An algorithm for online detection of temporal changes in operator cognitive state using real-time psychophysiological data

Jordan A. Cannon[1], Pavlo A. Krokhmal*[1], Russell V. Lenth[2], and Robert Murphey[3]

[1]Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA 52242
[2]Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242
[3]Air Force Research Laboratory, Munitions Directorate, Eglin AFB, FL 32542

February 2010

Abstract

We consider the problem of on-the-fly detection of temporal changes in the cognitive state of human subjects due to varying levels of difficulty of performed tasks using real-time EEG and EOG data. We construct the Cognitive State Indicator (CSI) as a function that projects the multidimensional EEG/EOG signals onto the interval [0, 1] by maximizing the Kullback-Leibler distance between distributions of the signals, and whose values change continuously with variations in cognitive load. During offline testing (i.e. when evolution in time is disregarded) it was demonstrated that the CSI can serve as a statistically significant discriminator between states of different cognitive loads. In the online setting, a trend detection heuristic (TDH) has been proposed to detect real-time changes in the cognitive state by monitoring trends in the CSI. Our results support the application of the CSI and the TDH in future closed-loop control systems with human supervision.

Keywords: Cognitive state; psychophysiological data; EEG; EOG; Kullback-Leibler distance; statistical analysis.

1 Introduction

The increasing complexity and sophistication of computing, sensing, and communication technologies paves the way for proliferation of unmanned autonomous systems and platforms, which will replace and/or assist humans in hazardous or resource consuming missions. Yet, the common consensus is that despite the ever growing need for increased autonomy of various unmanned systems and vehicles, the human supervision of such systems is indispensable and critical for mission success.

On the other hand, the large amounts of information produced by these complex systems can place a high demand on a human operator's cognitive load, potentially overwhelming him/her and leading to degraded performance. Thus, the ability of an automated control system to estimate the

*Corresponding author, krokhmal@engineering.uiowa.edu.

current functional state of a human operator/supervisor and supply information that is conducive to human decision-making at the given cognitive level is essential to robust and successful system performance.

The objective of the present endeavor is to advance methods for real-time detection of changes in human operator's functional state that meet real-time or online requirements, and which can be utilized in closed-loop autonomous control systems with human supervision. Generally, operator functional state (OFS) can be defined as the momentary ability of an operator to meet task demands with their cognitive and physiological resources. In the context of this work, the OFS is associated with cognitive load experienced by the operator; with this caveat in mind, we will use both terms interchangeably.

OFS is commonly measured indirectly, e.g., by using overt performance metrics on tasks; if performance is declining, a subpar OFS is assumed. Another indirect measure is the subjective estimate of mental workload, where an operator narrates his/her perceived functional state while performing tasks (Wilson and Russell, 2007).

More direct estimation of OFS can be accomplished via psychophysiological measurements such as electroencephalogram (EEG), electrocardiogram (ECG), and electrooculography (EOG). The responsiveness of psychophysiological signals to changes in operational conditions and, particularly, cognitive load, as well as their capability of measuring OFS are well-known in the literature (see, e.g., Byrne and Parasuraman, 1996; Gevins et al., 1998, 1997; Wilson and Fisher, 1991, 1995, and references therein).

Reviews of methods for assessing OFS using psychophysiological signals can be found in, e.g., Wilson and Russell (2003a, 2007). Many of these methods employ various data mining and pattern recognition techniques to classify mental workload into one of several discrete categories. For instance, given an experiment with easy, medium, and difficult tasks, and assuming that tasks induce varying degrees of mental workload, pattern recognition classifies which task is being performed by the operator for each epoch of physiological data. The most common classifiers are artificial neural networks (ANN) and multivariate statistical techniques, such as stepwise discriminant analysis. Artificial neural networks have proved especially effective at classifying OFS as they account for the non-linear and higher order relationships often present in physiological data, and often achieve classification accuracy greater than 80%. Also, data mining methods have been successfully employed for detecting changes in brain behavior during epileptic seizures (see, e.g., Pardalos et al., 2004, and references therein).

In this paper, we pursue an approach to detection of temporal changes in operator's cognitive state, which does not rely on ANN techniques that can be relatively computationally expensive, and is based on monitoring trends of the Cognitive State Indicator (CSI). The CSI is constructed as a function of the real-time psychophysiological data that maps the current cognitive state to a continuous range, e.g. $[0, 1]$, and as such indicates the degree to which the operator is engaged. We show that CSI can also serve as a classifier for discrete states (e.g., cognitive states corresponding to "low", "medium", or "high" task load, etc.)

The indicator-based techniques have been used for assessment of OFS in a number of studies. One such indicator, termed the Engagement Index (EI), has been studied extensively for its ability to detect real-time changes in task engagement (Pope et al., 1995; Freeman et al., 1999; Prinzel et al., 2000). Pope et al. (1995) introduced the EI via EEG wave bands alpha, beta, and theta as $\beta/(\alpha+\theta)$ to be "maximally sensitive to changes in task demand", and used the relative trending of the EI to

detect change. Freeman et al. (1999) used the absolute value of EI, instead of its relative changes, to estimate the OFS. By considering tasks with two different load levels (medium and high), Prinzel et al. (2000) observed that the EI was not successful in discriminating OFS under the different task loads. It can be argued that this may have been a result of the EI not accounting for between-subject differences, for it is constructed using features that are sensitive for engagement on average across all subjects.

A participant-, or subject-specific index of task loading that takes into account the specifics of each subject's response has been employed by Smith et al. (2001), who sought to establish an index that varied consistently with task load, and whose performance was reliable even under unstructured, complex task conditions. It was observed that the index increased consistently when the task load increased, and its values differed significantly between the different task loads.

The present work advances the index-based methods for estimation of OFS using psychophysiological data. Namely, we develop an *online* algorithm capable of detecting changes in the operator cognitive load dynamically, or "on the fly", by interpreting real-time EEG signals via the CSI measure, which is similar to the participant-specific index of Smith et al. (2001). By an online algorithm we mean a computational procedure that may not necessarily meet the real-time requirements, in the sense that its processing times exceed the system's characteristic time unit, but the corresponding time delay is relatively short so that the computational results retain their time value. In our specific case, the proposed computational procedures are generally amenable to real-time implementation, but the delays can be introduced by the need to stabilize the highly volatile EEG readings by aggregating them over time windows of several (usually three to five) seconds. To achieve desired detection accuracy, the classifier then needs to wait until newer data, which corresponds to the cognitive state of interest, passes through the window and older data phases out. The effect of window size on the system latency is discussed in Coyle et al. (2005) and Coyle (2009).

It is worth mentioning that the presented approach can find applications in the broad area of brain-computer interfaces (BCI). For a review of BCI methods and applications, see, among others, Lotte et al. (2007); Coyle (2009), and references therein.

