

Psychophysiological Monitoring of Aerospace Crew State

Grace C. Wusk
Virginia Tech
Blacksburg, Virginia
gracew@vt.edu

Andrew F. Abercromby
NASA Johnson Space Center
Houston, Texas
andrew.f.abercromby@nasa.gov

Hampton C. Gabler
Virginia Tech
Blacksburg, Virginia
gabler@vt.edu

ABSTRACT

As next-generation space exploration missions necessitate increasingly autonomous systems, there is a critical need to better detect and anticipate crewmember interactions with these systems. The success of present and future autonomous technology in exploration spaceflight is ultimately dependent upon safe and efficient interaction with the human operator. Optimal interaction is particularly important for surface missions during highly coordinated extravehicular activity (EVA), which consists of high physical and cognitive demands with limited ground support. Crew functional state may be affected by a number of variables including workload, stress, and motivation. Real-time assessments of crew state that do not require a crewmember's time and attention to complete will be especially important to assess operational performance and behavioral health during flight. In response to the need for objective, passive assessment of crew state, the aim of this work is to develop an accurate and precise prediction model of human functional state for surface EVA using multi-modal psychophysiological sensing. The psychophysiological monitoring approach relies on extracting a set of features from physiological signals and using these features to classify an operator's cognitive state. This work aims to compile a non-invasive sensor suite to collect physiological data in real-time. Training data during cognitive and more complex functional tasks will be used to develop a classifier to discriminate high and low cognitive workload crew states. The classifier will then be tested in an operationally relevant EVA simulation to predict cognitive workload over time. Once a crew state is determined, further research into specific countermeasures, such as decision support systems, would be necessary to optimize the automation and improve crew state and operational performance.

KEYWORDS

psychophysiology; machine learning; extravehicular activity

ACM Reference Format:

Grace C. Wusk, Andrew F. Abercromby, and Hampton C. Gabler. 2019. Psychophysiological Monitoring of Aerospace Crew State. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct)*, September 9–13, 2019, London.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6869-8/19/09...\$15.00

<https://doi.org/10.1145/3341162.3349309>

United Kingdom. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3341162.3349309>

1 PROBLEM STATEMENT

Human exploration missions to the moon and Mars will challenge crewmembers with never-before-seen physical and cognitive workloads. For journeys to Mars, communication latencies with Earth will require many functions to be controlled by in-flight hardware, software, and crewmembers. In addition, the frequency of extravehicular activity (EVA), or spacewalks, for future surface exploration operations will far exceed the hours of Apollo surface operations. Next-generation missions will require increasingly autonomous systems to lighten crewmembers' heavy task loads and to facilitate smooth operations with limited ground support. Safe and optimal operations necessitate a better understanding of the human component of the system. There is a critical need to characterize the effects of physical and cognitive workload on crew functional state, specifically during surface EVA. Operator functional state is defined as the ability of an operator to complete a task at a moment in time, and is affected by an operator's cognition and affect. Monitoring functional state is especially important for crewmembers as errors in spaceflight may have particularly drastic consequences. EVA operations are highly coordinated events and require extensive communication between the extravehicular and intravehicular crewmembers. The physical and cognitive workloads vary between these two positions. Through better characterization of these workloads, countermeasures such as decision support systems can be optimized to assist crewmembers with limited ground support. Ultimately, closing the loop between human and system with real-time crew state monitoring and analysis would allow the system to provide feedback, transfer control, and judge fitness for work, enhancing the safety and reliability of human exploration missions.

The physical response of EVA has been well studied in simulated and analog environments including simulations in the Active Response Gravity Offload System (ARGOS) and trainings in the Neutral Buoyancy Laboratory (NBL). Metabolic and heart rate data are also available from flight operations. However, compared to the current microgravity EVA on the International Space Station (ISS), physiological profiles will differ for the altered gravity environments of the lunar and Martian surfaces with 1/6 and 1/3 of Earth's gravity, respectively, and require further characterization. The cognitive and interactive cognitive-physical responses of EVA, on the other hand, have not been as well studied. Apollo crewmembers have highlighted the importance of understanding the cognitive load during surface EVA, "Consider mental and physical fatigue here separately, Although there was not a log of physical fatigue [during the lunar activity], the mind was being used quite a bit. You can sometimes wear your brain out before your body is fatigued," [Scheuring et al. 2007]. Currently, cognitive and stress

