33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Physiological Cognitive State Assessment: Applications for Designing Effective Human-Machine Systems

Justin R. Estepp, *Member, IEEE*, James C. Christensen

***Abstract*—Significant growth in the field of neuroscience has occurred over the last decade such that new application areas for basic research techniques are opening up to practitioners in many other areas. Of particular interest to many is the principle of neuroergonomics, by which the traditional work in neuroscience and its related topics can be applied to non-traditional areas such as human-machine system design. While work in neuroergonomics certainly predates the use of the term in the literature (previously identified by others as applied neuroscience, operational neuroscience, etc.), there is great promise in the larger framework that is represented by the general context of the terminology. Here, we focus on the very specific concept that principles in brain-computer interfaces, neural prosthetics and the larger realm of machine learning using physiological inputs can be applied directly to the design and implementation of augmented human-machine systems. Indeed, work in this area has been ongoing for more than 25 years with very little cross-talk and collaboration between clinical and applied researchers. We propose that, given increased interest in augmented human-machine systems based on cognitive state, further progress will require research in the same vein as that being done in the aforementioned communities, and that all researchers with a vested interest in physiologically-based machine learning techniques can benefit from increased collaboration. We thereby seek to describe the current state of cognitive state assessment in human-machine systems, the problems and challenges faced, and the tightly-coupled relationship with other research areas. This supports the larger work of the Cognitive State Assessment 2011 Competition by setting the stage for the purpose of the session by showing the need to increase research in the machine learning techniques used by practitioners of augmented human-machine system design.**

I. Introduction

METHODS by which physiological data are used in combination with pattern recognition and machine learning algorithms have become prevalent, no doubt in large successes in the field of and brain-computer interfaces (BCI;[1]-[3]). The various types of physiological data, themselves, that have been incorporated in a wide variety of BCI research including, but are certainly not limited to, modalities such as electroencephalography (EEG; [4]) electrocorticography (ECoG; [5]), functional near-infrared spectroscopy (fNIRS; [6]) and functional magnetic resonance imaging (fMRI; [7]). While each of these physiological measurements are superior and inferior to each other in a variety of ways (spatial resolution, temporal resolution, relationship to neuronal activity, associated hardware cost, invasiveness, etc.), EEG has been used by many as a popular compromise for overall practicality. As such, there has also been a large focus in signal processing [8] and machine learning [9] as they relate specifically to EEG.

Given the tremendous successes of BCI, it's highly plausible that the techniques inherent in these domains are also applicable to providing great improvements in other areas. Indeed, this is the case, and is the central focus of the area of physiological cognitive state assessment, whereby an individual's physiology is monitored and assessed, either post-hoc or in real-time, as it relates to facets of cognitive state, with examples such as workload [10]-[14], vigilance [15], fatigue [16], and emotion [17]. In highly-operational environments, this has often been referred to as operator functional state assessment (OFS; [13]-[14]).

There is converging evidence, as should not be surprising, that there is a wide overlap in both methods and challenges between BCI and cognitive state assessment. While each is certainly unique in its intended application, there is also a high degree of commonality; in regard to researchers of cognitive state assessment, there is an overwhelming amount of knowledge that can be leveraged from experts in BCI. For practitioners of augmented systems interface design using physiological cognitive state assessment as the link between human and machine, it is critical that they approach signal processing, feature extraction and machine learning with the same rigor and level of expertise observed in the neuroengineering community at large.

As an effort to bring the two communities closer together, the Cognitive State Assessment Competition (2011) was organized to bring together researchers from the fields of cognitive and computational neuroscience, novel sensor design, BCI and cognitive state assessment in order to explore the commonalities (and differences) between their own domains of expertise. A competitive analysis of a dataset from a cognitive state assessment study (where the manipulation of state was workload) was also organized to demonstrate to the larger BCI community the techniques used in cognitive state assessment, in hopes of spurring creative discussion and collaboration between researchers in both fields.

II. Application Areas for Human-Machine Systems

As noted in a recent literature review [18-19], the application of neuroergonomics, the "the study of brain and

Manuscript received 15 April, 2001. This work was supported in by the 711th Human Performance Wing Chief Scientist's Seedling Grant Program.

J.R. Estepp and J.C. Christensen are with the 711th Human Performance Wing, Air Force Research Laboratory at Wright-Patterson Air Force Base, 45433-7022 USA (email: {justin.estepp, james.christensen}@wpafb.af.mil).

behavior at work" [20], is highly applicable to many areas of United States Air Force (USAF) operations. While [18]-[19] focused their efforts on USAF operations (as this was their directive), there is certainly a wide-ranging domain to which these methods could (and should) be applied, as evidenced in the literature review contained within. Regardless of the domain of application, the major topic areas contained within the review for which cognitive state assessment has potential benefit are reviewed.

