Fall 2009

Statistical analysis and algorithms for online change detection in real-time psychophysiological data

Jordan Cannon
University of Iowa

STATISTICAL ANALYSIS AND ALGORITHMS FOR ONLINE CHANGE DETECTION IN REAL-TIME PSYCHOPHYSIOLOGICAL DATA

by

Jordan Cannon

A thesis submitted in partial fulfillment
of the requirements for the Master of
Science degree in Industrial Engineering
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisor: Assistant Professor Pavlo Krokhmal

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

MASTER'S THESIS

_____

This is to certify that the Master's thesis of

Jordan Cannon

has been approved by the Examining Committee
for the thesis requirement for the Master of Science
degree in Industrial Engineering at the December 2009 graduation.

Thesis Committee: _____
Pavlo Krokhmal, Thesis Supervisor

_____
Yong Chen

_____
Andrew Kusiak

To Lauren

ACKNOLWEDGEMENTS

I would like to thank my advisor, Professor Pavlo Krokhmal, for offering me his unparalleled expertise and valuable insight. Professor Krokhmal inspired me to pursue graduate studies and was the driving force behind my successful bid for a National Science Foundation Graduate Research Fellowship. His devotion and encouragement was instrumental in the completion of this research.

I would like to thank Professor Yong Chen and Professor Andrew Kusiak for serving as members of my examining committee. They contributed to my learning as an engineering student, and they provided valuable feedback on this thesis. Thanks to Professor Russell Lenth for providing consultation on the statistical results. Also, thanks to Zhaohan Yu for donating his time in support of my efforts.

I would like to acknowledge the National Science Foundation for their financial support. Thanks to the Department of Mechanical and Industrial Engineering, the Center for Computer Aided Design and The University of Iowa for generously offering me their resources. A special thanks to the University of Florida and the Air Force Munitions Directorate for sponsoring my research in Florida during the summer of 2009.

Finally, I would like to thank my family for their support throughout my education. Thanks especially to my brother Jake, for his wise counsel and advice. Finally, I would like to thank my wife, Lauren, for her love, kindness, and patience—I look forward to our bright future together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A01-F02	Letter signifies subject A, E or F, and number signifies trial, 01 or 02
ALL	All electrodes
ANN	Artificial neural network
BH	Back hemisphere electrodes (P_z, T_5, O_2)
BSS	Blind source separation
CL	Centerline
DFT	Discrete Fourier transform
ECG	Electrocardiogram
EEG	Electroencephalogram
EI	Engagement Index
EOG	Electrooculography
ERP	Event-related potentials
EVD	Eigenvalue decomposition
EWMA	Exponentially-weighted moving average
FA	False alarm
FH	Front hemisphere electrodes (VEOG, HEOG, F_7, F_z)
FIR	Finite impulse response
FPA	Factor pattern analysis
HEOG	Horizontal electrooculography
HL	High load task
ICA	Independent component analysis
KLD	Kullback-Liebler divergence
LCL	Lower control limit
LL	Low load task
MCEWMA	Moving-centerline exponentially-weighted moving average
MEWMA	Multivariate exponentially-weighted moving average

ML	Medium load task
NO	Non-ocular electrodes (F_z, P_z, T_5, O_2)
O	Ocular electrodes (VEOG, HEOG, F_7)
OFS	Operator functional state
PCA	Principle component analysis
PDM	Peak detection method
RLS	Recursive least-squares
SampEn	Sample entropy
SSI	Subject-Specific Index
SWDA	Stepwise discriminant analysis
TDM	Trend detection method
TVIC	Task-varying independent component
UAV	Unmanned aerial vehicle
UCL	Upper control limit
VEOG	Vertical electrooculography

CHAPTER 1
INTRODUCTION

Modern systems produce a great amount of information and cues from which human operators must take action. On one hand, these complex systems can place a high demand on an operator's cognitive load, potentially overwhelming them and causing poor performance. On the other hand, some systems utilize extensive automation to accommodate their complexity; this can cause an operator to become complacent and inattentive, which again leads to deteriorated performance (Wilson, Russell, 2003a; Wilson, Russell, 2003b). An ideal human-machine interface would be one that optimizes the functional state of the operator, preventing overload while not permitting complacency, thus resulting in improved system performance.