The remainder of the paper is organized as follows. In section 2 we discuss the dataset and data processing from which the CSI is formulated and evaluated. In section 3, the derivation of the CSI is presented in detail. Next, in section 4 we present methods and results regarding CSI's ability to discriminate between task load levels offline, or on average. The results support the conclusion that the CSI is a statistically significant indicator of task load, and thus can serve as an accurate measure of cognitive load. In section 5, we present methods and results regarding CSI's online detection of changes in OFS. The results support the conclusion that the CSI is capable of detecting changes in OFS, and furthermore, demonstrate that when coupled with the Trend Detection Heuristic (TDH), the CSI could facilitate dynamic task allocation in future adaptive aiding systems.

2 The dataset and data preprocessing

The dataset used in this study originated from experiments conducted at Wright-Patterson AFB in 2007. Data was available for three subjects (A, E, and F), each of whom performed two 14-minute trials, which consisted in supervising a simulated bombing mission that involved four unmanned aerial vehicles. The simulation was designed to present the subjects with tasks of 3 levels of dif-

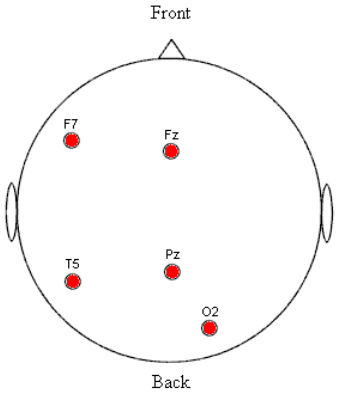

Figure 1: Locations of the EEG electrodes

ficulty: low (L0), medium (L1), and high (L2). The L0 was the baseline state and encompassed most of each trial. During a trial, the tasks of medium and high difficulty (L1 and L2) were present four times each, in a random order, each lasting approximately 20 seconds. Every trial began in the L0, and each period of L1 or L2 was followed by L0, thus allowing time for the subject to recover. The experimental design assumed that varying task loads induced corresponding levels of cognitive load on the subject. It is important to emphasize that since the tasks involved monitoring four UAVs executing a bombing mission, the tasks were very visual in nature and were expected to engage the visual processing centers of the brain; see Wilson and Russell (2007) for the complete experimental design and task details.

The psychophysiological data collected during each trial consisted of eight channels of ECG, EEG, and EOG, all recorded at a sampling frequency of 200 Hz. The EEG channels came from five scalp sites of the 10/20 International electrode system: F_7, F_z, P_z, T_5, and O_2 (see Fig. 1). In previous studies, these sites have been found to provide good discrimination between task levels in complex cognitive tasks (Wilson and Russell, 2007). In particular, Smith et al. (2001) have found that theta power measured from frontal-midline locations, e.g., F_z, increased with increasing cognitive load, and alpha power measured from parietal locations, e.g., P_z, routinely decreased with increasing cognitive load. Beta power was found to be strong in the parietal region of the brain, e.g., P_z and T_5, and the feature was weighted heavily in neural network methods (Gevins et al., 1998; Wilson and Russell, 2003b). Vertical and horizontal EOG signals, termed VEOG and HEOG respectively, were recorded for two purposes: as a measure of cognitive load, and for elimination of blink artifacts in affected EEG signals. In the present analysis, only the EEG and EOG data were used.

The data was pre-processed by an online adaptive filter that eliminated blink artifacts affecting the F_z electrode, whose close proximity to the eye rendered it most susceptible to contamination (see Figs. 1 and 2). The adaptive filter incorporated the VEOG (s_t^{v}) and HEOG (s_t^{h}) signals as reference inputs to de-contaminate the F_z signal, s_t^{fz}, for every time moment t. The artifact-free signal, $\hat{s}_t^{\mathrm{fz}}$, is obtained as

$$\hat{s}_t^{\mathrm{fz}} = s_t^{\mathrm{fz}} - \hat{s}_t^{\mathrm{v}} - \hat{s}_t^{\mathrm{h}},$$

where $\hat{s}_t^{\mathrm{v}}$ and $\hat{s}_t^{\mathrm{h}}$ are the filtered VEOG and HEOG reference signals, respectively:

$$\hat{s}_t^{\mathrm{v}} = \sum_{m=1}^{M} h_m^{\mathrm{v}} s_{t+1-m}^{\mathrm{v}}, \quad \hat{s}_t^{\mathrm{h}} = \sum_{m=1}^{M} h_m^{\mathrm{h}} s_{t+1-m}^{\mathrm{h}},$$

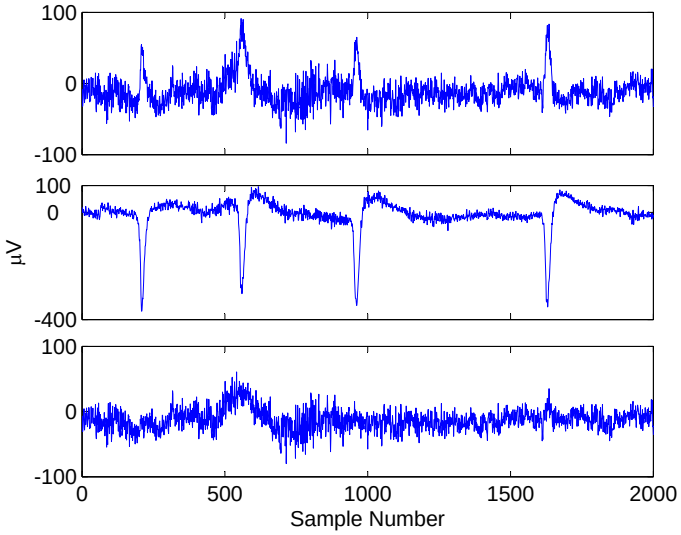

Figure 2: Unfiltered F_7 signal with blink artifacts (top); VEOG signal (middle); filtered F_7 signal (bottom)

and the filter coefficients h_m^{v} and h_m^{h} are obtained by minimizing the weighted squares

$$\min \sum_{i=M}^{n} \lambda^{n-i} \left[\hat{s}_i^{\mathrm{fz}}\right]^2,$$

where λ is the "forgetting factor". The filter coefficients are updated for each time instant n via a recursive least-squares (RLS) algorithm; see He et al. (2007) for complete details.

Once eye blink artifacts were removed, the psychophysiological signals were transformed to frequency domain using Fast Fourier Transform (FFT) for every epoch, usually three to five seconds, and the powers of the standard alpha, beta, gamma, etc., wavebands were computed. The powers of particular EEG/EOG wavebands are well known to correlate with changes in cognitive load: for instance, alpha waveband power, 8–12 Hz, increases with relaxation, while theta waveband power, 5–8 Hz, generally decreases with relaxation (see, for instance, Smith et al., 2001, and others). The powers of various wavebands computed from the EEG/EOG data are henceforth referred to as "features".