As these volumes of literature review are, by themselves, a wealth of knowledge, the seminal works discussed within will not be treated further in this work, but left to the reader to explore at their leisure. It is certainly the case that each area mentioned has the potential to, or has already been demonstrated to, benefit from robust augmented system design using physiological cognitive state assessment.

III. METHODS IN COGNITIVE STATE ASSESSMENT

While somewhat diverse in technique, many methods used in many cognitive state assessment paradigms are nearly identical to those using EEG for BCI applications. As an example in the context of cognitive workload as a state, [14] uses power in traditional clinical frequency bands (delta [1:3 Hz], theta [4:7 Hz], alpha [31:42 Hz], beta [14-30 Hz] and gamma [31-42 Hz]) of the EEG channels as the largest portion of their feature set. In addition to power in the frequency bands, they also used power in the frequency bands of the vertical and horizontal electrooculogram (VEOG and HEOG, respectively), an analog for heart rate (inter-beat interval, or IBI) collected from the electrocardiogram (ECG), an analog of blink rate (inter-blink interval, or IBLI) from the VEOG, and an analog of respiration rate (inter-breath interval, or IBRI) from the respiration signal. The classifier used was a 3-layer artificial neural network (ANN) using backpropagation training (hidden layer contained the same number of nodes as the input layer, output layer consisted of 3 nodes for a 3-class classifier). In this particular study, [14] achieved 84.9%, 82.0% and 86.0% mean classification accuracies of a baseline, low workload and high workload (respectively) cognitive task in real-time. Post-hoc analysis of the dataset that implemented feature selection based on the ranked saliency of the features from the training sets increased accuracy on the three cognitive tasks to 91.0%, 85.2% and 88.7%, respectively.

In general, these methods are a near perfect complement to methods that have been used in research focusing on BCI applications. One distinct difference, however, could be the inclusion of measures of cognitive state that are not directly neurophysiological, such as heart rate, blink rate, and respiration, as used in [14]. These peripheral measures, largely associated with the autonomic nervous system, have been demonstrated as salient for cognitive sates such as workload [12] but may not lend themselves well to the goals of BCI. In addition, other measures that are largely neurobehavioral, such as eye movements as recorded through eye tracking methodologies [21] could be useful for a variety of cognitive states. As such, the range of unique and (somewhat) independent metrics that can be used for cognitive state assessment are not limited to purely neurophysiological means, which has been demonstrated to be of benefit [22] depending on the domain of interest. Again, in some applications of BCI, these methods may not be reasonable, feasible, or even desirable, given the nature of BCI itself to be largely tied to neural processes. This represents one extension of traditional BCI techniques that cognitive state methodologies may utilize, although is certainly not meant to be interpreted as an advantage over BCI methods, but merely one possible distinction between the two. As evidenced by the great progress in BCI research, the lack of measures not directly related to neural processes is certainly not stifling advances in the state-of-the-art!

IV. FUTURE WORK FOR ROBUST COGNITIVE STATE ASSESSMENT

As with BCI, cognitive state assessment techniques suffer from many of the same fundamental flaws. A breakdown in the ability of a pattern classifier (using data from multiple sessions within an hour of each other) to accurately predict state can be observed over the course of days, and even hours, as time elapses between collection of data used in the machine learning training set at the test set of interest [23]. While the direct source of this variation is still unclear, a certain amount of it could a result of classifier overfitting [9] due to the limited time course represented in the training set. Another possibility, and an almost certainty at that, is that nonstationarity in the feature data over the course of time renders the trained weights and biases of the classifier unusable, as the relationship between input and output vectors has fundamentally changed [8], [24]-[25]. There could also be differences related to electrode placement, impedance and quality variation across multiple days, but some in-progress work of our own suggests that this variation is negligible in relation to the potential confounds of overfitting and nonstationarity. Lastly, but not to be discounted, are differences between individuals that can be observed in both their task performance on a given platform [26] and general categorization of the most salient features used in their individualized model of cognitive state [27].