testing is performed on crewmembers using subjective measures, such as visual analog scales, and objective measures, such as the Psychomotor Vigilance Test (PVT) and the Cognition Test Battery [Williams 2016]. While these standard measures provide valuable metrics for assessing crew state, they require a crewmember's time and attention to complete and cannot assess real-time operations. Additionally, flight surgeons closely watch and constantly assess crewmembers on board the ISS. Every minute of work scheduled beyond a crewmember's allowable amount of work time must be approved by the flight surgeon. This subjective decision is informed by the flight surgeon's collective knowledge of the crewmember's psychological and physical well-being. However, for future exploration missions it will be extremely difficult for flight surgeons to constantly assess crewmembers' workloads due to communication delays and distance from Earth. There is a need for a passive, objective tool to assess crew state in real-time in the increasingly autonomous environments. A psychophysiological monitoring approach allows for real-time classification of an operator's cognitive state using features derived from non-invasive biosignals, with the ultimate goal of adapting the system to fit the needs of the operator.

2 RELATED WORK

Physiological computing builds upon the field of psychophysiology to classify psychological state from physiological signals in real-time using data fusion methods. Physiological responses, primarily autonomic nervous system (ANS), hemodynamic, and electrophysiological responses, are recorded through physiological signals. Psychophysiological features are then extracted from the physiological signals to train a model.

The psychophysiological monitoring approach has been used to predict stress and workload in many previous studies [Bonarini et al. 2008; Das et al. 2017; Saha et al. 2017]. The United States Air Force Research Laboratory has used psychophysiological features to predict functional state during multiple flight tasks, including the Multi-Attribute Task Battery (MATB) and an air traffic control task [Christensen et al. 2012; Wilson and Russell 2003a,b]. Crew state monitoring using multi-modal physiological sensing has also been tested at NASA Langley Research Center in flight simulation cockpits [Harrivel et al. 2016, 2017]. Specifically, the crew state monitoring team was able to achieve an average multi-state prediction accuracy of 88.6% using electroencephalography (EEG), galvanic skin response (GSR), and heart rate variability (HRV) with a subject-dependent model. A series of Attention-related Human Performance Limiting States (AHPLS) defined the classes and were trained by a set of benchmark tasks including the MATB. The crew states were chosen in the experiment because they have been determined to cause pilots to lose airplane state awareness.

Real-time physiological assessment is also of interest to the military. The US Army created and tested a tool to assess physical compensatory reserve in real-time, using machine learning and a blood volume pulse (BVP) signal [Convertino et al. 2015]. From an individual's BVP signal, the system estimates physical reserve. While this work focuses on physical state, similar machine learning techniques may be used to classify cognitive state.

While there is a large body of literature on physiological computing, both cognitive and affective, a large knowledge gap remains in

applying these methods to an operationally-relevant surface EVA scenario, particularly in a simulation with integrated performance metrics to validate the labeled crew states.

3 METHODOLOGY

3.1 Compile Sensor Suite

Compilation of the sensor suite is largely completed. There are a number of physiological signals that measure ANS response including electrocardiography (ECG), photoplethysmography (PPG), electrodermal activity (EDA), skin temperature, and respiration measurements. Additionally, electroencephalography (EEG) provides a more direct measure of brain activity. To facilitate real-time, continuous monitoring, this work will focus on non-invasive sensing techniques. Commercially available devices such as the Interaxon Muse and Empatica E4 offer inexpensive, unobtrusive means of measuring physiological signals. The Muse is a four-lead dry EEG system, capable of collecting an additional one-lead ECG. The E4 is a wrist-worn device with EDA, PPG, and skin temperature sensors. The Interaxon Muse and the Empatica E4 have been synchronized in the sensor suite using the Lab Streaming Layer (LSL). LSL is an open source system for unifying time series data collection in research experiments. The commercial off-the-shelf (COTS) devices provide potential to apply the findings of this work and the development of the classification tool to other human-computer interaction research. However, data quality is a concern and laboratory grade equipment may also be tested. After collecting the physiological signals, a set of psychophysiological features are extracted including EEG bandpowers, heart rate variability (HRV) time-domain and frequency-domain features, and phasic and tonic skin conductance features from EDA. In addition to ANS and electrophysiological responses, psychophysiological features may be extracted from position, velocity, and acceleration signals recorded by the body-worn devices.