No single waveband (feature), however, has been found capable of serving as a robust indicator of a subject's cognitive load during the trials. Thus, the present endeavor is concerned with identifying a subset of features, along with the corresponding functional form that would be able to detect changes in the subject's cognitive state in real time.

3 Cognitive State Indicator function

The presented approach to real-time change detection of operator cognitive state consists in constructing a scalar measure, or score that takes values within a prescribed range, e.g. [0, 1], such that its smaller values indicate a low cognitive load, and larger values indicate an increased cognitive

load. Presuming that different levels of task difficulty induce the corresponding changes in the operator cognitive state, it is required that the score assumes smallest values during L0 periods, and largest values during L2 periods.

In effect, the sought measure of cognitive load should be a robust classifier for distinguishing the L0 and L2 states; notice however, that this does not automatically imply an ability to discriminate the intermediate L1 from the L2 or the L0. In reality, subjects must perform tasks on a continuum of difficulty, thus, a robust index should not be tailored to a particular "intermediate" level task. Rather, it is reasonable to define the CSI on the extremes of task load, and then assume that tasks of intermediate difficulty, like the L1, will fall within the range of the score.

The Cognitive State Indicator function is constructed by selecting a subset of features that exhibit the greatest statistical difference in the L0 and L2 states; then a projection of the corresponding feature vector is obtained that further maximizes the difference between distributions of features in the aforementioned states.

A candidate set of features used for selection of the corresponding subset, subsequently employed for construction of the CSI, is shown in Table 1. The candidate set mainly contained sub-band powers of theta, measured from the frontal lobe, and sub-band powers of alpha, measured from both the frontal and parietal lobes; these features have been shown to vary with task load in many studies (Gevins et al., 1997; Inouye et al., 1994; Ishii et al., 1999; Klimesch et al., 1993). In addition to these traditional features, a measure of theta at VEOG was included in the candidate set. This feature has been shown to vary with task load in similar UAV simulations, and has successfully been used in OFS classifiers (Wilson and Russell, 2003b). Lastly, two beta features were included to further augment the candidate set, as numerous studies have consistently demonstrated its sensitivity to task load (Freeman et al., 1999; Wilson and Fisher, 1995; Wilson and Russell, 1999; Ginter et al., 2005).

Table 1: Candidate feature set

Electrode	**Features**
VEOG	theta (5–8 Hz)
F_z	theta (5–6, 6–7, 7–8 Hz), alpha (8–10, 10–12 Hz)
P_z	theta (5–6, 6–7, 7–8 Hz), alpha (8–10, 10–12 Hz)
F_7	theta (5–6, 6–7, 7–8 Hz), alpha (8–10, 10–12 Hz)
T_5	alpha (8–10, 10–12 Hz), beta (14–30 Hz)
O_2	alpha (8–10, 10–12 Hz), beta (14–30 Hz)

In this study, we used the traditional cross-validation technique in construction of the CSI, whereby one of the two trials ("training trial") of each subject was used to select a subset of features from the candidate set, which were then optimized with respect to each subject. The derived CSI was then tested using the data from the other trial ("testing trial") for its ability to detect changes in the cognitive state; the roles of the trials were then reversed.

The training trial data was partitioned into three data sets containing the data corresponding to the L2, L1, and L0 tasks, respectively (e.g., data with timestamps corresponding to the four L2 tasks were consolidated into a single L2 training set, similarly for the L1 and L0 data). For each of the three training subsets, the features from the candidate set were computed via FFT for every three to

four seconds of EEG/EOG data. An illustration of a distribution of feature realizations is furnished in Fig. 3, which displays the distribution of alpha powers of F_z and F_7 signals for subject A.

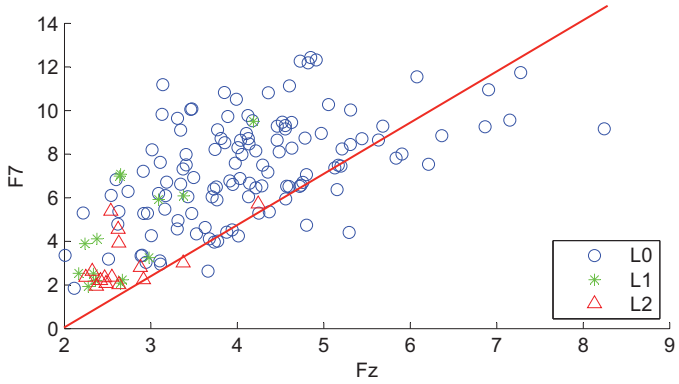

Figure 3: An example of the distribution of features F_z and F_7 in the alpha band

Then, discrimination among cognitive states induced by task loads L0, L1, and L2 can be facilitated when the distributions of the features are maximally different. As a measure of distance between distributions, we used the Kullback-Leibler (KL) distance, also known as KL divergence. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the vector of n EEG/EOG features, whose joint conditional distribution, given that it corresponds to operator's cognitive state $\omega_i \in \Omega$, is equal to $f_i(\mathbf{x}) = f_i(\mathbf{x}|\omega_i)$; in our case, we have $\Omega = \{\mathrm{L2}, \mathrm{L0}\}$. Then, the (symmetric) KL distance between distributions f_0 and f_2 is defined as (Kullback and Leibler, 1951; Kullback, 1968)

$$J_{02}[X_1, \ldots, X_n] = \int_{\mathbf{x}} [f_0(\mathbf{x}) - f_2(\mathbf{x})] \ln \frac{f_0(\mathbf{x})}{f_2(\mathbf{x})} \, \mathrm{d}\mathbf{x}. \tag{1}$$

Online computation of the KL distance J_{02} between the operator states $\{\omega_i, \omega_j\} = \{\mathrm{L0}, \mathrm{L2}\}$ can be facilitated by assuming that the joint distributions $f_i(\mathbf{x})$ are multivariate normal, $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i \in \{0, 2\}$, in which case it can be shown that

$$\begin{aligned} J_{02}[X_1, \ldots, X_n] = {} & \frac{1}{2} \operatorname{Tr}\left[(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_2^{-1} - \boldsymbol{\Sigma}_0^{-1})\right] \\ & + \frac{1}{2} \operatorname{Tr}\left[(\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_2^{-1})(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_2)^{\top}\right], \end{aligned} \tag{2}$$

where $\operatorname{Tr}\mathbf{A}$ is the trace of a square matrix $\mathbf{A}$.