As a point of paramount importance, these same problems have been observed in BCI literature. In [28], reported waveform shape instability in raw voltage recordings using microelectrode arrays in rhesus monkeys manifests over the course of hours and days, although, as the authors note, some of this variability is likely attributable to shifts in positioning of the microelectrode array over time. [8] observes the same phenomenon in distributions of training and test features and also discusses unsupervised methods for reducing the detrimental effects of nonstationarity. To the authors' knowledge, no such attempt to mitigate nonstationarity in feature distributions in the context of cognitive state has been attempted at the time of publication, although these methods warrant a high degree of merit and are, in fact, the largest motivation for the work presented as part of the Cognitive State Assessment Competition 2011 session.

As far as transitioning work largely done in the laboratory environment to users in the applied, operational environment, there is also a great amount of work to be done in regard to sensor technologies. Ideally, sensor technologies that are deployed to end-users should be non-invasive (minimally invasive may be acceptable in some instances, given adequate performance improvement as a trade-off), robust, and minimal in form factor and quantity with respect to maintaining high performance accuracy. With respect to EEG, there has been a large investment, and also progress, in this effort over the last several years. The authors encourage interested readers to see the review of the state of dry-contact and non-contact biopotential electrodes authored by [29].

V. COGNITIVE STATE ASSESSMENT COMPETITION 2011

As the main focus of this session, a competitive analysis of a common dataset was proposed to outline and demonstrate current machine learning techniques in cognitive state assessment, their relationship to current techniques in BCI, and common findings and problems between the two areas. This competition, called the Cognitive State Assessment Competition, was facilitated by distributing a common dataset to contributing session authors.

The dataset provided for the competition was collected while study participants completed the Multi-Attribute Task Battery (MATB; [30]). There were 8 participants in total, and each participant completed 3 trials on 5 separate days. In each trial, segments of task difficulty intended to produce low, medium and high workload were presented in a random order, with 'transition' time between workload segments. This transition time was 60 seconds between low/high and high/low segments and 30 seconds between low/medium, medium/low, medium/high, and high/medium segments. Each segment was 5 minutes in length. Two examples (random orders ‘A’ and ‘C’) is shown in Figures 1 and 2, respectively.

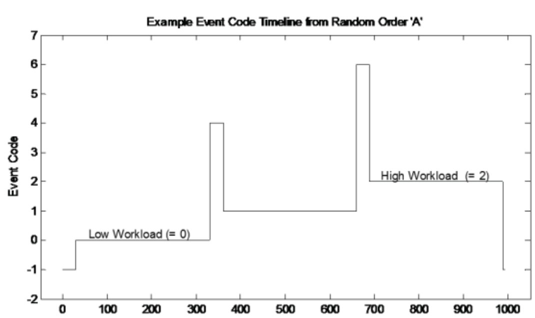

Fig. 1. Example event code timeline from random order ‘A’. In random order ‘A’, order of workload segments was low, medium, and then high. Between each different workload segment, a 30 second transition period was used to gradually increase workload between the starting and ending levels of the transition.

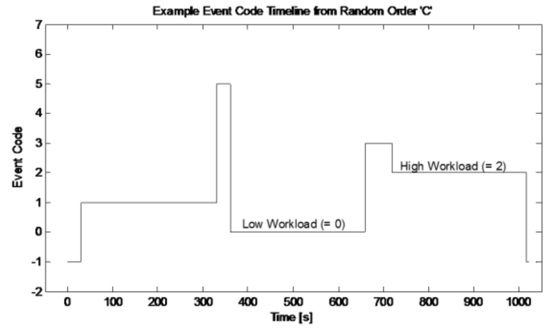

Fig. 2. Example event code timeline from random order ‘C’. In random order ‘C’, order of workload segments was medium, low, and then high. Between medium and low workload, a 30 second transition period was used. However, between low and high workload, a 60 second transition period was used.

The five days of data collection for each participant were not sequential, but spread out over the course of one month. The data collection days were randomly distributed such that each study participant had data collection days that were one day, one week (two instances) and two weeks apart. Deviations from this paradigm were minimized to the extent possible, though some accommodations were made due to participant availability and scheduling conflicts, thus resulting in minor deflections from this ideal schedule for some participants. Table I depicts this data collection schedule for two of the eight participants.

TABLE I
EXAMPLE DATA COLLECTION SCHEDULES

SUN	MON	TUE	WED	THU	FRI	SAT
	Day 1		Day 1	Day 2		
				Day 3		
	Day 2					
	Day 3			Day 4		
	Day 4	Day 5		Day 5		

For each trial 19 channels of EEG (according to the International 10-20 System were collected, as well as peripheral measures such as ECG, VEOG, HEOG and respiration. For the competition dataset, most of the peripheral measures were omitted so that participants could concentrate their efforts on creating features from only the EEG data. Both VEOG and HEOG were included in the dataset in the event any method required them for artifact correction. In total, there were 21 channels of data (19 channels of EEG from the 10-20 System, VEOG, and HEOG) available.