3.2 Propose Target Crew States for Classification

One of the challenges in crew state monitoring is defining a suitable set of target states. The ideal classifier for human functional state in operational environments has many characteristics. First and foremost, the ideal classifier classifies cognitive workload using psychophysiological features in real-time. There are many variables that can affect cognitive workload, including stress and motivation. Therefore, the ideal classifier is also able to account for emotional state. Ideally, a target state is generalizable across time, subjects, and situations [Haynes and Rees 2006]. However, subtle contextual differences may alter mental states. This requires classification algorithms to maintain a degree of flexibility and ignore irrelevant differences between mental states. For this project, the target crew states for classification are high and low cognitive workload. Performance metrics and subjective assessments will be used to validate the low and high cognitive workload target crew states for specific tasks. In addition, timeline analysis will be used to estimate workload percentage over time based on the average number of tasks per unit time. The quantifiable performance metrics, subjective assessments, and timeline analysis estimates will be extremely important

to characterize the binary, discrete labels of high and low cognitive workload used to train the supervised machine learning models.

3.3 Collect Training Data

Physiological signals will be recorded using the non-invasive sensor suite during a set of benchmark tasks to train the model. Benchmark tasks include high and low workload cognitive and functional tasks. The Paced Stroop Test (PST) will be used as the simple cognitive task to train the model. The PST is a widely used neuropsychological test designed to assess the ability to inhibit cognitive interference. Color-words (i.e. blue) are presented in either congruent or incongruent conditions and participants are required to name the color of the word rather than reading the word. The incongruent condition, in which the color-word and color of the word do not match, requires participants to inhibit cognitive interference and represents the high cognitive workload condition. The congruent condition represents low cognitive workload. In addition to the PST, the Multi-Attribute Task Battery (MATB) will be used as a more complex, functional benchmark task. The MATB is a desktop flight simulator in which operators simultaneously control a joystick and a mouse. The MATB is a well-validated tool complete with system monitoring, tracking, resource management, and communications tasks that can be adjusted in frequency and difficulty to simulate high and low workload flight events [Comstock and Arnegard 1992]. Both the PST and the MATB have been used in previous psychophysiological monitoring studies [Das et al. 2017; Harrivel et al. 2016; Saha et al. 2017; Wilson and Russell 2003b]. As mentioned previously, subjective assessments, namely the NASA Task Load Index (TLX), and performance metrics of speed and accuracy from the PST and MATB trials will be used to validate the high and low workload labels. Preliminary data has been collected from nine subjects completing the high and low workload MATB events.

3.4 Develop Classifiers

The first step to develop a classifier from the training data is to extract psychophysiological features. Ideally, these features are then normalized and reduced. From the preliminary MATB data collection using the Muse, a total of 26 features were extracted from the EEG and ECG signals in 10-second 50% overlap windows, including relative bandpowers from the four EEG channels and time-domain and frequency-domain HRV features from the ECG signal. The bandpowers included delta (1-4 Hz), theta (4-8 Hz), alpha (7.5-13 Hz), beta (13-30 Hz), and gamma (30-44 Hz). The HRV features included mean R-R intervals, standard deviation of R-R intervals, square root of mean squared difference of successive R-R intervals, low frequency power (0.04-0.15 Hz), high frequency power (0.15-0.4 Hz), and low to high frequency power ratio. The features were normalized using three techniques including subtracting baseline features, dividing baseline features, and adding baseline features to the feature set as a baseline matrix.

Nine different binary classification models have been evaluated on the preliminary MATB-Muse dataset, including Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Classification and Regression Tree (CART), Random Forests (RF), Naive Bayes (NB), Support Vector Machines linear (SVML),

SVM Gaussian kernel (SVMG), and SVM sigmoid kernel (SVMS). Subject-specific and population-based training schemes were evaluated for the preliminary dataset. While population-based training is more generalizable, creating a participant-specific model would still be useful for the application area as there is a small sample of crewmembers. The binary (high/low workload) classifiers were cross-validated with five, stratified folds. The subject-dependent logistic regression model performed the best on the preliminary dataset. Average logistic regression accuracy across the 5-folds varied between 51% and 88% for the nine subjects. Principal Component Analysis (PCA) was used to visualize the high dimensional dataset.