The problem reduces, therefore, to identifying a subset of features $\{X_{i_1}, \ldots, X_{i_m}\}$ from the candidate set, whose distributions in states L0 and L2 have the largest KL distances. Alternatively, a simpler heuristic method can be employed for selection of such a subset, which is based on the well-known property of the KL distance that in the case when the features $X_1, \ldots, X_n$ are mutually independent, the distance function $J_{ij}[X_1, \ldots, X_n]$ is additive (Kullback and Leibler, 1951):

$$J_{ij}[X_1, X_2, \ldots, X_n] = \sum_{k=1}^{n} J_{ij}[X_k]. \tag{3}$$

Thus, by rank-ordering the features in the candidate set with respect to their individual KL divergences in states L0 and L2:

$$J_{02}[X_{(1)}] \geq J_{02}[X_{(2)}] \geq \ldots \geq J_{02}[X_{(n)}], \tag{4}$$

and selecting the top several features one obtains a subset with the most discriminatory information between the states L0 and L2 in the training data. Although the additivity property (3) does not fully hold for the training data, since the features are not independent, it has been observed that the subsets selected based on (4) led to superior results during testing of the constructed CSI. This may be explained by the fact that feature selection based on maximization of $J_{02}[X_{i_1}, \ldots, X_{i_k}]$ results in "overfitting" the training data, which leads to inferior results during the testing stage. In view of this, the latter method was utilized to select the features for the CSI. For the dataset used in this study, the best results have been achieved with the subset of selected features containing four or five features, which is also consistent with earlier similar studies (Pope et al., 1995; Smith et al., 2001).

Once the vector of m features $\mathbf{X} = (X_{i_1}, \ldots, X_{i_m})$ with the largest KL distances between states L0 and L2 has been identified from the candidate set in Table 1, the sought scalar measure of cognitive load can be obtained as a projection of the vector $\mathbf{X}$ onto some hyperplane in $\mathbb{R}^m$:

$$Y = \mathbf{a}^\top \mathbf{X}, \tag{5}$$

where $\mathbf{a} \in \mathbb{R}^m$ is some constant vector. If the features are jointly normally distributed as described above, it is easy to see that the projection Y of the vector $\mathbf{X}$ in states induced by L0 and L2 has normal distributions $N(\mu_i, \sigma_i^2)$, whose means and variances are given by

$$\mu_i = \mathbf{a}^\top \boldsymbol{\mu}_i, \quad \sigma_i^2 = \mathbf{a}^\top \boldsymbol{\Sigma}_i \mathbf{a}, \quad i = 0, 2.$$

The KL distance between the L2 and L0 projections is then computed by

$$J_{02}[Y] = \frac{1}{2}\left(\frac{\sigma_0^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_0^2}\right) + \frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_2^2}\right)(\mu_0 - \mu_2)^2 - 1. \tag{6}$$

The vector $\mathbf{a}$ in (5) is selected so as to maximize the KL distance $J_{02}[Y]$ in (6) between distributions of the image Y in states L0 and L2 (see Fig. 3):

$$\max_{\mathbf{a} \in \mathbb{R}^m} \frac{\mathbf{a}^\top \boldsymbol{\Sigma}_0 \mathbf{a}}{\mathbf{a}^\top \boldsymbol{\Sigma}_2 \mathbf{a}} + \frac{\mathbf{a}^\top \boldsymbol{\Sigma}_2 \mathbf{a}}{\mathbf{a}^\top \boldsymbol{\Sigma}_0 \mathbf{a}} + \frac{\mathbf{a}^\top \mathbf{M}\, \mathbf{a}}{\mathbf{a}^\top \boldsymbol{\Sigma}_0 \mathbf{a}} + \frac{\mathbf{a}^\top \mathbf{M}\, \mathbf{a}}{\mathbf{a}^\top \boldsymbol{\Sigma}_2 \mathbf{a}}, \tag{7}$$

where the matrix $\mathbf{M} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_0)^\top$ is positive semidefinite. Problem (7) is generally non-convex, but it simplifies significantly if one assumes that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, in which case (7) reduces to

$$\max_{\mathbf{a} \in \mathbb{R}^m} \frac{\mathbf{a}^\top \mathbf{M}\, \mathbf{a}}{\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}}. \tag{8}$$

It is easy to see that the optimal objective value in (8) equals to the maximum eigenvalue of the (positive semidefinite) matrix $\boldsymbol{\Sigma}^{-1}\mathbf{M}$, and thus the optimal $\mathbf{a}^*$ is given by the corresponding eigenvector. In our computational studies it was observed that such an optimal solution to (8) represents a good quality solution to (7), and also produces comparably good results in the testing stage.

The exponentially weighted moving average (EWMA) was employed to smooth the projection Y of the selected features $\mathbf{X}$:

$$\widetilde{Y}_t = \lambda Y_t + (1 - \lambda)\widetilde{Y}_{t-1}, \tag{9}$$

where Y_t represents the image (5) of $\mathbf{X}_t$ at time t, and $\widetilde{Y}_t$ is the corresponding smoothed value; the smoothing parameter λ has been chosen at $\lambda = 0.35$, and the initial smoothed value $\widetilde{Y}_0$ was initialized to the mean $\mathsf{E}[Y]$ of the training data.

Once smoothed, the projection Y was scaled to range from 0 (the lowest cognitive load) to 1 (the highest cognitive load). The scaling was achieved through range restriction and normalization:

$$S_t = (Y^* - Y_*)^{-1}\big[Y_* + \max\{0, \widetilde{Y}_t - Y_*\} - \max\{0, \widetilde{Y}_t - Y^*\}\big], \tag{10}$$

where S_t is the final expression for the Cognitive State Indicator (CSI) function, and Y^*, Y_* are the upper and lower limits, respectively. We truncated the outliers by defining Y^* and Y_* that allow $\widetilde{Y}$ to vary two standard deviations above the sample mean $\overline{Y}_2$ of the L2 training set, and two standard deviations below the sample mean $\overline{Y}_0$ of the L0 training set; any Y that occurred beyond this range was truncated:

$$Y_* = \overline{Y}_0 - \sqrt{2(\sigma_2^2 + \sigma_0^2)}, \quad Y^* = \overline{Y}_2 + \sqrt{2(\sigma_2^2 + \sigma_0^2)}. \tag{11}$$

Here, σ_2^2 and σ_0^2 are the sample variances of Y over the L2 and L0 training sets, accordingly.

It must be emphasized that the parameters of the normalization and range restriction were all defined on a subject's training data; these parameters did not change during testing.

4 Offline classification of cognitive load using CSI

The objective of this section is to demonstrate the ability of the CSI to discriminate between the cognitive states of the operator under the corresponding task loads, L0, L1, and L2. Although the following analysis has been conducted "a posteriori", i.e., not in real time, it will be seen in section 5 that it possesses a substantial explanatory power to elucidate the performance of the CSI in detecting changes in OFS in real time.

Since two sets of data based on two trials have been collected for each subject, we performed the standard cross-validation by using one of the trials as the training dataset for construction of the CSI, and the other for testing, and then reversing the roles of the trials. The CSI has been defined completely, including the subset of EEG/EOG features which best discriminate between the L2 and L0, the projection (5), and the scaling and smoothing parameters, on the training data. Once the CSI was defined, its performance was evaluated on the test data; to this end, the test data was parsed into three testing sets, corresponding to the task load (L2, L1, and L0), during which the data was recorded.