All 21 of these data channels were collected using the MICROAMPS system from SAM Technologies, Inc. (San Francisco, CA, USA). MICROAMPS has default high-pass and low-pass filters at 0.05 [Hz] and 100 Hz, respectively, and a sampling rate of 256 [Hz]. Aside from these filters, no other processing was performed on the dataset. All values are in [μV]. The 19 EEG channels were referenced to a single (left) mastoid. VEOG was a bipolar channel with electrodes placed above and below the left eye. HEOG was also a bipolar channel with electrodes placed outside the outer canthus of each eye. All electrodes were tin cup electrodes (9 [mm]). Impedances for the EEG channels were all below [5 kΩ], and impedances for the VEOG and HEOG channels were all below 15 [kΩ].

Additional information about the competition structure can be referenced in the work by [31], which was a contribution submitted to this session as well. In brief, the session organizers created labeled training and blind test sets on which the competition participants were allowed to perform an initial analysis. This was meant to serve as a baseline (for reasons also discussed in [31]) whereby no additional information about participant or data collection session was available.

At the time of this manuscript submission, all competition participants are active in this preliminary analysis, and many of their initial results should be available in their draft submissions. As a follow-on to this initial analysis, the full, labeled dataset (including information about participants and data collection session) was also distributed in order to permit competition participants to iterate their analysis with knowledge of subject and data collection session (which, the authors feel, is of critical importance to develop robust models of cognitive state from machine learning techniques). In addition, competition participants have been encourage and are pursuing the explore other recommended analyses with the dataset (such as using multiple days in training to improve resiliency to nonstationarity and developing learning techniques that are capable of learning group-level patterns).

REFERENCES

[1] Wolpaw,. J.R., Birbaumer, N., Heetderks, W.J., McFarland, D.J., Peckham, P.K. Schalk. G. et al., “Brain-computer interface technology: a review of the first international meeting,” *IEEE Trans. Rehab. Eng.,*vol 8, pp. 164-173, June 2000.

[2] Birbaumer, N., “Breaking the silence: brain-computer interfaces (BCI) for communication and motor control,” *Psychophysiology*, vol. 43, pp. 517-532, September 2006.

[3] Schwartz, A.B., Cul, X.T., Weber, D.J. and Moran, D.W., “Brain-controlled interfaces: movement restoration with neural prosthetics,” *Neuron*, vol. 52, pp. 205-220, October 2006.

[4] Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F. and Arnaldi, B., “A review of classification algorithms for EEG-based brain-computer interfaces,” *J. Neural Eng.*, vol 4, pp. R1-R13, January 2007.

[5] Wilson, J.A., Felton, E.A., Garell, P.C., Schalk, G., and Williams, C., “ECoG factors underlying multimodal control of a brain-computer interface,” *IEEE Trans. Neural Sys. Rehab. Eng.*, vol. 14 (2), pp. 246-250, June 2006.

[6] Luu, S. and Chau, T., “Decoding subjective preference from single-trial near-infrared spectroscopy signals,” *J. Neural Eng.*, vol. 6, pp. 1-8, December 2008.

[7] Pereira, F., Mitchell, T. and Botvinick, M., “Machine learning classifiers and fMRI: a tutorial overview,” *NeuroImage*, vol. 45, pp. S199-S209, November 2008.

[8] Krusienski, D.J., Grosse-Wentrup, M., Galan, F., Coyle, D., Miller, K.J., Forney, E. et al., “Critical issues in state-of-the-art brain-computer interface signal processing,” *J. Neural Eng.*, vol. 8, pp. 1-8, March 2011.

[9] Lemm, S., Blankertz, B., Dickhaus, T., and Muller, K., “Introduction to machine learning for brain imaging,” *NeuroImage*, to be published.

[10] Berka., C., Levendowski, D.J., Cvetinovic, M.M., Petrovic, M.M., Davis, G., Lumicao, M.N. et al., “Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset,” *Int. J. of Human-Computer Interaction*, vol. 17, pp, 151-170, 2004.

[11] Freeman, F.G., Mikulka, P.J., Prinzel, L.J., and Scerbo, M.W., “Evaluation of an adaptive automation system using three EEG indices with a visual tracking task,” *Biological Psychology*, vol. 50, pp. 61-76, May 1999.

