which performance (i.e., airspeed, altitude, heading, and vertical speed) was recorded. The three progressively challenging scenarios (S1, S2, and S3) were defined as follows: i) S1 (low demand): The goal was to maintain the aircraft's current altitude, heading, and airspeed while maintaining a straight and level course. The weather was defined by the absence of clouds, precipitation, and wind with unlimited visibility; ii) S2 (medium demand): The goal was to maintain the aircraft's current heading, airspeed, and a "wings-level" attitude while continuously making assigned altitude changes within a specified ascent and descent rate. The weather contained no precipitation or wind, but visibility was zero; iii) S3 (high demand): The goal was to maintain the aircraft's current airspeed, while constantly changing the heading at a 15-degree angle of bank, ascending while turning right and descending while turning left. Visibility was zero, as in S2, with no precipitation, but with a moderate wind. These three scenarios were selected from predefined flight training challenges with minor alterations developed by experienced pilots from the USNA.

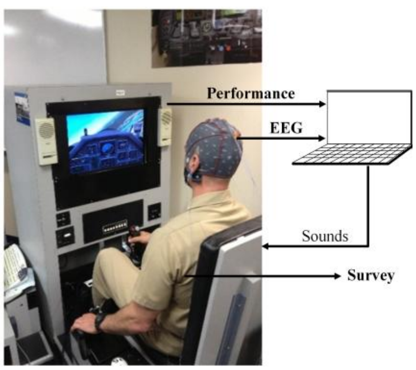

Figure 1: Experimental set-up. Participants executed three flight scenarios in a simulator while wearing the dry EEG cap, ECG sensor, and ear-bud speakers. Participants controlled the aircraft using the control stick (between knees) with the right hand, the throttle with left hand, and the rudder pedals with both legs. Performance, EEG and ECG were recorded while probe sounds (i.e., to-be-ignored) were delivered.

The sequence of scenarios was counter-balanced. Novel sounds were generated similarly to the method used by Miller et al. [8] while using ear-buds instead of external speakers.

First, participants filled out an informed consent form approved by the Institutional Review Board of the University of Maryland and the USNA and a handedness survey. Then, during a familiarization period, participants practiced the task for 5 minutes along with exposure to the novel sounds. Participants were then prepared for the placement of the electrocardiogram (ECG) sensor and dry EEG cap according to the 10-20 system. An initial scenario (challenge) was then provided with relevant instructions to the participants. After this setup period, the first 10-minute scenario with presentation of the novel sound stimuli was executed. Participants were then provided the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) survey to report their subjective experience of cognitive load. Also, a visual analog scale (VAS) was used to assess perceived mental load and task difficulty. The same order of procedures was followed until all three challenge levels were completed. The order of conditions was counterbalanced.

B. Data Acquisition

In addition to the subjective reports (NASA TLX, VAS) and the flight performance captured by a computerized log system, the ECG and EEG were continuously recorded during the entire study. The EEG system was calibrated, and electrode impedances were maintained below 5 kOhm. Both EEG and ECG were sampled at a rate of 512 Hz. The right mastoid was employed as the ground and the EEG recording was accomplished using the left ear as the reference. An online band-pass filter was applied with a range of 0.01 Hz to 40 Hz. Four dry-sensor EEG (g-Tech™, Schiedlberg, Austria) signals were collected from sites along the frontal (Fz), fronto-central (FCz) central (Cz), and parietal (Pz) midline. The ECG record was achieved by placing a sensor below the bottom left rib using a unipolar configuration (one electrode) with common reference and ground to the EEG.

C. Data Processing: NASA TLX and VAS

A series of one-way ANOVAs was employed to test the participants' subjective mental workload measured with the NASA TLX and VAS for each scenario. For all subsequent ANOVAs reported in this article, the Greenhouse-Geisser correction was employed when sphericity was violated.

D. Data Processing: Flight performance

In each scenario, acceptable performance criteria were predefined as deviations of the flight status (e.g., altitude, airspeed, heading, bank angle, etc.) from the tolerance limits (i.e., goals) defined by the experienced naval pilots. For each metric (airspeed, altitude, heading, and vertical speed), the behavioral performance was reconstructed and a composite performance index scaled between 0 (worst) and 1 (best) was computed. Also, change in the performance index between S3 and S1 as well as between S2 and S1 were calculated. One-way ANOVAs were subsequently calculated to assess any changes in both the performance index and for the percentage drop between the various levels of challenge.