The features which most frequently comprised the CSI across all trials were, in order, VEOG (5–8 Hz), O_2 (14–30 Hz), O_2 (10–12 Hz), and P_z (10–12 Hz). Notice that the alpha, beta, and theta wave bands were represented by at least one of the top four features. Also recall that VEOG (5–8 Hz) and O_2 (14–30 Hz) were two of the features indigenous to our candidate set. Interestingly, they were the most consistently divergent features. The VEOG (5–8 Hz) feature was especially prominent as it was included in the CSI of every trial. This feature is primarily a measure of eye activity and its importance is a reflection of the visual nature of the tasks performed by subjects in the trials; the issue of utilizing EOG data for measuring of cognitive load in this study is further discussed below.

Table 2 displays the means of the CSI across the three task loads for each trial by subject. Greater index scores correspond to higher cognitive loads while lower index scores indicate lower cognitive

loads. Observe that the average CSI monotonically increased from the L0 to the L2 for Subjects A and E, but not for Subject F.

Table 2: Mean values of the CSI across task loads

Subject	Testing trial	L2	L1	L0
A	1	0.807	0.683	0.265
	2	0.725	0.462	0.185
E	1	0.762	0.407	0.150
	2	0.901	0.482	0.247
F	1	0.626	0.634	0.591
	2	0.546	0.639	0.319

To determine which results in Table 2, if any, were statistically significant, a mixed-effect model was formulated for each subject as

$$S = L_i + T_j + L_i T_j + \epsilon_{ij}, \tag{12}$$

where S is the CSI function predicted by: the task load L_i, the trial T_j, the interaction $L_i T_j$ of task load and trial, and finally, an error term, ϵ_{ij}. The sole fixed-effect, task load L_i, was expected to explain the variability in the CSI. The remaining two terms, trial and the task load by trial interaction, specify random effects assumed to be normally distributed with a zero mean and variance unique to each term.

Table 3 contains the ANOVA results for each subject. For Subjects A and E, task load was a significant predictor of the CSI, with p-values well below the .05 threshold. In contrast, there is no evidence that Subject F's CSI varied with task load.

Table 3: Type III test for task load L

Subject	Numerator d.f.	Denominator d.f.	F	p-value
A	2	2	53.36	0.0184
E	2	2	161.21	0.0062
F	2	2	1.85	0.3508

Next, it was of interest to examine whether the CSI can discriminate not only the "extreme" L0 and L2 tasks, but also the "intermediate" L1 task. Recall that L1 data has not been used at all for constructing and training the CSI. To this end, post-hoc analyses were conducted to reveal for which task loads the corresponding CSI means differed significantly from each other. Table 4 contains the p-values to pair-wise tests for the difference of task load means with the type I error controlled by the Bonferroni adjustment. For Subjects A and E, the tests comparing the L2 and L0 CSI means were significant at the .05 threshold. Again, Subject F's CSI was not distinguishable by task load.

These results support CSI's ability to discriminate between the L2 and L0 for Subjects A and E. For these subjects, the CSI also discriminated the L1 with significance or near significance in three of the four pair-wise tests; this despite the index not being optimized to delineate intermediate level

Table 4: Bonferroni-adjusted p-values in pair-wise tests for the difference of task load means

Subject	L2 vs. L0	L2 vs. L1	L0 vs. L1
A	0.030	0.237	0.072
E	0.009	0.033	0.064
F	0.928	1.000	0.620

tasks. The next section will address whether these statistically significant results prove practically significant in real-time OFS change detection. We expect the CSI, which discriminates between task loads with significance, to be proficient in detecting real-time changes in OFS; in contrast, the CSI, which discriminated task loads poorly, like that of Subject F, is expected to be incapable of detecting changes in real-time.

Now we would like to comment on utilization of the EOG data, along with the EEG data, in construction of the CSI and evaluation of the cognitive load. Recall that the VEOG theta feature was included in the CSI in all trials, due to its having the highest KL divergence between the states L0 and L2. This can be viewed as a direct consequence of the predominantly visual nature of the tasks performed by the subjects in the course of the trials (monitoring of a simulated bombing mission on a computer screen). Namely, in such a setup the subjects were unable to engage effectively in tasks of elevated cognitive load without an increase in eye activity required for, e.g., target identification/tracking, etc. In view of that, it was of interest to investigate the effect of exclusion of the VEOG feature from the candidate set during construction of the CSI. Tables 5–7 present the results analogous to those reported in Tables 2–4, respectively, with the distinction that the VEOG feature was not utilized.

As one can observe, exclusion of the VEOG theta feature leads to deterioration of the results, particularly, loss of significance for Subject A (although barely, see Table 6). This is of no surprise, as the VEOG represents the most "valuable" feature in the context of the presented method, due to its having the greatest KL divergence between states L0 and L2. The ability to discriminate with significance among task loads L0, L1, and L2 for Subject A is also compromised (Table 7). Yet, the results of online testing reported in the next section indicate that loss of statistical significance for Subject A due to exclusion of the VEOG data does not translate into a markedly inferior performance, as compared to Subject E. Moreover, it will be seen that online results for all subjects suffer relatively mildly from the removal of the VEOG, which can be considered as an indication that the CSI is relatively robust with respect to feature selection.

5 Online change detection in OFS using CSI

The ultimate objective of the presented endeavor was to devise a method for robust detection of changes in operator's functional (cognitive) state that are induced by varying levels of task difficulty, using the real-time psychophysiological data provided by EEG and EOG sensors. The CSI measure of cognitive load introduced in section 3 allows for reducing the problem of detecting "changes" in the multidimensional time series as provided by the EEG/EOG data to same problem for univariate time series, embodied by the CSI function S_t.

Table 5: Mean values of the CSI function across task loads (VEOG excluded)

Subject	Testing trial	L2	L1	L0
A	1	0.773	0.678	0.382
	2	0.715	0.311	0.150
E	1	0.684	0.248	0.228
	2	0.860	0.394	0.272
F	1	0.600	0.557	0.460
	2	0.566	0.575	0.378

Table 6: Type III test for task load L (VEOG excluded)

Subject	Numerator d.f.	Denominator d.f.	F	p-value
A	2	2	10.5	0.0869
E	2	2	55.11	0.0178
F	2	2	8.89	0.1011

The CSI has been construed so as to assume higher values when the task load elevates, and smaller values when the task load reduces. Thus, periods of increased cognitive load can be identified with "peaks" or "plateaus" in the graph of S_t as a function of time t, and, conversely, reduced levels of cognitive load can be identified by "valleys" in S_t.