[12] Wilson, G.F. and Fisher, F., “The use of cardiac and eye blink measures to determine flight segment in F4 crews,” *Aviation, Space and Environmental Medicine*, vol. 62, pp. 959-961, October 1991.

[13] Wilson, G.F. and Russell, C.A., “Operator functional state classification using psychophysiological features in an air traffic control task,” *Human Factors*, vol. 45(3), pp. 381-389, Fall 2003.

[14] Wilson, G.F. and Russell, C.A., “Real-time assessment of mental workload using psychophysiological measures and artificial neural networks,” *Human Factors*, vol. 45(4), pp. 635-643, Winter 2003.

[15] Mikulka, P.J., Scerbo, M.W., and Freeman, F.G., “Effects of a biocybernetic system on vigilance performance,” *Human Factors*, vol. 44(4), pp. 654-664, Winter 2002.

[16] Wilson, G.F., Caldwell, J.A. and Russell, C.A., “Performance and psychophysiology measures of fatigue effects on aviation related tasks of varying difficulty,” *Int. J. of Aviation Psychology*, vol. 17(2), pp. 219-247, 2007.

[17] Lin, Y., Wang, C., Jung, T., Wu, T., Jeng, S., Duann, J. et al, “EEG-based emotion recognition in music learning,” *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 1798-1806, July 2010.

[18] Christensen, J., Doczy, E., Durbin, M., Finomore, V., Funke, M., McKinley, R. et al., “Neuroergonomics deep dive literature review, volume I: neuroergonomics and cognitive state,” Air Force Research Laboratory Technical Report AFRL-RH-WP-TP-2011-0001, November 2010.

[19] Chan, V., Christensen, J., DelRaso, N., Funke, G., Knott, B., Panganiban, A. et al., “Neuroergonomics deep dive literature review, volume 2: neuroergonomics and performance: prediction, assessment and facilitation,” Air Force Research Laboratory Technical Report AFRL-RH-WP-TP-2011-0002, November 2010.

[20] Parasuraman, R. and Wilson, G.F., “Putting the brain the work: neuroergonomics past, present and future”, *Human Factors,* vol. 50(3), pp. 468-474, Fall 2008.

[21] Marshall, S.P., “Identifying cognitive state from eye metrics,” *Aviation, Space, and Environmental Medicine*, vol. 78(5), pp. B165-B175, May 2007.

[22] Laine, T.I., Bauer, K.W., Jr., Lanning, J.W., Russell, C.A. and Wilson, G.F., “Selection of input features across subjects for classifying cremember workload using artificial neural networks,” *IEEE Trans. Sys., Man, and Cyber. Part A*, vol. 32(6), pp. 691-704, November 2002.

[23] Christensen, J.C., Estepp, J.R., Wilson, G.F. and Russell, C.A., “The effects of day-to-day variability of physiological data on operator functional state classification,”, *NeuroImage*, submitted for publication, April 2011.

[24] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G. and Vaughan, T.M., “Brain-computer interfaces for communication and control,” *Clinical Neurophysiology*, vol. 113, pp. 767-791, April 2002.

[25] Haynes, J.-D. and Rees, G., “Decoding mental states from brain activity in humans,” *Nature Reviews Neuroscience*, vol. 7, pp. 523-534, July 2006.

[26] Wilson, G.F. and Russell, C.A., “Performance enhancement in a UAV task using psychophysiologically determined adaptive aiding,” *Human Factors*, vol. 49(6), pp. 1005-1019, December 2007.

[27] Russell, C.A. and Gustafson, S.G., “Selecting salient features of psychophysiological measures,” Air Force Research Laboratory Technical Report AFRL-HE-TP-2001-0136.

[28] Chestek, C.A., Cunningham, J.P., Gilja, V., Nuyujukian, P., Ryu, S.I. and Shenoy, K.V., “Neural prosthetic systems: current problems and future directions,” *31st Annual Int. Conf. IEEE EMBS*, pp.2269-3375, September 2009.

[29] Chi, Y.M., Jung, T.-P., and Cauwenberghs, G., “Dry-contact and noncontact biopotential electrodes: methodological review,” *IEEE Revies in Biomed. Eng.,* vol. 3, pp. 106-119, October 2010.

[30] Comstock, J.R. and Arnegard, R.J., “The multi-attribute task battery for human operator workload and strategic behavior research,” NASA Technical Memorandum 104174, January 1992.

[31] Estepp, J.R., Crossen, S. L. and Christensen, J.C., “Nonstationarity in physiological cognitive state assessment using artificial neural networks,” *33rd Anuual Int. Conf. IEEE EMBS*, submitted for publication.