E. *Data Processing: EEG*

The data were re-referenced to an averaged-ears montage and then filtered using a 20-Hz low pass filter, 48 dB rolloff. Next, ERPs to the novel sounds were generated to estimate attentional reserve. One-second epochs that were time-locked to the auditory stimuli were extracted from the EEG signals and were mean baseline-corrected using the pre-stimulus interval (i.e. -100-0ms). The transformed data were then visually inspected and those epochs retaining significant artifact were excluded from further analyses. The remaining epochs were averaged for each of the three conditions, resulting in three components derived from the ERPs generated for each individual. Finally, the average amplitude for the P3 (270-370 ms) component was derived. One-way ANOVAs were computed separately for each electrode to assess differences in ERP amplitudes for each challenge.

To estimate cognitive workload, spectral analysis was used using a Fast Fourier transform (1-Hz resolution). The spectral power was then computed for each scenario. The frequency bins were log-transformed and summed to obtain spectral power for the Theta (3-8 Hz) / Alpha (8-13 Hz) ratio. A one-way ANOVA was employed to assess differences in the power ratios for the three levels of challenge.

F. Data Processing: HRV

The HRV was computed according to previous methods used by the Task Force of The European Society of Cardiology and The North American Society of Pacing. Namely, the square root of the mean squared successive differences (RMSSD) was calculated. The mean squared differences before and after each interval were calculated and then the square root value was computed. A one-way ANOVA was used to assess differences in the HRV for the three levels of challenge. To assess the HRV in relationship to perceived workload and task difficulty, one-tailed t-tests between the conditions were used for each of the VAS scales.

III. RESULTS

A. Behavioral assessment

As the demand increased, the performance index significantly decreased ($p<0.001$) (Fig. 2A).

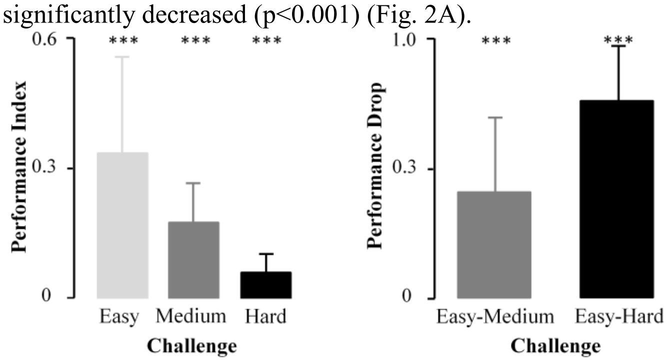

Figure 2: Pilots' performance based on the index generated during each scenario (S1: low demand; S2: medium demand and S3: high demand). Left panel: Performance index is a composite of all the metrics computed for each scenario. Right panel: Changes in performance expressed as percentage drop relative to S1. For this figure as well as Figures 3 and 4, stars without fork indicate that all the comparisons between the levels of challenge are significantly different; otherwise the fork indicates the specific significant contrast. *:p<0.05; **:p<0.01; ***:p<0.001.

Pilots' relative performance (percentage change) for S2 and S3 with respect to S1 significantly decreased ($p<0.001$) (Fig. 2B). The findings revealed that participants' responses to challenge generally differed across the three scenarios on the items of VAS and NASA TLX. Post-hoc analyses showed that participant ratings of load increased, as expected, as the challenge increased. However, comparisons between the easy and medium levels on Ease and Physical Demand were not statistically significant (Fig. 3).

B. Physiological assessment

Although differences in the RMSSD were not significant when comparing the three scenarios ($p>0.05$), it was found to be significantly lower for the individuals who perceived the task as demanding a high degree of concentration (Fig. 4A,B; $p<0.05$), which indicates a reduction of vagal influence on the heart under elevated load.

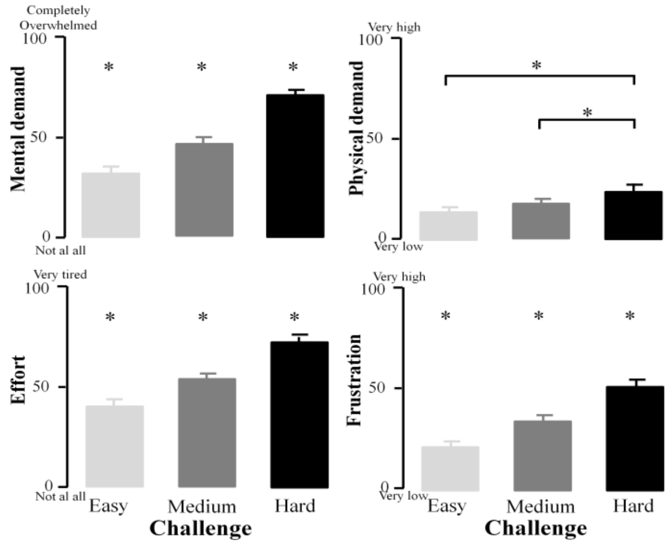

Figure 3: Participants' responses for each of the three levels of challenge for the assessment to the level of mental demand, physical demand, effort and frustration using the NASA TLX.