A well-known method of finding irregularities or anomalies in time series is the "double-window" technique, where for any given epoch t the most recent $2k$ data points are partitioned into two sequences, or windows of length k: $z_{t-2k+1}, \ldots, z_{t-k}$ and $z_{t-k+1}, \ldots, z_t$ (the lengths of windows need not be the same). Then, by computing an appropriate function of the data in each window, $h_1(t) = h(z_{t-k+1}, \ldots, z_t)$ and $h_2(t) = h(z_{t-2k+1}, \ldots, z_{t-k})$, changes in time series z_t can be detected by spikes in the value of $\Delta(t) = |h_1(t) - h_2(t)|$.

This method works reasonably well on data exhibiting stable patterns, such as ECG, but the EEG/EOG data used in the present study is rather poorly suited for the approaches based on double-window idea since it does not exhibit regular patterns.

Instead, we exercise the idea that detection of elevated (reduced) levels of operator's cognitive load can be accomplished by identifying *trends* in the CSI S_t, i.e. by detecting periods of consistent increase (decrease) in a series of observed values of S_t.

To this end, we present the Trend Detection Heuristic (TDH), which detects positive (negative) trends in the CSI that are initiated with the onset (end) of task loads. The TDH has two main components: an adaptive threshold and a trend detector. The adaptive threshold adjusts to the non-stationary behavior occasionally exhibited by the CSI, where its magnitude during a task may be very different for an identical task occurring later in time. The TDH's second component, the trend detector, identifies task-induced trends in the index. Detecting trends accomplishes two things: first, coupled with information on the index's magnitude, a trend helps discriminate between true task-induced peaks and irregular transient spikes. Second, depending on the length of the trend, tasks

Table 7: Bonferroni-adjusted p-values in pair-wise tests for the difference of task load means (VEOG excluded)

Subject	L2 vs. L0	L2 vs. L1	L0 vs. L1
A	0.133	0.439	0.480
E	0.028	0.050	0.897
F	0.204	1.000	0.249

occurring below the threshold can be identified and the threshold can be adapted accordingly.

Algorithm 1 illustrates the TDH on the particular case of identifying positive trends in the CSI, which are expected to indicate an increase in the cognitive load. Namely, the TDH judges the value S_t of CSI at time t against two criteria: its magnitude relative to an adaptive threshold and how well the previous points, $S_{t-1}, S_{t-2}, \ldots, S_{t-n}$, have trended. Trends are identified by tracking the positive "slopes" between consecutive points over a period of time. The slopes need not be consecutive since noise may cause temporary "reversals" in the trend; reversals will only dissolve a trend if, after a period R_{lim} of time, no point falls above the trend's last highest point. If R_{lim} is not exceeded at some point, tracking of positive slopes and the trend detection continues until a predefined number K_* of positive slopes has been observed in the current trend.

Now that an emerging trend is identified, an elevated level of CSI is signaled if its current observation exceeds the adaptive threshold Z, to determine if a task-induced cognitive load has occurred. If $S_t > Z$, the trend is classified as task-induced. If $S_t < Z$, the emerging trend remains unclassified but continues to be monitored until $\kappa > K^*$, where K^* is a newly defined limit. In summary, the second phase of the algorithm classifies trends in one of two ways: the first is when the trend outright breaches Z_t, and the second, is when the trend is occurring below Z_t, but is sustained long enough to be regarded as being task-induced. K^* quantifies "long enough" by accounting for the distance S_t falls below Z_t; the further below Z, the longer trending must continue in order to be classified. The motivation here is to account for non-stationarity while preventing the classification of trends not associated with a task load. If a task-induced trend is classified below the current Z_t, the threshold is updated using exponentially weighed moving average that weights the magnitudes of previously classified trends with the most recent trend, through a smoothing parameter λ. This mechanism adapts the threshold so it can detect task-induced trends occurring at different magnitudes in the future.

As usual, the parameters of the TDH were calibrated through training and subsequently used during testing. Particularly, the value of K_* ranged between 4 and 6, α ranged in $[0.0,\ 0.35]$, and β ranged in $[5, 6]$; the values of λ and R_{lim} were chosen at 0.3 and 5, respectively. Again, cross-validation was performed, producing two sets of results per subject by swapping training and testing trials.

The metric used to evaluate the performance of TDH as applied to CSI was the percentage of tasks identified correctly by the TDH. Tasks were considered as identified successfully when the TDH signaled a task-induced trend directly following a task's onset. Occasionally, the TDH signaled a trend when there was no associated task load; such instances were classified as false alarms and were tabulated as another metric of performance.

Figures 4 and 5 illustrate the performance of the TDH method on trials of subjects A and F, respec-

Algorithm 1 Trend detection heuristic (TDH)

Input: Testing data $\{S_1, \ldots, S_t\}$, K_*, R_{lim}, λ, α, β
Initialize: $\nu := 0,\ \kappa := 0,\ K^* = \infty,\ S_{\max} := \mu_{\text{test}},\ Z = \mu_{\text{test}} + \alpha\sigma_{\text{test}},\ \mathcal{V} = \emptyset,\ \mathcal{I} = \emptyset$
repeat
  $t := t + 1$
  if $S_t > S_{t-1}$ **then**
    $S_{\max} := S_t,\ \kappa := \kappa + 1,\ \nu := 0$
  else
    $\nu := \nu + 1$
  end if
until $\nu \geq R_{\text{lim}}$ or $\kappa \geq K_*$ or $t = T$
if $\nu \geq R_{\text{lim}}$ **then** /* Case 1: trend dissolves due to a sustained reversal in the trend */
  $\kappa := 0,\ L_{\max} := S_t$
else /* Case 2: trend is sufficiently long and/or is above the adaptive threshold */
  if $S_t \geq Z$ or $\kappa \geq K^*$ **then**
    $\mathcal{V} := \mathcal{V} \cup S_t,\ \mathcal{I} := \mathcal{I} \cup t,\ \kappa := 0,\ \nu := 0,\ T_{\text{up}} := \infty,\ S_{\max} := \mu_{\text{test}},$
    $Z := \min\{Z,\ (1-\lambda)Z + \lambda S_{\max}\}$
  else /* Case 3: trend exists but does not yet meet Case 2 conditions */
    if $\kappa = K_*$ **then**
      $K^* := \left\lceil \frac{Z - S_t}{\sigma_{\text{test}}}\beta \right\rceil + K_* + 1,\ \kappa := \kappa + 1$
    end if
  end if
end if
if $t < T$ **then**
  go to step 3
end if
Output: set $\mathcal{V}$ of trends and set $\mathcal{I}$ of trend detection time epochs