An operator's functional state (OFS) is the momentary ability of an operator to meet task demands with their cognitive resources. A high OFS indicates that an operator is vigilant and aware, with ample cognitive resources to achieve satisfactory performance. A low OFS, however, indicates a non-optimal cognitive load, either too much or too little, resulting in sub-par system performance (Wilson, Russell, 1999).

With the ability to measure and detect changes in OFS in real-time, a closed-loop system between the operator and machine could optimize OFS through the dynamic allocation of tasks. For instance, if the system detects the operator is in cognitive overload, it can automate certain tasks allowing them to better focus on salient information. Conversely, if the system detects under-vigilance, it can allocate tasks back to the manual control of the operator. In essence, this system operates to "dynamically match task demands to [an] operator's momentary cognitive state", thereby achieving optimal OFS (Wilson, Russell, 2007).

This concept is termed adaptive aiding and has been the subject of much research, with recent emphasis on accurately assessing OFS in real-time. OFS is commonly measured indirectly, like using overt performance metrics on tasks; if performance is

declining, a low 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, Russell, 2007). Unfortunately, indirect measures of OFS are often infeasible in operational settings; performance metrics are difficult to construct for highly-automated complex systems, and subjective workload estimates are often inaccurate and intrusive (Wilson, Russell, 2007; Prinzel et al., 2000; Smith et al., 2001).

OFS can be more directly measured via psychophysiological signals such as electroencephalogram (EEG) and electrooculography (EOG). Current research has demonstrated these signals' ability to respond to changing cognitive load and to measure OFS (Wilson, Fisher, 1991; Wilson, Fisher, 1995; Gevins et al., 1997; Gevins et al., 1998; Byrne, Parasuraman, 1996). Moreover, psychophysiological signals are continuously available and can be obtained in a non-intrusive manner, pre-requisite for their use in operational environments.

The objective of this study is to advance schemes which detect change in OFS by monitoring psychophysiological signals in real-time. Reviews on similar methods can be found in, e.g., Wilson and Russell (2003a) and Wilson and Russell (2007). Many of these methods employ pattern recognition to classify mental workload into one of several discrete categories. For instance, given an experiment with easy, medium and hard tasks, and assuming the tasks induce varying degrees of mental workload on a subject, these methods classify which task is being performed for each epoch of psychophysiological data. The most common classifiers are artificial neural networks (ANN) and multivariate statistical techniques such as stepwise discriminant analysis (SWDA). ANNs have proved especially effective at classifying OFS as they account for the non-linear and higher order relationships often present in EEG/EOG data; they routinely achieve classification accuracy greater than 80%.

However, the discrete output of these classification schemes is not conducive to real-time change detection. They accurately classify OFS, but they do not indicate *when*

OFS has changed; the change points remain ambiguous and left to subjective interpretation. Thus, the present study introduces several online algorithms which objectively determine change in OFS via real-time psychophysiological signals.

The following chapters describe the dataset evaluated, discuss the statistical properties of psychophysiological signals, and detail various algorithms which utilize these signals to detect real-time changes in OFS. The results of the algorithms are presented along with a discussion. Finally, the study is concluded with a comparison of each method and recommendations for future application.

CHAPTER 2
DATASET PROCESSING AND STATISTICAL ANALYSIS

2.1 Dataset Description

The dataset utilized in the following analyses originated from experiments conducted at Wright-Patterson AFB in 2007. Data was available for three subjects. Each subject performed two 14-minute trials with unmanned aerial vehicle (UAV) tasks presented at three levels: low, medium, and high, denoted as LL, ML, and HL, respectively. The LL was the baseline state and subsequently encompassed most of each trial. The ML and HL were presented four times each, in a balanced order, lasting approximately 20 seconds each time. Each trial began in the LL and after every ML or HL was presented, it returned to the LL. The experimental design assumed that varying task load induced corresponding levels of cognitive load on the subject. Throughout this study, cognitive load is assumed to be a proxy measure of OFS; thus detecting a change in task load, is synonymous with detecting a change in OFS.

The tasks involved monitoring four UAVs executing a bombing mission. During the ML and HL, subjects performed a visual search of a radar image uploaded at designated waypoints by the UAVs. The subjects panned the radar image, located potential targets, and marked six of them for bombing, according to pre-determined priorities. This had to be accomplished within the 20 second timeframe. The HL was more difficult than the ML as its radar image contained more distracter targets, in addition to imposing more complex targeting priorities. The baseline LL condition simply required the subject to monitor the UAVs flight paths until each waypoint, i.e. ML or HL, was encountered. Overall, the tasks were 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.