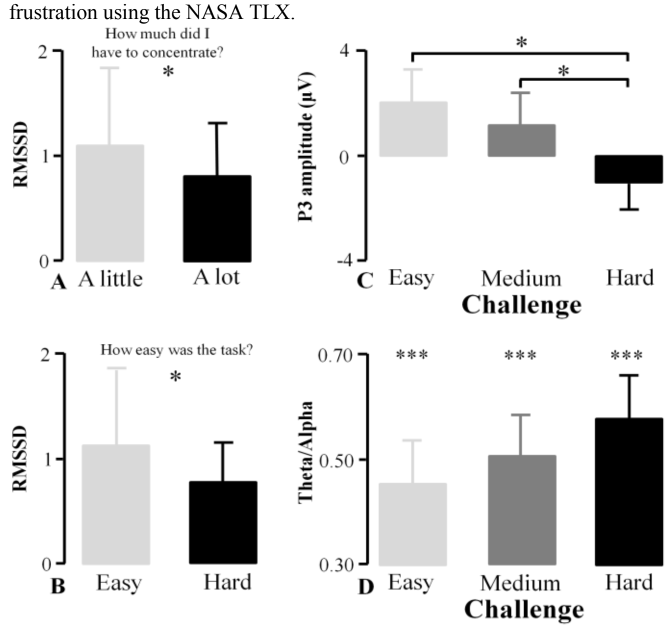

Figure 4: (A-B) RMSSD metrics to assess HRV during the easy, medium and hard scenarios. (C) Amplitude of P3 response for the site Pz and (D) Theta/alpha EEG power ratio for each of the three levels of challenge (i.e., easy, medium, and hard).

C. Brain dynamics assessment

The amplitude of the P3 component at site Pz revealed a robust significant difference ($p<0.05$). As expected, the magnitude of the ERP amplitude was higher for the Easy compared to Hard (S1 vs. S3: $p<0.05$) scenarios and Medium compared to Hard (S2 vs. S3: $p<0.05$) scenarios (Fig. 4C). EEG spectral analysis revealed that the theta/alpha ratio increased as the challenge increased; significant differences were found between all levels of challenge (S1 vs. S2:$p<0.05$; S2 vs. S3:$p<0.01$; S1 vs. S3:$p<0.001$;Fig. 4D)

IV. DISCUSSION

This study confirms and extends previous research by identifying multiple biomarkers, and particularly those derived from EEG by combining a newly developed experimental method and the use of dry EEG to assess cognitive workload and attentional reserve during progressive simulated piloting task demands. EEG power and HRV were used to assess cognitive workload while the ERPs assessed attentional reserve. Cognitive workload and attentional reserve were negatively related.

Specifically, elevations in flight simulator task demands were positively associated with subjective reports of mental workload and negatively related to performance. EEG provided two important markers. First, the theta/alpha power ratio was positively related to workload. This ratio was sensitive to the three levels of challenge and able to effectively discriminate between them. Secondly, the P300 amplitude recorded in the parietal region indicated lower attentional reserve in S2 (medium demand) compared to S1 (low demand) as well as in S3 (high demand) compared to S1 (low demand). Overall, these EEG findings confirm and extend those from previous studies in two directions: i) identifying EEG biomarkers for an aircraft piloting task using only four dry EEG sensors [2,7,9] and ii) extension of a recent method to assess attentional reserve, as previously used in a laboratory setting, to a more ecologically valid performance environment using few dry EEG sensors [8,10]. However, additional biomarkers must be considered in the future since it has been suggested that simple metrics such as EEG power (for various frequency bands and their ratios) can be modified by other mental states such as sleep deprivation. This concern suggests a potential distortion of the assessment of cognitive workload [2,11]. Also, compared to EEG, HRV was less sensitive to changes in absolute cognitive workload, but it was related to perceived workload as measured by self-report. In particular, elevated HRV was indicative of elevated vagal influence to the heart, believed to be protective of cardiac stress, and was characteristic of participants who perceived less effort during the imposed workload. The biomarkers derived from EEG recorded with only four dry EEG sensors, were sensitive to various levels of cognitive workload. This finding is important since a dry EEG system composed of only few sensors does not require using conductive gel and the very limited number of sensors can reduce the preparatory and computational burden, which is promising for operational settings (e.g., cockpits and unmanned aerial vehicles). A similar approach could be used for other fields of research such as assistive technology in a cognitive-motor rehabilitation context as with amputee populations.