The approach used on the preliminary dataset from the Muse device alone will be expanded in future studies to include other devices, including the E4. The process of feature extraction, normalization, and reduction will be repeated with future datasets. The algorithm development is expected to iterate over time.

3.5 Test Classifiers in EVA Simulation

To apply the knowledge gained in the model development, the classifier will be tested on an operationally relevant EVA simulation. A typical surface EVA may be divided into three parts, overhead activity such as exiting an airlock, translation activity such as ambulating to a destination, and station activity such as deploying a science instrument [Miller et al. 2017]. The translation activity will be the focus of this work. Specific measures and metrics of performance will be embedded into the simulated surface EVA translation task. The cognitive demands of the task will be mapped to specific macrocognitive functions and cognitive constructs. The cognitive demands will be manipulated to simulate high and low cognitive workload conditions, with workload percentage estimated using timeline analysis. Physiological data, performance metrics, and subjective assessments will be recorded from participants completing the high and low cognitive workload translation tasks to validate the two workload conditions. The population-based and subject-dependent crew state models trained on the PST and MATB benchmark tasks will be used to predict workload during the translation tasks.

4 EVALUATION

In terms of data analysis, the models are ultimately limited by the quality of the data from the wearable devices. In part, better understanding the limits of the devices is included in the research objectives. However, the main goal is not to validate the devices but to validate the psychophysiological features. The assumption is that the signal quality will continue to improve with new and better technology.

The crew state classification model will be evaluated during the training phase with cross-validation and during the testing phase using the simulated EVA environment. The performance of the classifiers will be assessed by multiple metrics including accuracy, precision, recall, and F1 score of the test data. The workload will be estimated using timeline analysis and validated with performance metrics and subjective assessments, supporting the ground truth labels of crew functional state.

The human subject population is a limitation of this work. Crewmembers are highly trained individuals, whereas participants in the proposed studies may not be as experienced with piloting and surface EVA tasks. Of course, future exploration missions will undoubtedly challenge crewmembers with unforeseen tasks, as it is impossible to prepare for every situation. Training trials will be included in all of the studies to mitigate effects of learning and experience. In addition to training, the level of cognitive workload and performance on the tasks is highly influenced by motivation. An extremely high level of performance is demanded in every spaceflight mission to ensure the success of the mission and the safety of the crew. It is difficult to simulate this level of pressure in the laboratory.

5 EXPECTED CONTRIBUTION

Compiling the sensor suite will result in a robust, non-invasive set of sensors to collect accurate and precise physiological measurements. The exact set of sensors used for classification is subject to change depending on the test environment and the results of the dimension reduction. The goal is to use the training data collection with the benchmark tasks to prepare for the more EVA-relevant simulations. Iteration will be important to build the classification algorithm.

Training and testing the model will provide a continuous classification variable of crew functional state. Trends in the physiological data during high and low cognitive workload tasks will be evaluated. Population-based and subject-specific models will be compared.

Using this continuous variable of crew state, future research could assess the means and efficacy of altering the automation to provide feedback, transfer control, and make judgments about 'fitness for work'. Overall, these strategies have been proposed and tested to various extents in operationally-relevant environments. However, each of these strategies rely on information of crew cognitive state in real-time. Given an estimate of crew functional state, increasingly autonomous systems could be employed to train crewmembers on the ground and assist them on future surface EVAs on the moon and Mars. In addition to training and aiding crewmembers, general knowledge of crew functional state could be used to test and evaluate new systems, such as heads-up-displays (HUDs), and to better understand the phenomena behind the occurrence of sub-optimal crew state. In this way, the engineering and design community, as well as, the scientific community would benefit from the real-time, objective assessment of crew cognitive state and performance.

6 BIOGRAPHICAL SKETCH

Grace C. Wusk, B.S., is a Ph.D. candidate in Biomedical Engineering and Mechanics at Virginia Tech. Grace is a NASA Space Technology Fellow (NSTRF) collaborating with the Human Physiology, Performance, Protection, and Operations (H-3PO) Laboratory at NASA Johnson Space Center (JSC). Grace began her Ph.D. in fall 2016 and is expected to complete her Ph.D. in spring 2021. Hampton C. Gabler, Ph.D., is Grace's faculty advisor at Virginia Tech within the Department of Biomedical Engineering and Mechanics. Andrew F. J. Abercromby, Ph.D., is Grace's NASA research collaborator and lead of the H-3PO Laboratory at NASA JSC.