The data collected for each trial consisted of eight physiological channels of EEG, EOG, and electrocardiogram (ECG) recorded at a sampling frequency of 200 Hz. The

EEG channels were recorded from five electrodes: F_7, F_z, P_z, T_5, and O_2. These electrodes were affixed to the subject's scalp according to the 10/20 international electrode system, as shown in Figure 2.1. Vertical and horizontal EOG data, termed VEOG and HEOG respectively, were collected for two purposes: primarily, as measures of cognitive load, and secondly, to eliminate blink artifacts in the EEG signals. Finally, one channel of ECG was collected to measure heart rate. Only the EEG and EOG data were used in the following analyses.

From this point forward, each trial will be denoted by a letter followed by a number; the letter represents the subject A, E, or F, and the number identifies whether it was the subject's first or second trial. For instance, A01 denotes data from the first trial of subject A.

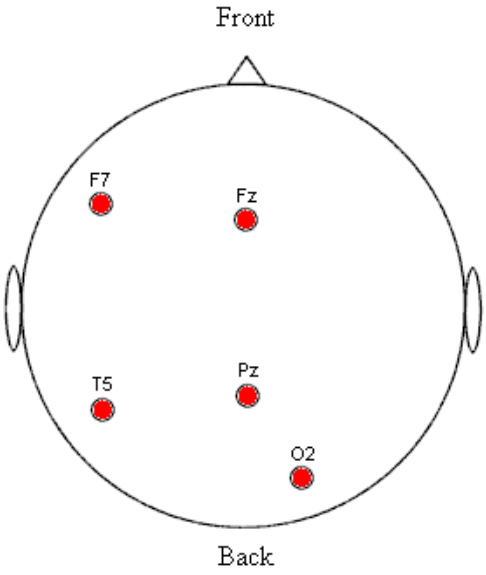

Figure 2.1 EEG electrode diagram

2.2 Pre-Processing the Data

Blink artifacts contaminate EEG signals when the electrical activity from a subject's eye is recorded by the EEG electrodes affixed to their scalp. This interference greatly distorts the EEG data and negatively affects subsequent analyses. Many filters have been developed to remove blink contamination, of interest here however, are filters that remove the artifacts online, rather than in a post-processing scheme. One form of online artifact removal is adaptive filtering. The adaptive filter incorporates the VEOG and HEOG signals as reference inputs to de-contaminate an EEG signal, $s(i)$, for every time moment i. The artifact-free signal, $e(i)$, results by

$$e(i) = s(i) - \hat{r_v}(i) - \hat{r_h}(i) \tag{2.1}$$

where,

$$\begin{aligned} \hat{r_v}(i) &= \sum_{m=1}^{M} h_v(m) r_v(i+1-m) \\ \hat{r_h}(i) &= \sum_{m=1}^{M} h_h(m) r_h(i+1-m) \end{aligned} \tag{2.2}$$

are filtered VEOG and HEOG reference signals, respectively. The $h_v(m)$ and $h_h(m)$ terms represent finite impulse response (FIR) filters of length *M*, which are updated for every time period *i,* to filter the raw VEOG and HEOG signals.

Updating the FIRs is accomplished through a recursive least-squares (RLS) algorithm presented in Table 2.1. The underbars denote column vectors and $R(i)$ is a matrix. In the present analysis, the forgetting factor, λ, was set to .9999 and *M* was set to a length of three. See He, Wilson, and Russell (2007) for complete details.

Figure 2.2 (a) displays an EEG signal recorded at the F_7 electrode, and (b) displays the VEOG signal recorded during the same period of time. Blink contamination is clearly present in (a), where sharp peaks indicate eye blinks. Notice these same

blinks are reflected in (b), the VEOG signal, whose intent is to detect eye activity. The adaptive filter uses the VEOG signal to identify the blinks and to remove them from affected EEG signals, like F_7, where blink activity is considered interference. Figure 2.3 depicts the contaminated F_7 signal in (a), but this time with the corresponding filtered F_7 signal in (b), after adaptive filtering. Notice that the sharp peaks characteristic of eye blinks are no longer present.