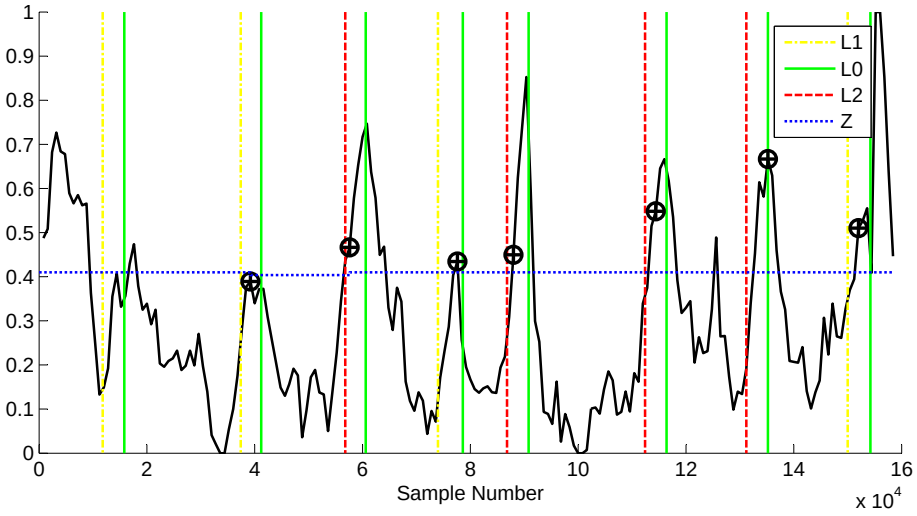

Figure 4: Results of the Trend Detection Heuristic applied to CSI of Subject A, testing trial 2

tively. The vertical lines denote the onsets of the L2, L1, and L0 tasks, and the cross-hair markers identify the points where the TDH has detected an increase in the cognitive load. The horizontal dotted line illustrates the adaptive threshold Z; recall that the initial value of Z is determined during the training trial, and can be adjusted during testing. An instance of threshold adjustment can be seen in Fig. 4, where the first detection was made prior to the CSI exceeding the current threshold value Z, because a positive slope of sufficient duration was observed. Due to this fact, the adaptive threshold was lowered, and, after the subsequent detection occurred well above the threshold value, it was increased.

Table 8 displays the results of evaluating the CSI via the TDH for each trial by subject. The CSI of Subjects A and E performed with an average task detection rate of 81.25%. The CSI of subject F proved rather ineffective at detecting task-induced changes in cognitive load. Potential reasons for this failure are considered below. If the results of subject F are included, the average task detection rate of the CSI was 58.33%, with less than one false alarm per trial on average. With the results of Subject F excluded, it took on average 9.67 seconds to detect a change in the cognitive load; recall that elevated-load tasks, L1 and L2, lasted for about 20 seconds. This delay is due to the TDH algorithm, which needs the CSI to maintain a trend over a certain number of epochs to register a change in cognitive state, as well as due to the size of the moving window. Since shorter windows lead to more volatility in the CSI values and thus jeopardize robust detection, the window size was selected in this study so as to achieve a balance between detection accuracy and detection latency. In addition, statistical significance of the CSI as a discriminator of cognitive load makes it amenable to online classification of the cognitive state, once a change has been detected. Namely, for L1 and L2 tasks, whose onset was correctly identified by the TDH, only one misclassification has occurred across trials of Subjects A and E.

Table 9 reports the correponding results in the case when the VEOG signal is excluded from consideration. In comparison to the results in Table 8, exclusion of the VEOG feature leads to deterioration of detection rates for Subjects A and E, while the detection rates for Subject F remain similarly poor. Considering that the VEOG feature had the highest KL divergence between states L0 and L2,

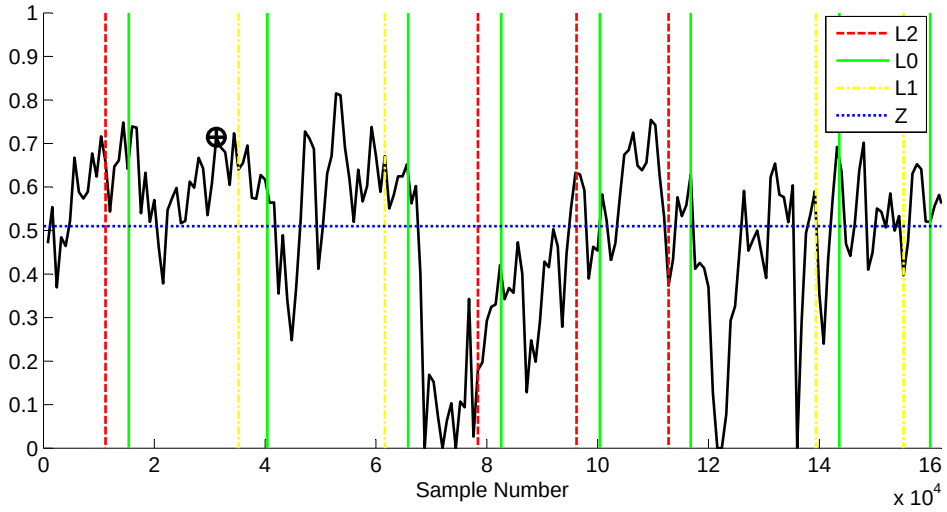

Figure 5: Results of the Trend Detection Heuristic applied to CSI of Subject F, testing trial 1

Table 8: Change detection rates due to TDH applied to the CSI

Subject	**Trial**	**Detection rate (%)**	**# of false alarms**	**Online classification accuracy (%)**
A	1	75.0	1	100.0
	2	87.5	0	100.0
E	1	75.0	0	100.0
	2	87.5	2	85.71
F	1	0.0	1	0.0
	2	25.0	1	0.0

Table 9: Change detection rates due to TDH applied to the CSI (VEOG excluded)

Subject	Trial	Detection rate (%)	# of false alarms	Online classification accuracy (%)
A	1	75.0	2	83.33
	2	62.5	0	100.00
E	1	62.5	0	100.00
	2	75.0	2	83.33
F	1	25.0	3	50.00
	2	0.0	2	0.00

whereby it was the most "valuable" feature with respect to construction of the CSI function (see section 3), it can be concluded that the reduction in detection rates due to exclusion of the VEOG feature is relatively mild for Subjects A and E. This can also be regarded as an indication that the proposed technique is relatively robust with respect to selection of features.

The specific features of the dataset used in our study have placed certain limitations on the procedure of testing the TDH heuristic. For instance, due to the relatively short durations of the periods of elevated task load (20 seconds), TDH was used to detect only positive trends in the CSI (recall that the EEG data was processed in 3 to 5 second windows). Thus, the return of a subject to the baseline L0 state was assumed when the CSI S_t dropped below the level S^*_{t*} at which the last elevated state was detected.

Also, it is of interest to note that the positive trends associated with increase of task loads in the trials of subjects A and E had sometimes originated before the actual start of L2 or L1 tasks (see Fig. 4). This can be explained as artifacts of the smoothing filter applied to CSI; another possibility is that due to the cyclical structure of the trials, the subjects may have "anticipated" the onset of the next task of elevated difficulty towards the end of baseline period L0. Such an interpretation is indirectly supported by the fact that the transitions between L1 or L2 periods and the baseline (L0) periods coincide very well with sharp dropoffs in CSI values.

