Defence Research and Development Canada
DRDC-RDDC-2014-P147
December 2014

Untangling Operator Monitoring Approaches When Designing Intelligent Adaptive Systems for Operational Environments

Ming Hou[1] and Cali M. Fidopiastis[2]

[1] Defence Research & Development Canada-Toronto, Toronto, Canada
Ming.Hou@drdc-rddc.gc.ca
[2] University of Alabama at Birmingham, Birmingham, AL, USA
cfidopia@uab.edu

Abstract. An Intelligent Adaptive System (IAS) is a synergy between an intelligent interface and adaptive automation technologies capable of context sensitive interaction with operators. A well-designed IAS should enable flexible task allocation between the operator and the machine. Research suggests that the integration of real-time operator state assessment (e.g., performance, psychophysiology) can create a true 'human-in-the-loop' system, thereby minimizing deleterious performance effects such as overlooking automation failures and slowly reorienting to tasks. However, these research approaches apply a variety of methodologies to determine sensors, metrics, and overall system design when applied to real world tasks. This paper seeks to untangle these issues such that a more comprehensive framework for systematically evaluating the utility of cognitive state detection methods is attainable.

Keywords: Intelligent tutoring systems, adaptive automation, augmented cognition, psychophysiological measures, cognitive state.

1 Introduction

The goal of a well-developed Intelligent Adaptive System (IAS) is to reduce operator workload and fatigue by automatically assigning tasks to the machine (i.e., adaptive automation) during times of high-stress (e.g., time-pressured or emergency situations)[1]. Alternatively, during times of under load, the IAS should reengage the operator to improve situation awareness by either returning the tasks to operator control (i.e., keeping the operator in-the-loop) or by providing stimuli to refocus operator attention (i.e., keeping the operator on-the-loop). An ideal IAS design should allocate tasks collaboratively (i.e., intelligent adaptive automation) while emphasizing the needs of the operator. Research suggests that the integration of real-time operator state assessment (e.g., performance, psychophysiology) can create a true 'human-in-the-loop' system, thereby minimizing deleterious performance issues such as overlooking automation failures and slowly reorienting to task [2]. Technological advances in artificial intelligence and augmented cognition have the potential to realize this type of intelligent support.

Technologies that monitor real-time psychophysiological changes provide an unobtrusive means to assess operator internal status without interrupting the task performance. Operator state monitoring technologies enable an IAS to invoke dynamic task allocation based on operator needs and adapt accordingly. Thus, these technologies have great potential to facilitate intelligent adaptation in simulation based training and real world settings. For example, electroencephalography (EEG) can monitor cognitive workload by measuring electrical activity through scalp-mounted sensors. Electrocardiography (ECG), electrodermal response (EDR), and respiration rate can successfully determine mood [3],[4],[5], while eye-trackers and blink monitors ascertain fatigue levels in automobile drivers [6],[7],[8].

While there is reasonable correlation between operator state constructs (e.g., cognitive workload) and psychophysiological measures, there are challenges to using such approaches. First, psychophysiological relationships are heavily task dependent. For example, tasks relying on physical activity can cause inaccuracy in cardiovascular measures of operator state. Second, there is considerable variation in how individuals react to systems under different conditions. There is even further variation when one considers the differences between individuals. Individual differences in traits such as spatial and attentional abilities can also affect human performance [9]. IASs that are able to adapt to these differences have the potential to reduce operator error and make the overall system easier to learn and operate. Thus, IAS designers must consider the merits of operator monitoring technologies that facilitate intelligent adaptation. An IAS can intelligently adapt to unique situations and individual operators by obtaining behavioral, psychophysiological, contextual, and subjective data of the operator in real-time. Clearly, the result of employing operator monitoring techniques is more robust operator-machine interaction and increased efficiency.

