

Psychophysiological Sensing and State Classification for Attention Management in Commercial Aviation

Angela R. Harrivel¹, Charles Liles², Chad L. Stephens³, Kyle K. Ellis⁴, Lawrence J. Prinzel⁵, and Alan T. Pope⁶
NASA Langley Research Center, Hampton, Virginia, 23681

Attention-related human performance limiting states (AHPLS) can cause pilots to lose airplane state awareness (ASA), and their detection is important to improving commercial aviation safety. The Commercial Aviation Safety Team found that the majority of recent international commercial aviation accidents attributable to loss of control inflight involved flight crew loss of airplane state awareness, and that distraction of various forms was involved in all of them. Research on AHPLS, including channelized attention, diverted attention, startle / surprise, and confirmation bias, has been recommended in a Safety Enhancement (SE) entitled “Training for Attention Management.” To accomplish the detection of such cognitive and psychophysiological states, a broad suite of sensors has been implemented to simultaneously measure their physiological markers during high fidelity flight simulation human subject studies. Pilot participants were asked to perform benchmark tasks and experimental flight scenarios designed to induce AHPLS. Pattern classification was employed to distinguish the AHPLS induced by the benchmark tasks. Unimodal classification using pre-processed electroencephalography (EEG) signals as input features to extreme gradient boosting, random forest and deep neural network multiclass classifiers was implemented. Multi-modal classification using galvanic skin response (GSR) in addition to the same EEG signals and using the same types of classifiers produced increased accuracy with respect to the unimodal case (90% vs. 86%), although only via the deep neural network classifier. These initial results are a first step toward the goal of demonstrating simultaneous real time classification of multiple states using multiple sensing modalities in high-fidelity flight simulators. This detection is intended to support and inform training methods under development to mitigate the loss of ASA and thus reduce accidents and incidents.

I. Introduction

ATTENTION-related human performance limiting states (AHPLS) can cause pilots to lose airplane state awareness (ASA), and their detection is important to improving commercial aviation safety. The Commercial Aviation Safety Team (CAST) reviewed international airplane accidents between 2001 and 2010 and attributed 18 of them to loss of control inflight (LOC-I). They found that 13 of those 18 accidents, accounting for more than half of the fatalities resulting from LOC-I, involved flight crew loss of airplane state awareness. Further, they found that distraction was involved in all 18 of the LOC-I events (CAST, 2014a). Distraction was divided into two types: channelized and diverted attention. When diverted, crew were distracted from aviating by actions and thoughts associated with decision making (sometimes under high workload). When channelized, crew focused on one instrument or response to the exclusion of other important sources of information. Additionally, confirmation bias was described as making a decision based on faulty information or incorrect reasoning which favors one understanding of an event. As a result, research on AHPLS including channelized attention, diverted attention, and confirmation bias has been recommended in a Safety Enhancement (SE) entitled “Training for Attention Management” (CAST, 2014b). The startle reflex and the surprise response of the sympathetic nervous system

¹ Researcher, Crew Systems & Aviation Operations Branch, Mail Stop 152, AIAA Senior Member

² Researcher, Comprehensive Digital Transformation, Big Data Team, Booz Allen Hamilton, Mail Stop 148

³ Researcher, Crew Systems & Aviation Operations Branch, Mail Stop 152

⁴ Researcher, Crew Systems & Aviation Operations Branch, Mail Stop 152

⁵ Researcher, Crew Systems & Aviation Operations Branch, Mail Stop 152

⁶ Researcher, Crew Systems & Aviation Operations Branch, Mail Stop 152

(Rammirez-Moreno & Sejnowski, 2012) are also included in the SE. To accomplish the detection of such cognitive and physiological states, the Crew State Monitoring team at the NASA Langley Research Center (LaRC) has implemented a broad suite of sensors to simultaneously measure their psychophysiological markers in high fidelity flight simulation studies. The goal is to demonstrate simultaneous real time classification of multiple states using multiple sensing modalities in high-fidelity flight simulators. This detection is intended to support and inform training methods under development to mitigate the loss of ASA and thus reduce accidents and incidents.