With regard to Subjects A and E, it can be claimed that the CSI can accurately detect real-time changes in OFS. However, this is not the case for Subject F, whose CSI failed to respond consistently to changing task load. The results of online analysis are in accord with the offline analysis presented in section 4, where it was demonstrated that CSI can discriminate between L0 and L2 conditions with statistical significance for Subjects A and E, but not for Subject F. Thus, it can be conjectured that the efficiency of CSI in detecting real-time changes in cognitive state is predicated on it being a statistically significant discriminator of load levels on average.

Several explanations can be offered with regard to poor performance of the CSI for Subject F. Instances when certain OFS measures exhibit markedly inferior results on a particular subject have been observed in the literature (see, e.g., Smith et al., 2001), and are usually attributed to individual differences. It is also possible that a subject may have been unwilling to exert the effort necessary in the L2 condition, or that the difficulty of L2 tasks was not high enough induce the desired cognitive load state. In general, if the implicit assumption that subjects increased their effort proportional to task demands is violated, any attempt to correlate task load to cognitive load is futile.

The TDH algorithm was also accountable for a portion of the inaccuracy attributed to the CSI. On several occasions, the CSI clearly signaled a change in task load, but since it did not trend "long enough", the TDH did not classify the changes as task-induced. An example of this can be seen in Fig. 4, where the CSI clearly increases during the first task, but the TDH fails to indentify it.

6 Conclusions

In this paper, we presented an approach for online detection of temporal changes in operator cognitive state based on psychophysiological data. The introduced cognitive load indicator function (CSI) is capable of accurately measuring the cognitive load and facilitates the detection of its changes in real-time by means of the trend detection heuristic (TDH). When trained to discriminate between cognitive states resulting from low and high difficulty tasks, the CSI has been shown to perform this with statistical significance for two out of three subjects; moreover, for the same subjects the CSI demonstrated near significant results in discriminating the intermediate load tasks as well. The statistical significance of offline discrimination has been demonstrated to translate into robust performance in online change detection, when coupled with the TDH.

7 Acknowledgements

J. Cannon acknowledges support from the National Science Foundation's Graduate Research Fellowship. P. Krokhmal and R. Murphey acknowledge support from the Air Force Office of Scientific Research.

The authors would also like to thank the Human Effectiveness Directorate of the Air Force Research Lab for providing the dataset used in this study.

References

Byrne, E. A. and Parasuraman, R. (1996) "Psychophysiology and adaptive automation," *Biological Psychology*, **42** (3), 249–268.

Coyle, D. (2009) "Neural network based auto association and time-series prediction for biosignal processing in brain-computer interfaces," *IEEE Computational Intelligence Magazine*, **4** (4), 47–59.

Coyle, D., Prasad, G., and McGinnity, T. (2005) "A time-series prediction approach for feature extraction in a brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **13** (4), 461–467.

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., and Scerbo, M. W. (1999) "Evaluation of an adaptive automation system using three EEG indices with a visual tracking task," *Biological Psychology*, **50** (1), 61–76.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., and Du, R. (1998) "Monitoring working memory load during computer-based tasks with EEG pattern recognition methods," *Human Factors*, **40** (1), 79–91.

Gevins, A., Smith, M. E., McEvoy, L., and Yu, D. (1997) "High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice," *Cerebral Cortex*, **7** (4), 374–385.

Ginter, J., Blinowska, K. J., Kamiński, M., Durka, P. J., Pfurtscheller, G., and Neuper, C. (2005) "Propagation of EEG Activity in the Beta and Gamma Band during Movement Imagery in Humans," *Methods of information in medicine*, **44** (1), 106–113.

He, P., Wilson, G., Russell, C., and Gerschutz, M. (2007) "Removal of ocular artifacts from the EEG: a comparison between time-domain regression method and adaptive filtering method using simulated data," *Medical and Biological Engineering and Computing*, **45** (5), 495–503.

Inouye, T., Shinosaki, K., Iyama, A., Matsumoto, Y., Toi, S., and Ishihara, T. (1994) "Potential flow of frontal midline theta-activity during a mental task in the human electroencephalogram," *Neuroscience Letters*, **169** (1–2), 145–148.

Ishii, R., Shinosaki, K., Ukai, S., Inouye, T., Ishihara, T., and Yoshimine, T. (1999) "Medial prefrontal cortex generates frontal midline theta rhythm," *Neuroreport*, **10** (4), 675–679.

Klimesch, W., Schimke, H., and Pfurtscheller, G. (1993) "Alpha frequency, cognitive load, and memory performance," *Brain Topography*, **5** (3), 241–251.

Kullback, S. (1968) *Information Theory and Statistics*, Dover, Gloucester, MA.

Kullback, S. and Leibler, R. A. (1951) "On Information and Sufficiency," *The Annals of Mathematical Statistics*, **22** (1), 79–86.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007) "A review of classification algorithms for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, **4**, R1–R13.

Pardalos, P. M., Chaovalitwongse, W., Iasemidis, L. D., Sackellares, J. C., Shiau, D.-S., Carney, P. R., Prokopyev, O. A., and Yatsenko, V. A. (2004) "Seizure Warning Algorithm Based on Optimization and Nonlinear Dynamics," *Mathematical Programming*, **101** (2), 365–385.

Pope, A. T., Bogart, E. H., and Bartolome, D. S. (1995) "Biocybernetic system evaluates indexes of engagement in automated task," *Biological Psychology*, **40** (1–2), 187–195.

Prinzel, L. J., Freeman, F. C., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. (2000) "A closed-loop system for examining psychophysiological measures for adaptive task allocation," *International Journal of Aviation Psychology*, **10** (4), 393–410.

Smith, M. E., Gevins, A., Brown, H., Karnik, A., and Du, R. (2001) "Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction," *Human Factors*, **43** (3), 366–380.

Wilson, G. F. and Fisher, F. (1991) "The use of cardiac and eye blink measures to determine flight segments in F4-crews," *Aviation Space and Environmental Medicine*, **62** (10), 959–962.

Wilson, G. F. and Fisher, F. (1995) "Cognitive task classification based upon topographic EEG data," *Biological Psychology*, **40** (1–2), 239–250.

Wilson, G. F. and Russell, C. (1999) "Operator Functional State Classification Using Neural Networks with Combined Physiological and Performance Features," in: "Proceedings of the Human Factors and Ergonomics Society 43th Annual Meeting," .

Wilson, G. F. and Russell, C. A. (2003a) "Operator functional state classification using multiple psychophysiological features in an air traffic control task," *Human Factors*, **45** (3), 381–389.

Wilson, G. F. and Russell, C. A. (2003b) "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Human Factors*, **45** (4), 635–643.

Wilson, G. F. and Russell, C. A. (2007) "Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding," *Human Factors*, **49** (6), 1005–1018.