While these capabilities are encouraging, creating a standardized framework from which to design, assess, and validate the sensor and metric system integration is elusive. This paper provides an overview of frameworks used to assess operator state monitoring methods and sensor technologies currently used in IAS. Because these frameworks approach the issue of biosensor selection, evaluation, and validity from different, yet specific underlying assumptions, they may provide divergent or conflicting results when generally applied to operational tasks. Thus, the aim of this paper is to bring these issues to light and provide examples and recommendations to improve IAS design and implementation in operational environments.

2 Augmented Cognition Framework and Sensor Selection

Defining the human state of the operator through psychophysiological sensors and associated metrics has a long history. From the eye tracking work conducted by [10] to the multi-modal computerized airborne research platform at the University of Iowa [3], biosensor integration within these simulated real world settings provide objective measures of task relevant human state detection. The task context is only one important factor in determining the relevance of including psychophysiological metrics within an IAS [11]. There are other critical issues that preclude the successful straightforward application of psychophysiological measures to intelligent and adaptive systems for field use.

Cummings [12] provided a review of the Augmented Cognition (AugCog) research domain and concluded that psychophysiological based adaptive triggers suffered from the lack of construct and internal validity (highly correlated independent measures), statistical validity (low sample sizes), missing data, and inherently noisy sensor data. St. John et al. [13] do cite similar issues in their discussion of results from the AugCog integration feasibility study, a proof of concept of demonstration of using this framework in field based experiments. Cummings' suggested remedy for the AugCog "dilemma" is the development of better sensors and signal processing tools. This suggestion is naïve to the fact that biosensor choice and their respective metrics derive from unstandardized operational definitions of cognitive constructs, which in turn suffer from the biases inherent in the brain theory chosen to support the operational definition [14], [15], [16].

Left open to the IAS designer are the specifics on how to operationally define the human state, calculate the sensor metric, apply data fusion techniques, and define appropriate methods for state classification. Because there are no standards, basic assumptions on how to proceed are anyone's best guess. In an effort to apply guidelines to sensor and metric selection, IAS researchers utilize Technology Readiness Levels (TRL) to evaluate sensors and metrics for use in military applications [17], [18]. The US Department of Defense (DOD) determined a method to categorize the maturity of evolving technologies for field [19]. In general, TRLs range from level one, where the basic utility of a technology is documented, to mid-levels four and five, where assessment of the technology in the laboratory and relevant field experiments shows promise for field application, to a high level of nine, where a full-scale system is demonstrated as successful in an operational setting. Stanney and Hale [18] further evaluate the TRL level based on the sensor technologies ability to measure, diagnose, and/or augment the IAS. Table 1 presents a TRL readiness level for sensor metrics and their associated needs when applying such measures using an AugCog framework (adapted from [18]).

Artifact detection and reduction in a mobile, hostile environment also introduce unique challenges to the use of psychophysiological measures [20]. Dorneich et al. note that successful cognitive state classifiers in mobile operational environments are ones that can discriminate the cognitive construct through the noise and improve their accuracy as new data accrues throughout the operation. This necessitates a classifier that determines cognitive state moment-to-moment, and not by averaging responses over time. Thus, sensor selection further depends on the nature of the training task, the operational definition of the cognitive state defined by the task parameters, and the context of the operational environment.

Thus, the effective and the valid design of a general IAS proceeds in two required stages: 1) simulator based experiments using a multi-modal approach to narrow sensors and metric selection and 2) field based experiments that test and validate the IAS approach for field use. In the simulator stage high resolution sensing systems such as EEG and fNIR are appropriate for use in determining cognitive states important to the training task, as well as the robust measurability of the state.