The detection of these states will aid understanding of the phenomena behind their occurrence in flight, and enable the development of ground-based training methods to mitigate attentional performance decrement (Hockey, et al., 2003). Attentional effort can be increased to improve attentional performance (Sarter, et al., 2006) if pilots know when to apply these efforts to support situational awareness without causing other detrimental effects. Analogous to training to learn to recognize one's own reaction to and symptoms of hypoxia or fatigue and respond appropriately, training methods will be devised to improve self-monitoring of and response to one's own attentional performance.

Channelized attention may be detected using the combination of neurological indicators of attention and differences between actual versus expected overt behavior during specific phases of flight. The combination of information from such separate measures or converging indicators (Cowings et al. 2007) is intended to improve the discrimination of optimal states of attention from non-optimal states of attention. Thus, psychophysiological and behavioral data fusion efforts are a key to success. The use of multi-modal psychophysiological measures as classifier input features is predicted to produce classifiers with greater accuracy of state prediction than those produced using uni-modal measures. This paper reports the initial results from our first human subject flight simulation study.

II. Background

The use of multiple simultaneous psychophysiological measures for emotional and cognitive state prediction, as reviewed recently by Novak, et al. (2012), is emerging in the literature but not yet widespread. Multi-modal wireless sensor systems are commercially available, and much work has been done to predict workload from physiological measures, e.g. Wilson et al., 2003. However, prior work has not fully investigated the classification accuracy for the particular states of interest to CAST based on multi-modal sensing in a real-time system using operationally-relevant, realistic flight scenarios. Further, data fusion methods for classifying psychological states from psychophysiological measures have not yet matured to universal acceptance (Novak, et al., 2012). Although researchers applying psychophysiological measures to the study of emotional experience in the real world have specified key considerations to serve as guidelines for studying psychological states in operational contexts (Wilhelm & Grossman, 2010), much work remains to be done regarding the validation of state detection and classifier generalizability from controlled laboratory environments to more realistic situations (Kingstone, et al., 2003). Furthermore, a balance must be struck between implementation in realistic uncontrolled operational environments and the use of laboratory-based psychological tasks so that the contribution of each sensor to the measurement of each state may be understood. This is important to determine the generalizability of a classification system across tasks, to the objective selection of sensors for operational systems designed for use outside the laboratory, and to the understanding of the phenomena behind the induction of AHPLS.

Data for the present report were collected during a human subject study executed in a fixed base flight simulator. Multiple simultaneous psychophysiological measures were integrated as a means of detecting AHPLS during benchmark tasks for classifier model training and during experimental flight scenarios for prediction testing. The experimental flight scenarios will be revised for future presentation in a motion base flight simulator. Analysis efforts begin with an initial assessment of these methods in the greater context of plans to further explore classification and analysis methods toward testing this technology in higher-fidelity flight simulators. Data fusion and state classification will be implemented in multiple ways for comparison purposes and for informing future operational system development. A state prediction will be made from among the states of interest using (1) each sensing modality (uni-modal classifier models), (2) all sensing modalities (a single multi-modal classifier model), and (3) the use of multiple uni-modal and multimodal classifier outputs in higher-level state classification. Prior work classifying state based on EEG signals (Li, 2015) may be leveraged to address the problems of individual variation among participants and to address situations where truth labels are unavailable. The current paper initially reports the ability of classifier models to discriminate between cognitive states as induced by the benchmark tasks. Initial uni-modal and multi-modal results were compared, both using a high-level combination of the outputs from three different classifiers.

If successful, a prototype system will be developed for integration and hardening toward further validation and potential use in relevant operational environments. Also, knowledge of which classifier input feature carried the most weight in creating each classifier model can inform sensor down-selection. For example, such weighting can be used to determine the location of a single-site electroencephalography (EEG) electrode instead of a multi-site EEG array. Similarly, accuracy can be determined with the input from each sensing modality left out in turn from higher-level classification. In this way, decisions may be made for each sensing modality and feature regarding further investment in sensor obtrusiveness, crew acceptance efforts and training simulator or eventually cockpit integration based on the value it brings to overall state prediction accuracy. Reducing sensor obtrusiveness and computational overhead is desirable for implementation in operational contexts.