ACKNOWLEDGMENTS

This work was supported by a NASA Space Technology Research Fellowship.

REFERENCES

- Andrea Bonarini, Luca Mainardi, Matteo Matteucci, Simone Tognetti, and Roberto Colombo. 2008. Stress Recognition in a Robotic Rehabilitation Task. In *"Robotic Helpers: User Interaction, Interfaces and Companions in Assistive and Therapy Robotics"*, A Workshop at ACM/IEEE HRI. Amsterdam, Netherlands, 41–48.
- James C. Christensen, Justin R. Estep, Glenn F. Wilson, and Christopher A. Russell. 2012. The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage* 59, 1 (1 2012), 57–63. <https://doi.org/10.1016/j.neuroimage.2011.07.091>
- Jr. Comstock, J. Raymond and Ruth J. Arnegard. 1992. The multi-attribute task battery for human operator workload and strategic behavior research. (1 1992). <https://ntrs.nasa.gov/search.jsp?R=19920007912>
- Victor A. Convertino, Jeffrey T. Howard, Carmen Hinojosa-Laborde, Sylvain Cardin, Paul Batchelder, Jane Mulligan, Gregory Z. Grudic, Steven L. Moulton, and David B. MacLeod. 2015. Individual-specific, beat-to-beat trending of significant human blood loss: The compensatory reserve. *Shock* 44, 8 (2015), 27–32. <https://doi.org/10.1097/SHK.0000000000000323>
- Deepan Das, Tanuka Bhattacharjee, Shreyasi Datta, Anirban Dutta Choudhury, Pratyusha Das, and Arpan Pal. 2017. Classification and Quantitative Estimation of Cognitive Stress from In-Game Keystroke Analysis using EEG and GSR. In *IEEE Life Sciences Conference*. IEEE, 286–291. <https://doi.org/10.1109/LSC.2017.8268199>
- Angela R. Harrivel, Charles Liles, Chad L. Stephens, Kyle K. Ellis, Lance J. Prinzel, and Alan T. Pope. 2016. Psychophysiological Sensing and State Classification for Attention Management in Commercial Aviation. In *AIAA Infotech @ Aerospace*. 1–8. <https://doi.org/10.2514/6.2016-1490>
- Angela R. Harrivel, Chad L. Stephens, Robert J. Milletich, Christina M. Heinich, Mary Carolyn Last, Nicholas J. Napoli, Nijo Abraham, Lance J. Prinzel, Mark A. Motter, and Alan T. Pope. 2017. Prediction of Cognitive States during Flight Simulation using Multimodal Psychophysiological Sensing. In *AIAA Information Systems-AIAA Infotech @ Aerospace*. 1–10. <https://doi.org/10.2514/6.2017-1135>
- John-Dylan Haynes and Geraint Rees. 2006. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* 7, 7 (7 2006), 523–534. <https://doi.org/10.1038/nrn1931>
- Matthew James Miller, Austin Claybrook, Jessica J Marquez, and Karen M Feigh. 2017. Operational Assessment of Apollo Lunar Surface Extravehicular Activity. *NASA TP2017-21*, July (2017). <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170007261.pdf>
- Deba Pratim Saha, R. Benjamin Knapp, and Thomas L. Martin. 2017. Affective feedback in a virtual reality based intelligent supermarket. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing and ACM International Symposium on Wearable Computers*. 646–653. <https://doi.org/10.1145/3123024.3124426>
- Richard A Scheuring, David B Gillis, Josef Schmid, James M Duncan, Jeffrey R Davis, and Joseph D Novak. 2007. The Apollo Medical Operations Project: Recommendations to Improve Crew Health and Performance for Future Exploration Missions and Lunar Surface Operations. (2007). <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20070030109.pdf>
- Thomas Williams. 2016. Behavioral Health and Performance Laboratory Standard Measures. (2016). <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170005421.pdf>
- Glenn F. Wilson and Christopher A. Russell. 2003a. Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 45, 3 (9 2003), 381–389. <https://doi.org/10.1518/hfes.45.3.381.27252>
- Glenn F. Wilson and Christopher A. Russell. 2003b. Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 45, 4 (2003), 635–644. <https://doi.org/10.1518/hfes.45.4.635.27088>