Table 1. TRL Readiness Levels for AugCog Relevant Technologies

	TRL 3-5	TRL 5-6	TRL 6-8	Design Needs
Measure	Functional near infrared imaging (fNIR), posture, pupilometry	EDR, EEG, Respiration	ECG, Eye/Gaze Tracking	Data fusion & classifier techniques
Diagnose	Independent Components Analysis using psychometrics			Define machine learning techniques, define expert models that predict performance failures
Augment	Presentation, Schedule, System autonomy			Define individualized schemes for task change and task interruption

3 Understanding Research Assumptions for IAS Design

Table 2 describes the differences between the research fields of Neuroergonomics and AugCog when implementing psychophysiological measures. The importance of distinguishing between the two research fields is rarely regarded as important; however, each field has different end goals for the use of such metrics. More importantly, these implementation end goals are potentially divergent. Thus, the results based upon a Neuroergonomic approach may not translate into a viable AugCog approach and vice versa. As [21] suggested Neuroergonomics is predominantly a strategy for adaptive automation (AA), while AugCog is for intelligent and adaptive systems for training. Further, the underlying biocybernetic closed-loop of AA [22] is also different from the information processing closed-loop of AugCog (see [16]). While the cognitive state constructs are the same in name, why and how they are measured and utilized are different based upon these closed-loop architectures. Most literature reviews regarding cognitive state constructs, types of sensors, and associated metrics do not distinguish or cite the underlying assumptions or the assessment framework. This oversight does not allow for proper evaluation of the sensors, metrics, and analyses as they apply to each respective field of use.

One major difference between Neuroergonomics and AugCog is the data fusion needs across multimodal sensor arrays within augmented cognition based systems. The multimodal aspect allows prediction of more than one cognitive state. In addition, the automated mitigations within the AugCog framework are also more in number and variety. These mitigations not only pertain to the on and off state of the system, but also the system content, interface, and tasking. Augmented cognition based systems are, therefore, more complex and may take longer to design and validate.

Psychophysiological measures can also assist in determining whether static versus adaptive automation is more conducive for optimal task performance [23]. In essence, the task and the task environment make a difference as to whether psychophysiological measures are even necessary or appropriate. Fairclough [22] suggests that an adaptive response system follows three broad categories: 1) offering assistance when the

Table 2. Difference between Neuroergonomics and Augmented Cognition

	Neuroergonomics	**Augmented Cognition**
Sensor number	Single Sensor	Multiple sensors
Field of use	Genral adaptive automation	Mainly adaptive automation for training; Shifting toward adaptive automation for unmanned systems
Mitigation triggers	On/Off	Triggers rescheduling of tasks, changes in display characteristics, and switching information format (e.g., sound, sight, haptics)
Psychophysiological constructs or cognitive state	Workload and/or engagement	Effort, arousal, engagement, workload

user is frustrated, overloaded, or stuck, 2) adapting the level of the user's interaction to increase engagement or decrease boredom, and 3) reinforcing positive emotion. The AA is typically optimized for one of the categories, not all as in AugCog based IAS. As Fairclough points out, psychophysiological measures are a way to operationalize the operator's state and are variable. As such, there can be a many-to-one mapping of psychophysiological measures to a particular state (e.g., pupil dilation and EEG for cognitive workload). However, a single sensor may be sensitive to more than one psychological state. For example, EEG can measure both engagement and workload [24]. There are also overlapping metrics with several psychophysiological states, such that there is a many-to-many mapping. Thus, the ideal one-to-one mapping of sensor metric to operator state *does not exist*.

In the adaptive automation literature, the use of many-to-one (multiple sensors mapped to a single operator state, workload or engagement) is now typical [25]. This convergent methodology can improve the predictive value of the cognitive state metric. More importantly, the testing environment plays a critical role in accurately determining the validity, reliability, sensitivity, and diagnosticity of the metric. For example, most studies on psychophysiological measures for adaptive systems are laboratory based [26]. How these measures transfer to real world contexts is unknown. Boucsein et al [26] argue that at the very least a simulator that matches the controls, context, and tasking of the real world should provide the testing environment for the metrics. Schaefer, Haarramann, Boucscein [27] further contend that while important operator states such as a decrement in vigilance are measurable using EEG, ECG, and EDR, only ECG and EDR would be acceptable to use on long duration tasks (e.g., flight). Thus, the choice of sensor must match the task process and context.