V. Conclusion

As previously mentioned, this work is a first step towards a programmatic effort to include additional biomarkers to assess cognitive workload in operational settings. Although this present work includes multiple markers within data modalities such as the time and frequency domains (ERP and spectral metrics, respectively), as well as between HRV and EEG, future research in this area will also include eye-tracking technology (e.g., gaze behavior and pupillometry) and functional near-infrared spectroscopy to assess the state of the brain since it was shown that both systems were able to provide biomarkers sensitive to cognitive workload fluctuation [7,12-15]. The idea that a multiplicity of markers could lead to a sensitive and robust composite biomarker to assess the cognitive workload and attentional reserve forms the basis of this research. Such a composite biomarker could then be used for the classification of cognitive workload and operator state using machine learning. This work will include data mining and machine learning algorithms combined for feature selection and classification while considering multiple time scales [16].

References

[1] C. Wickens, A. Kramer, L. Vanasse, E. Donchin, "Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources" *Science* vol.221, pp1080–82, 1983.

[2] C. Berka, D.J. Levendowski, M.N. Lumicao, A. Yau, G. Davis, V.T. Zivkovic, R.E. Olmstead, P.D. Tremoulet, and P.L. Craven, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviat Space Environ Med*, vol. 78(5), pp. B231-44, 2007.

[3] C. D. Wickens, "Situation awareness and workload in aviation," *Current Directions in Psychological Science*, vol.11, pp.128-33, 2002.

[4] F. Di Nocera, M. Camilli, and M. Terenzi, "A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload," *Journal of Cognitive Engineering and Decision Making*, vol. 1(3), pp 271-285, 2007.

[5] C. Dussault, J.C. Jouanin, M. Philippe, and C.Y. Guezennec, "EEG and ECG changes during simulator operation reflect mental workload and vigilance," *Aviat Space Environ Med*, vol 76(4), pp. 344-51, 2005.

[6] D. Schmorrow, K.M. Stanney, G.F. Wilson, and P. Young, "Augmented cognition in human-system interaction", in G. Salvendy, Ed. *Handbook of human factors and ergonomics*, John Wiley: New York, 2005.

[7] G. F. Wilson, "An analysis of mental workload in pilots during flight using multiple psychophysiological measures," *International Journal of Aviation Psychology*, vol. 12, pp. 3–18, 2002.

[8] M.W. Miller, J.C. Rietschel, C.G. McDonald, and B.D. Hatfield, "A novel approach to the physiological measurement of mental workload". *Int. J. Psychophysiol*, vol 80(1), pp. 75–78, 2011.

[9] L.J. Prinzel, F.G. Freeman, M.W. Scerbo, P.J. Mikulka, and A.T. Pope, "A closed-loop system for examining psychophysiological measures for adaptive task allocation," *International Journal of Aviation Psychology*, vol. 10, pp. 393–410, 2000.

[10] J.C. Rietschel, M.W. Miller, R.J. Gentili, R.N. Goodman, C.G. McDonald and B.D. Hatfield, "Cerebral-cortical networking and activation increase as a function of cognitive-motor task difficulty," *Biol Psychol*, vol. 90(2), pp. 127–133, 2012.

[11] M.E. Smith, and A. Gevins, "Neurophysiologic monitoring of mental workload and fatigue during operation of a flight simulator," *Proc. SPIE Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations*, Orlando, 2005, pp. 116–26.

[12] H. Ayaz, P.A. Shewokis, S. Bunce, K. Izzetoglu, B Willems, and B. Onaral, "Optical brain monitoring for operator training and mental workload assessment," *Neuroimage*, vol. 59(1), pp. 36-47, 2012.

[13] G.F. Wilson, and C.A. Russell, "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Hum Factors*, vol. 45, pp. 635-643, 2003.

[14] L.J. Prinzel, F.G. Freeman, M.W. Scerbo, P.J. Mikulka, and A.T. Pope, "Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component," *Hum Factors*, vol. 45, pp. 601–13, 2003.

[15] C.A. Russell, G.F. Wilson, M.M. Rizki, T. S. Webb, and S.C. Gustafson, "Comparing classifiers for real time estimation of cognitive workload", in *Found. Aug. Cog.*, D.D. Schmorrow, Ed. Lawrence Erlbaum Associates: Mahwah, 2005, vol. 11, pp. 396-404.

[16] G. Davis, D. Popovic, R. R. Johnson, C. Berka and M. Mitrovic, "Building dependable EEG classifiers for the real world – It's not just about the hardware", in *Found. Aug. Cog.*, D.D. Schmorrow et al., Eds. Springer: Heidelberg, vol. 5638, 2009, pp. 355-364.