III. Human Subject Study Methods

Twelve regional air transport pilots (one female) were asked to perform tasks in a fixed base flight simulator. All participants consented to take part in the study as approved by the Institutional Review Board of the NASA Langley Research Center. The tasks included: resting tasks, benchmark tasks designed to induce AHPLS and low/high workload conditions, and experimental flight scenarios. The experimental flight scenarios were designed with variations in task load to induce variations in situation awareness, workload, and stress that may be associated with AHPLS.

A. Data Collection

To probe both the brain and the sympathetic nervous system, psychophysiological sensors were applied to measure electroencephalography (EEG) signals using an Advanced Brain Monitoring X24 EEG System (Berka, et al., 2007), and to measure electrocardiogram (EKG), respiration (Resp.), and galvanic skin response (GSR) via a Mind Media, BV. Nexus-10 Mark II system passively throughout all task and simulated flight performance. Pilot behavior was recorded in the flight simulator via a Smart Eye, AB. eye tracking system and simulator flight control inputs. Two self-report questionnaires were administered after each task: the subjective NASA Task Load Index was used to evaluate each participant’s workload during each task (Hart & Staveland, 1988), and the qualitative NASA Situation Awareness Rating Technique (SART) (Taylor, 1990) was used to assess situation awareness.

B. Signal Pre-Processing

All measured time series were recorded using MAPPS (EyesDx, Inc., Coralville, IA), a software suite designed to collect aircraft/simulator state, event marker, and pilot psychophysiological and behavioral data. The software time synchronizes all data channels for real-time review of data, as well as post-hoc analysis. Eight mono-polar EEG signals were reduced to examine frequency domain components via spectral analysis, heart rate variability analysis was derived from the EKG (Hamilton, 1987), and skin conductance responses were taken from GSR signal to generate normalized classifier input features from the time series measurements..

C. Benchmark Tasks

Use of the benchmark tasks was modeled after the methods of Hirshfield, et al. (2009). The selected tasks, listed in Table 1, were used to induce AHPLS under controlled conditions, and were chosen for high likelihood to induce these experiences in isolation and with the full knowledge of the participant (except for the startle / surprise task and the high versus low workload condition). The benchmark tasks lasted for 6 minutes each. For baseline rest, the participants were asked to sit quietly without daydreaming. For Channelized Attention, a personal-computer-based version of the game Tetris (Fairclough & Gilleade, 2012) was employed and participants were instructed to remain completely focused on playing the game. To induce startle and surprise,

<i>AHPLS</i>	<i>Task</i>
Baseline rest	Rest, Eyes Open looking at crosshair
Channelized Attention	Tetris
Startle / Surprise	unexpected primary flight display inversion
Diverted Attention	Vigilance Task with Math
High Workload, flight-like complex multi-task	MATB High Workload
Low Workload, flight-like complex multi-task	MATB Low Workload

Table 1. Benchmark Tasks.

the primary flight display in the simulator was inverted unexpectedly during a flight navigation task. For Diverted attention, the participants were asked to respond with a button press when bars of different heights were presented on the left-hand side of the screen (Parasuraman & Davies, 1977). On the right-hand side, periodically (every 20 to

30 seconds) a math task appeared which required their immediate attention prior to their return to monitoring for bar-height targets. The bar height patterns to be compared were presented simultaneously, or successively with a short pause between them necessitating the activation of working memory. Finally, high and low workload was induced with a flight-like multi-task (the MATB^{*}) used to simulate nominal flight in an optimal attentional state (Santiago-Espada, Myer, Latorella, & Comstock, 2011).