For example, using a flight simulator, [27] showed the nonspecific SCR and interbeat heart rate interval (corrected with respiration) successfully predicted vigilance decrements and consequently applied automation that assisted the pilot during periods of high turbulence. The researchers compared these results to those pilots operating using yoked controls and found that these measures taken on 1-minute intervals were able to provide the support needed to improve task performance during events experienced in the field. The importance of this study is that the metrics did not need special processing or data fusion approaches. The metrics integrated with the simulator easily and mapped to system sensors (e.g., turbulence) without the overhead of building a software interface.

4 Efficacy versus Effectiveness

Efficacy within AA refers to how well psychophysiological measures perform in laboratory based experiments [28]. IAS imposes a different problem, namely that of effectiveness or the ability of the measure to generalize to the real world. Fuchs et al. [29] suggest that laboratory-based testbeds may not provide effective field deployable IAS solutions. However, efficacy of sensor selection and algorithm implementation is best determined within a simulation based testbed as part of the initial IAS development cycle. Linking the simulation based experiments to the actual task requires sensors and associated algorithms integrated within the real world context. An example of the potential connection between efficacy and effectiveness is the Cognitive Avionics Tool Set (CATS) is a multi-sensor integration, analysis, assessment, and visualization tool that can integrate with flight simulators, as well as aircrafts [30]. The researchers took an iterative stepwise approach to the development of CATS such that the measures for evaluating the simulator-based training are equally relevant for airborne training.

The problem space of pilot training, regardless of vehicle (e.g., car, plane, etc.), necessitates a cognitive state assessment battery along with a methodology to collect high quality data and a technique to evaluate the data meaningfully in the operational environment. CATS implement an augmented cognition adaptive closed-loop framework that provides a means to conduct an iterative design cycle with the Soldier-in-the loop. The researchers [17] also chose a battery of sensors with a technical readiness level of at least six (sensors show relevance in a simulator) and refined any measure below 6 with a comprehensive experimental plan. These sensors included eye tracking (estimated TRL 7), ECG (estimated TRL 7), respiration (estimated TRL 6), EDR (estimated TRL 6), and EEG (estimated TRL 5). The interface software and architecture was further developed within a simulator environment, and then tested in field flight training. Of importance is that refinement of the sensors, algorithms, and integration techniques took several years to evaluate pilot readiness, as well as adapting the training system. A generalized IAS framework for a class of related training tasks provided a foundation for appropriate application of the sensor implementation to other related operational fields.

5 Summary and Future Directions

Both Neuroergonomics and AugCog research rely on psychophysiological measures and suffer from the same issues when applying these measures within the operational environment. For example, both research domains necessitate a cognitive map of neural processes (e.g., attention and memory) reliably represented by patterns of neuronal firing across brain regions that are correlated with task performance [31]. Further, the fields utilize medical technologies that monitor central and peripheral nervous system responses, (i.e., EEG and ECG, respectfully). Tools for validating these translated algorithms for use within real-time dynamical human-machine systems are still under development.

This development process necessitates a more diligent approach to the choice of operational definitions for cognitive constructs, better understanding of underlying framework assumptions, and a comprehensive research agenda that includes efficacy and effectiveness studies. While CATS provides a successful example of translating the AugCog approach across different navigation activities, this method may not be appropriate for all IAS implementations.

Some general recommendations for IAS design:

1. Define clearly and accurately the task context and the task requirements such that they transfer into an effective operational protocol.
2. Perform a cognitive task analysis thereby matching tasking with motor, cognitive, perceptual, and multi-modal aspects of the domain;
3. Determine from the task specification and cognitive task matrix if the system needs to be adaptive and/or intelligent;
4. Define operationally the important cognitive state constructs;
5. Perform simulator-based experiments with a multi-modal sensor approach to narrow sensors and metric selection. The main goal of this step is to separate task-independent cognitive states and functions from those that are task relevant. This stage may require high-resolution sensors such as EEG and fNIR to provide verification that the operator state is measurable and has the resolution necessary for accuracy;
6. Validate the integrated closed-loop system with field operators within the simulator. Note that this step requires an assessment of the efficacy of the IAS.
7. Perform effectiveness studies in the field to verify generalizability of the IAS approach.