2.3 Processing the Data

Once eye blink artifacts were removed, the psychophysiological signals were subjected to the Discrete Fourier transform (DFT) for every epoch, usually three to five seconds long. The DFT transforms signals from the time domain to the frequency domain. To accomplish this, the DFT assumes that the time domain signal is a sum of many sinusoids; this assumption is generally deemed appropriate for psychophysiological signals. The DFT and its inverse are computed by

$$X(\Omega) = \sum_{n=-\infty}^{\infty} x[n]e^{-jn\Omega}$$
$$x[n] = \frac{1}{2\pi}\int_{0}^{2\pi} X(\Omega)e^{jn\Omega}\, d\Omega \tag{2.3}$$

where Ω is the discrete-frequency variable and $x[n]$ is a discrete time series, in this case obtained by sampling a continuous EEG\EOG signal (Phillips et al., 2007).

Utilizing the DFT, a frequency spectrum was created for each epoch where the powers of particular wavebands were computed. Waveband power was computed as the average power for all frequencies falling within the waveband. The powers of particular EEG\EOG wavebands have been shown 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 (Smith et al., 2001).

Table 2.1 Adaptive filtering algorithm

1. **Initialize**:
   1.1. $\underline{H}(i-1) = 0$ where $\underline{H} = \begin{bmatrix} \underline{H_v} \\ \underline{H_h} \end{bmatrix}$ and $\underline{H_v}$ and $\underline{H_h}$ are vectors of the filter coefficients, $h_v(m)$ and $h_h(m)$, respectively
   1.2. $[R(i-1)]^{-1} = \boldsymbol{I}/\sigma$ where $\boldsymbol{I}$ is a $2M$ x $2M$ identity matrix and $\sigma = 0.01$
   1.3. M and λ are user-defined
   1.4. $i = M$
   1.5. n is the length of the signal to be filtered
2. **Calculate** $\underline{K}(i)$:

$$\underline{K}(i) = \frac{[R(i-1)]^{-1}\underline{r}(i)}{\lambda + \underline{r}(i)^T[R(i-1)]^{-1}\underline{r}(i)}$$

where $\underline{r}(i) = \begin{bmatrix} \underline{r_v}(i) \\ \underline{r_h}(i) \end{bmatrix}$ where $\underline{r_v}(i)$ and $\underline{r_h}(i)$ are vectors of the VEOG and HEOG signal, respectively:

$$\underline{r_v}(i) = [r_v(i), r_v(i-1), \ldots, r_v(i+1-M)]^T$$

$$\underline{r_h}(i) = [r_h(i), r_h(i-1), \ldots, r_h(i+1-M)]^T$$

3. **Calculate** $e\left(\frac{i}{i-1}\right)$:

$$e\left(\frac{i}{i-1}\right) = s(i) - \underline{r}(i)^T\underline{H}(i-1)$$

4. **Calculate** $\underline{H}(i)$:

$$\underline{H}(i) = \underline{H}(i-1) + \underline{K}(i)e\left(\frac{i}{i-1}\right)$$

5. **Update** $[R(i)]^{-1}$:

$$[R(i)]^{-1} = \lambda^{-1}[R(i-1)]^{-1} - \lambda^{-1}\underline{K}(i)\underline{r}(i)^T[R(i-1)]^{-1}$$

6. **Calculate** $e(i)$:

$$e(i) = s(i) - \underline{r}(i)^T\underline{H}(i)$$

Note: This is the vector equivalent of (2.1)