D. Classification Methods

Measurements made during the benchmark tasks provided required ground truth for training state classifiers using supervised machine learning techniques. Classifier models will be trained to recognize pilot state during the experimental flight scenario trials based on patterns of the physiological signals measured during the benchmark tasks. In this way, the benchmark data will be used without a preconceived model of expected physiological signal change. Classifier model training data will be that which was collected during both baseline resting and benchmark-task-induced states, enabling binary (two-class: resting vs. channelized) classification at each time point, or multi-class distinction of the AHPLS (channelized or diverted) from the low-workload state, as reported here. The state classifiers then will be used to predict operator state during each experimental flight scenario or segment. The true or known state will be used to determine classifier prediction accuracy. This ground truth for cognitive state during the experimental scenarios will be indicated by the convergence of physical behavior, missed manipulation checks, subjective self-report, eye tracking results, and flight technical performance (taking into account what is expected for the specific phase of flight based on subject matter expertise). In this way, the effectiveness of the flight scenario to induce the intended state can be verified.

Three classifiers were trained with the first 50 % of the signals. The next 25 % was used for validation, and the final 25 % was used for testing. The three classifiers were: gradient boosting (XGB[†]), a deep neural network (DNN) (Keras[‡]), and a Random Forest classifier (Scikit-Learn[§]). Gradient boosting is an ensemble machine learning technique which fits many classification and regression trees (CART) to the input data (Friedman, 2001). CARTs are trained via the boosting method where multiple weak models are eventually combined into a single, effective model. DNNs are supervised machine learning models made up of many layers of artificial neural network nodes (Srivastava et al., 2014). These deep layers are capable of learning features from training data which improve model classification performance. The DNNs were trained with a technique known as dropout to prevent overfitting on training data. Random forest is another ensemble model which trains multiple weak decision tree classifiers that are combined into a single, robust model (Breiman, 2001). Random forest decision trees are trained using feature bootstrap aggregation, or bagging. The overall prediction was based on a simple average of the three results.

IV. Results and Discussion

Using multi-modal psychophysiological measures as classifier input features was predicted to produce classifiers with greater accuracy of cognitive state prediction than those produced using uni-modal measures. To determine if the results are in agreement with the prediction, the accuracy of classification for each of three states (channelized attention, diverted attention and low workload flight) for the uni-modal physiological measurement modality of EEG was reported and compared to the accuracy obtained using multi-modal classification for EEG and GSR. Confusion matrices were generated to assess the false positive rates and predictive power of the classifiers. These are presented in Tables 2 and 3. Plotted in Fig. 1 for one participant are the true positive and false negative rates with and without multi-modal features. The addition of the GSR signal as a feature to the EEG features resulted in a 7% improvement in correct classifications for the deep neural network model. The EEG-only model was 82 % accurate, while the addition of the GSR feature improved the accuracy to 89%. As a result, overall accuracy rose from 86% to 90%. The addition of GSR to the other two classifiers did not improve accuracy. In the GSR case, only the diverted attention successive case was used.

Both models are reliable for distinguishing the low workload and channelized attention benchmark states. The classifier models are generally less reliable for diverted attention states. Diverted attention may be more difficult to predict because of the periodic nature of the diversion, and assignment of the truth labels to account for time spent diverted to performing the math task should be refined.

* Multi-Attribute Task Battery available at: <http://matb.larc.nasa.gov/>

† eXtreme Gradient Boosting (XGB) available at: <https://xgboost.readthedocs.org/en/latest/>

‡ Keras available at: <http://keras.io/#keras-theano-based-deep-learning-library>

§ SKLearn available at: <http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

V. Limitations and Future Work

Regarding AHPLS investigated, the analysis of startle / surprise and resting state prediction are left to future work. Also, an appropriate benchmark task to induce confirmation bias has yet to be designed. Regarding features, the inclusion of heart rate variability (HRV) and eye tracking metrics as classifier inputs are yet to be investigated.