As future system designs transition from passive machines to more proactive and perhaps even predictive machines able to adapt to new operators, environments, and scenarios, the use of operator feedback data to facilitate intelligent adaptation will provide a higher degree of system reliability and protection for the operator.

References

1. Hou, M., Kobierski, R., Brown, M.: Intelligent adaptive interfaces for the control of multiple UAVs. J. Cogn. Eng. Decis. Making 1(3), 327–362 (2007)
2. Kaber, D.B., Endsley, M.R.: The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task. Theor. Iss. Ergon. 5(2), 113–153 (2004)

3. McCraty, R., Atkinson, M., Tiller, W.A., Rein, G., Watkins, A.D.: The Effects of Emotions on Short-term Power Spectrum Analysis of Heart Rate Variability. Am. J. Cardiol. 76(14), 1089–1093 (1995)
4. Schnell, T., Keller, M., Macuda, T.: Pilot State Classification and Mitigation in a Fixed and Rotary Wing Platform. Aviat. Space Environ. Med. 78(3), 377 (2007)
5. Figner, B., Murphy, R.O.: Using Skin Conductance in Judgment and Decision Making Research. In: Schulte-Mecklenbeck, M., Kuehberger, A., Ranyard, R. (eds.) A Handbook of Process Tracing Methods for Decision Research, pp. 163–184. Psychology Press, New York (2011)
6. Barr, L., Howrach, H., Popkin, S., Carroll, R.J.: A Review and Evaluation of Emerging Driver Fatigue Detection, Measures and Technologies, US department of Transportation, Washington DC, USA (2009)
7. Dinges, D.F., Mallis, M., Maislin, G., Powell, J.W.: Final Report: Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and as the Basis for Alertness Management. Report No. DOT HS 808 762. National Highway Traffic Safety Administration, Washington, D.C. (1998)
8. Fraunhofer-Gesellschaft, `http://www.fraunhofer.de/en/press/research-news/2010/10/eye-tracker-driver-drowsiness.html`
9. Chen, J.Y.C., Terrence, P.I.: Effects of Tactile Cueing on Concurrent Performance of Military and Robotics Tasks in a Simulated Multitasking Environment. Ergon. 51(8), 1137–1152 (2008)
10. Fitts, P.M., Jones, R.E., Milton, J.L.: Eye Movements of Aircraft Pilots During Instrument-landing Approaches. Aero Engin. Rev. 9(2), 24–29 (1950)
11. Mathan, S., Whitlow, S., Dorneich, M., Ververs, P., Davis, G.: Neurophysiological Estimation of Interruptibility: Demonstrating Feasibility in a Field Context. In: Schmorrow, D.D., Nicholson, D.M., Drexler, J.M., Reeves, L.M. (eds.) Foundations of Augmented Cognition, 4th edn., pp. 51–58. Strategic Analysis, Arlington (2007)
12. Cummings, M.L.: Technology Impedances to Augmented Cognition. Ergon. Des., 25–27 (2010)
13. St. John, M., Kobus, D.A., Morrison, J.G., Schmorrow, D.: Overview of the DARPA augmented cognition technical integration experiment. International Journal of Human-Computer Interaction 17(2), 131–149 (2004)
14. Fidopiastis, C.M., Drexler, J., Barber, D., Cosenzo, K., Barnes, M., Chen, J.Y., Nicholson, D.: Impact of Automation and Task Load on Unmanned System Operator's Eye Movement Patterns. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS (LNAI), vol. 5638, pp. 229–238. Springer, Heidelberg (2009)
15. Fidopiastis, C.M., Nicholson, D.M.: Neuroergonomics: From Theory to Practice. In: Marek, T., Karwowski, W., Rice, V. (eds.) Advancing the Understanding of Human Performance: Neuroergonomics, Human Factors Design, and Special Populations, ch. 36, pp. 354–359. CRC Press, Boca Raton (2010)
16. Fidopiastis, C.M.: Theoretical Transpositions in Brain Function and the Underpinnings of Augmented Cognition. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) FAC 2011. LNCS (LNAI), vol. 6780, pp. 153–158. Springer, Heidelberg (2011)
17. Schnell, T., Cornwall, R., Walwanis, M., Grubb, J.: The Quality of Training Effectiveness Assessment (QTEA) Tool Applied to the Naval Aviation Training Context. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS (LNAI), vol. 5638, pp. 640–649. Springer, Heidelberg (2009)