HRV may be important to cognitive state determination, as it generally indicates arousal of the nervous system, and can contribute information to aid the detection of states of temporal stress (Nickel and Nachreiner, 2003). Eye tracking may be useful to objectively assess any state-specific nature of crew instrument scanning. Finally, correlations between the physiological measures and the flight control inputs or subject self-report responses are yet to be assessed. Also, true real time processing is yet to be implemented, although software is in place to aid time series recording, synchronization and visualization, and to interface with custom classification algorithm scripts written in-house.

There are many opportunities for future work in classification analysis. First, a weighted mean with weights based on each classifier's test performance should be explored to replace the simple averaging used to determine the overall state prediction. Second, the importance of each feature to classifier model performance may be used to inform the down-selection of sensor features employed in future tests or system prototypes (such as channels from multisite modalities such as EEG, or derived variables such as HRV). Third, individualized, participant-dependent models may offer greater accuracy while participant-independent models may obviate user-specific classifier training time.

Classification accuracy should be reported for participant-independent classifier models, allowing accuracy to be tested by predicting the state of a participant whose data were not used to generate the model (whose data were left out of the training data set). This will begin to answer the question of how much user-specific training time (if any) is required prior to use while still achieving acceptable results. However, additional participants may be needed to support participant-independent analyses. Employing adaptive on-line machine learning techniques is of interest to reduce classifier training time while maintaining useful positive predictive power. Regardless, participant-dependent classification also should be investigated using model

	<i>Low Wrkld Predicted</i>	<i>Chan Attn Predicted</i>	<i>Div Att Succ Predicted</i>	<i>Div Att Sim Predicted</i>	<i>Recall</i>
<i>Low Wrkld True</i>	22562	1141	2	152	0.946
<i>Chan Attn True</i>	1849	22026	195	1	0.915
<i>Div Att Succ True</i>	473	374	482	1485	0.171
<i>Div Att Sim True</i>	23	1192	430	671	0.290
<i>Precision</i>	0.906	0.891	0.435	0.291	<i>Accuracy: 0.862</i>

Table 2. Prediction with EEG alone, for one participant.

	<i>Low Wrkld Predicted</i>	<i>Chan Attn Predicted</i>	<i>Div Att Succ Predicted</i>	<i>Div Att Sim Predicted</i>	<i>Recall</i>
<i>Low Wrkld True</i>	21617	1372	867	-	0.906
<i>Chan Attn True</i>	985	22965	121	-	0.954
<i>Div Att Succ True</i>	433	1219	1170	-	0.414
<i>Div Att Sim True</i>	-	-	-	-	-
<i>Precision</i>	0.938	0.899	0.542	-	<i>Accuracy: 0.901</i>

Table 3. Prediction with EEG and GSR, for the same participant.

individualization methods to determine whether any improvement in accuracy warrants user-specific training time for field use.

Some additional limitations to this initial study stem from the need for additional experimental flight scenarios capable of inducing AHPLS. Ground truth for cognitive state during the experimental scenarios may be enhanced by a systematic ground truth finding procedure based on: the subjective evaluations of flight instructors, flight performance measures, and strong physiological indicators (Zhang, et al., 2010). Finally, collected data also may be subjected to analysis employing adaptive neuro-fuzzy inference system methods (MATLAB ANFIS editor) to determine whether continuous state labels can be assigned rather than categorical classification labels (Novak, et al., 2012; Malkawi and Murad, 2013; Lin, et al., 2006; Moon, et al., 2002). These concerns will be addressed and opportunities for investigation will be undertaken as resources permit in future years, starting with higher fidelity motion-based flight simulation trials already planned.

VI. Conclusion

In conclusion, multi-modal sensing has improved the accuracy of AHPLS detection. The results of this and future work will be useful for determining the value of simultaneous multi-modal psychophysiological measures, and the value each sensor brings to classifier accuracy. Sensors may then be chosen by weighing their value against the cost of using them in operational training contexts. Costs include pilot acceptance, obtrusiveness, comfort and privacy considerations, time spent training the classifier or applying sensors, and potential distraction from primary tasking – that of safe flight, real or simulated. However, such costs may not be appropriately weighed against the value of psychophysiological sensing until that value is adequately assessed and understood. The resulting state information can be fed to automated intelligence in the cockpit that can aid or alert the pilot to improve their performance or assist in the avoidance of errors. This work begins to determine the projected efficacy of such a crew state monitoring system and its potential future impact on the avoidance, detection, mitigation, and recovery from safety-critical human crew error.