18. Stanney, K.M., Hale, K.S.: Today's Competitive Objective: Augmenting Human Performance. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) FAC 2011. LNCS (LNAI), vol. 6780, pp. 628–635. Springer, Heidelberg (2011)
19. DOD, https://acc.dau.mil/CommunityBrowser.aspx?id=154268
20. Dorneich, M.C., Mathan, S., Ververs, P.M., Whitlow, S.D.: Cognitive State Estimation in Mobile Environments. In: Schmorrow, D.D., Stanney, K.M. (eds.) Augmented Cognition: A Practitioner's Guide, pp. 75–111 (2008)
21. Scerbo, M.W.: Adaptive Automation. In: Parasuraman, R., Rizzo, M. (eds.) Neuroergonomics: The Brain at Work, ch. 26, pp. 239–252. Oxford University Press, New York (2007)
22. Fairclough, S.H.: Fundamentals of Physiological Computing. Interact. Comput. 21(1-2), 133–145 (2009)
23. Sauer, J., Nickel, P., Wastell, D.: Designing Automation for Complex Work Environments Under Different Levels of Stress. App. Ergo. 44(1), 119–127 (2013)
24. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviat. Space Environ. Med. 78(suppl. 1), B231–B244 (2007)
25. Wilson, G.F., Russell, C.A.: Performance Enhancement in an Uninhabited Air Vehicle Task using Psychophysiologically Determined Adaptive Aiding. Hum. Fact. 49(6), 1005–1018 (2007)
26. Boucsein, W., Haarmann, A., Schaefer, F.: Combining Skin Conductance and Heart Rate Variability for Adaptive Automation During Simulated IFR Flight. In: Harris, D. (ed.) HCII 2007 and EPCE 2007. LNCS (LNAI), vol. 4562, pp. 639–647. Springer, Heidelberg (2007)
27. Schaefer, F., Haarmann, A., Boucscein, W.: The Usability of Cardiovascular and Electrodermal Measures for Adaptive Automation. In: Westerink, J., Ouwerkerk, M., Overbeek, T.J.M., Pasveer, W.F. (eds.) Probing Experience: From Assessment of User Emotions and Behaviour to Development of Products, vol. ch. 20, pp. 235–243. Springer, Netherlands (2008)
28. Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Parasuraman, R., Di Nocero, F., Prinzel III, L.J.: The efficacy of psychophysiological measures for implementing adaptive technology. In: National Aeronautics and Space Administration, vol. 211018, Langley Research Center (2001)
29. Fuchs, C., Aschenbruck, N., Martini, P., Wieneke, M.: Indoor Tracking for Mission Critical Scenarios: A Survey. J. of Perv. and Mobi. Comp. 7(1), 1–15 (2011)
30. Schnell, T., Macuda, T., Keller, M.: Sensor Integration to Characterize Operator State. In: Schmorrow, D.D., Stanney, K.M. (eds.) Augmented Cognition: A Practitioner's Guide, pp. 41–74. Human Factors and Ergonomics Society (HFES), Santa Monica (2008)
31. Sarter, N., Sarter, M.: Neuroergonomics: Opportunities and Challenges of Merging Cognitive Neuroscience with Cognitive Ergonomics. Theor. Iss. Ergon. Sci. 4(1-2), 142–150 (2003)