In the broader context of improved monitoring of an increasingly complex airspace, an extension of these methods can be made to include aviation system-wide events and the detection of accident-relevant precursors in ground, air (crewed and uncrewed) and control arenas. In combination with pilot state, this information may be used to detect a developing error chain before it results in an incident. Events patterns, known because they have been previously associated with incidents and accidents, may be monitored in real time to enable timely action for the purposes of aviation safety.

Acknowledgments

This research is supported under NASA's Airspace Operations and Safety Program's Technologies for Aircraft State Awareness Sub-Project. The authors thank Manjula Ambur for support and encouragement, Capt. Dan Kiggins, Capt. James Wilkerson and Randy Bailey for consultation and contributions to flight scenario development, Kiel Gilleade for the Tetris game, Katrina Colucci-Chang and Matthew Grey for signal pre-processing, Trey Arthur for benchmark task software development, and Steve Williams and Vince Houston for visual simulation support.

References

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., Craven, P. L. (2007) EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. *Aviation, Space and Environmental Medicine*, Vol. 78, No. 5, Section II, May 2007.

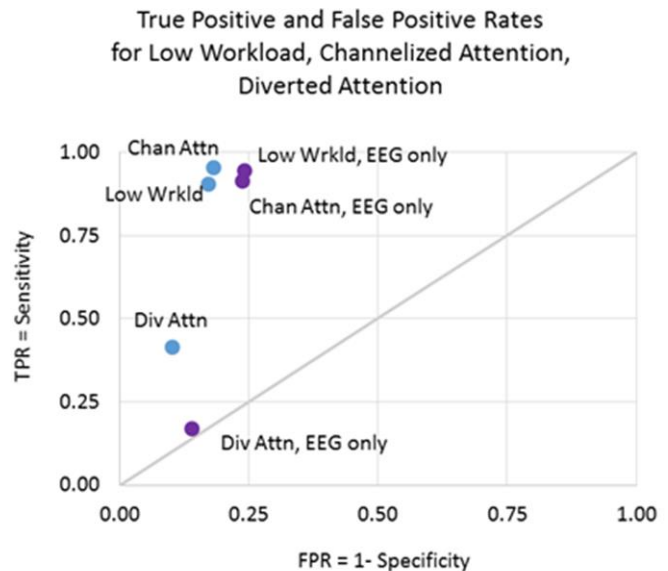


Figure 1. True and False Positive Rates with uni-modal (EEG only) vs multi-modal (EEG and GSR) classification.

Breiman, L. (2001). "Random Forests." Retrieved on October 23, 2015 from: <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>

Commercial Aviation Safety Team. (2014a) Airplane State Awareness Joint Safety Analysis Team Interim Report. Retrieved on March 3, 2015 from: http://www.skybrary.aero/index.php/Commercial_Aviation_Safety_Team_%28CAST%29_Reports

Commercial Aviation Safety Team. (2014b). SE211: Airplane State Awareness - Training for Attention Management (R-D). Available at [http://www.skybrary.aero/index.php/SE211:_Airplane_State_Awareness_-_Training_for_Attention_Management_\(R-D\)](http://www.skybrary.aero/index.php/SE211:_Airplane_State_Awareness_-_Training_for_Attention_Management_(R-D))

Cowings P, Toscano W, DeRoshia C, Taylor B, Hines AL, Bright A, Dodds A (2007) Converging indicators for assessing individual differences in adaptation to extreme environments. *Aviation, Space, and Environmental Medicine* 78 (5):B195-B215.

Fairclough, S., and Gilleade, K. (2013) Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation. *International Journal of Autonomous and Adaptive Communications Systems*. 6:1, 63–79.

Fairclough, S. H., & Gilleade, K. (2012). Construction of the Biocybernetic Loop: A Case Study. ICMI 2012 Grand Challenge - Brain-Computer Interfaces. Santa Monica, CA: ACM.

Friedman, J. H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine." Retrieved on October 23, 2015 from: <https://statweb.stanford.edu/~jhf/ftp/trebst.pdf>

Hamilton, Tompkins, W. J., "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database", *IEEE Trans. Biomed. Eng.*, BME-33, pp. 1158-1165, 1987.

Harrivel A. R., Weissman, D. H., Noll, D. C. and Peltier, S. J. (2013) Monitoring attentional state with fNIRS. *Front. Hum. Neurosci.* 7:861. doi: 10.3389/fnhum.2013.00861

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. NASA, Ames Research Center. Mountain View, CA: NASA.

Hirshfield, L. M., Solovey, E. T., Girouard, A., Kebinger, J., Jacob, R. J., Sassaroli, A., et al. (2009). Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload with Functional Near Infrared Spectroscopy. *Computer-Human Interaction*. Boston, MA: CHI.

Hockey, G. R. J. (2003) "Operator Functional State as a Framework for the Assessment of Performance Degradation" in Hockey, G. R. J., Gaillard, A. W. K, and Burov, O., *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks*. NATO Science Series, IOS Press, p. 9-11.

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., and Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science*. 12:5, 176-180.

Li, F. (2015) Improving Engagement Assessment by Model Individualization and Deep Learning. Dissertation, Old Dominion University.

Lin, C-T., Ko, L-W., Chung, I-F, Huang, T-Y, Chen, Y-C, Jung,T-P, and Liang, S-F. Adaptive EEG-Based Alertness Estimation System by Using ICA-Based Fuzzy Neural Networks. *IEEE Transactions on Circuits and Systems—I: Regular Papers*, VOL. 53, NO. 11, November 2006, 2469-2476.

Malkawi, M. and Murad, O. (2013). Artificial neuro fuzzy logic system for detecting human emotions. *Human-centric Computing and Information Sciences*, 3:3.

MATLAB, Statistics and Signal Processing Toolboxes, R2012b, The MathWorks, Inc., Natick, MA, United States.

Moon, B.S., Lee, H.C., Lee, Y.H., Park, J.C., Oh, I.S., and Lee, J.W. (2002). Fuzzy systems to process ECG and EEG signals for quantification of the mental workload. *Information Sciences* 142, 23–35.

Nickel, P., Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors* 45(4):575-590.

Novak, D., Mihelj, M., Munih, M., (2012). A survey of methods for data fusion and system adaption using autonomic nervous system responses in physiological computing. *Interacting with Computers* 24:154-172.

Parasuraman, R., & Davies, D. R. (1977). A Taxonomic Analysis of Vigilance. In R. R. Mackie, *Vigilance: Theory, Operational Performance, and Physiological Correlates*, 559-574. New York, NY: Plenum.

Pope, A.T., Bogart, E.H., Bartolome, E.S. (1995) Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*. 40, 187-195.

Ramirez-Moreno, D. & Sejnowski T. (2012) A computational model for the modulation of the prepulse inhibition of the acoustic startle reflex. *Biological Cybernetics*. 106(3), 169-176.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide. NASA, Langley Research Center. Hampton: NASA/TM-2011-217164, L-20031, NF1676L-12800.

Sarter, M., Gehring, W.J. and Kozak R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews*. 51:2, 145-160.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." Retrieved on October 23, 2015 from: <http://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>

Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew system design. AGAR, Situational Awareness in Aerospace Operations. AGAR.

Wilhelm, F., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, 84, 552–569.

Wilson GF & Russell CA (2003) "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," Hum. Factors, vol. 45, no. 4, pp. 635-643, 2003.

Zhang, G., Wang, W., Pepe, A., and Xu, R. A Systematic Approach for Real-Time Operator Functional State Assessment, Proceedings of the MODSIM World conference, Oct.13-15, 2010, Hampton, VA.