7. $i = i + 1$
8. **While** $i < n$ **repeat** steps 2 through 8

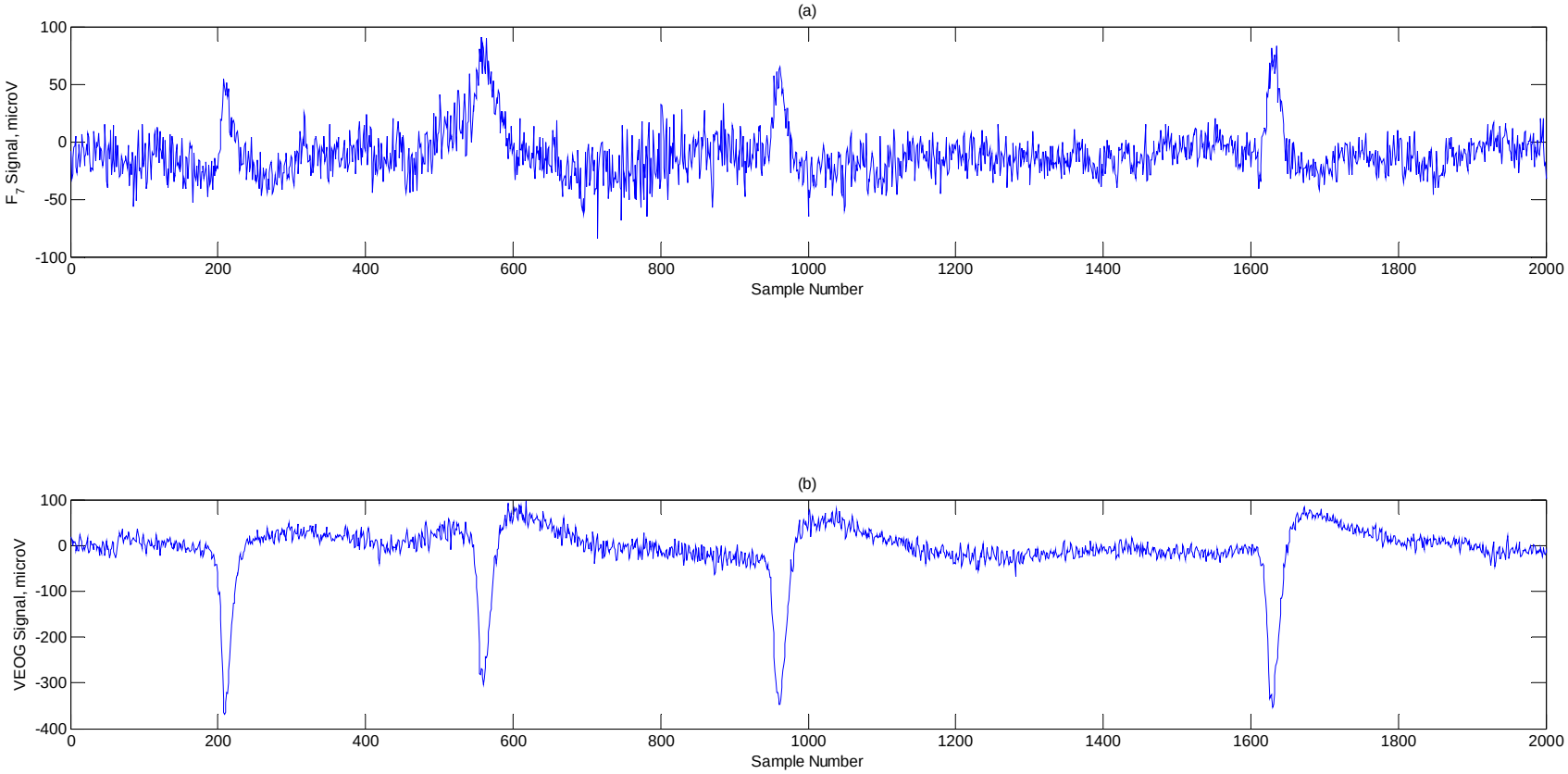

Figure 2.2 Signals with blink artifacts: (a) F_7, (b) VEOG

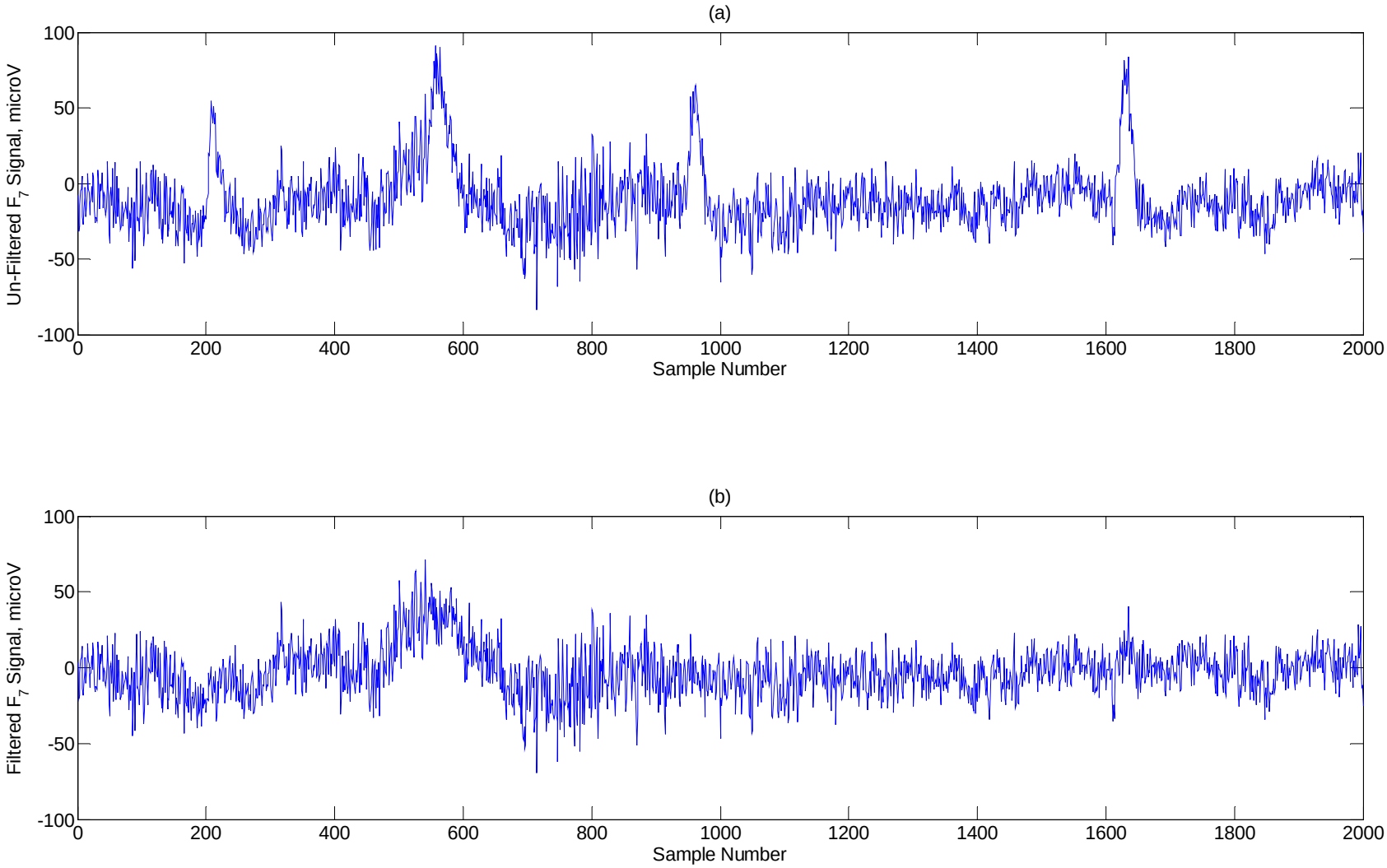

Figure 2.3 F_7 signal: (a) un-filtered, (b) filtered

From this point forward, the waveband powers computed from psychophysiological signals recorded at particular electrodes are referred to as *features*. It is possible to compute hundreds of features for each epoch, as there are many different combinations of wavebands and electrodes. Methods which measure and detect real-time changes in OFS must determine which of these features are indicators of cognitive load, and which are noise.

2.4 Characterizing the Post-Processed Data

Empirical analysis was conducted on several psychophysiological features to characterize their properties and behavior with respect to subject and task load. For this analysis, the features were computed in three-second epochs. It is common to find one-second epochs in the literature, however, they often exhibit highly variable and erratic behavior. Moreover, short epochs induce autocorrelation in the features. In contrast, longer epochs are more stable, but if they are too long, they will be infeasible in operational settings where change occurs instantaneously. Three-second epochs were found to achieve the best balance between short and long epochs.

The features chosen for this analysis are traditional features used in OFS classification methods. The wavebands of these features are: delta, 2-4 Hz, theta, 5-8 Hz, alpha, 9-13 Hz, beta, 14-32 Hz, and gamma, 33-43 Hz. Features from these wavebands have shown to correlate strongly with cognitive load, thus they are often used to classify OFS.

Figure 2.4 displays the distributions of theta, alpha, and beta waveband powers computed across all electrodes for every epoch. The data from every trial were combined to form the distributions, after the data were standardized to zero mean and unit variance with respect to each trial. As shown, the distributions are non-normal, displaying a strong positive skew. The beta waveband power is especially skewed and nearly resembles a Poisson distribution. All three distributions were subjected to statistical tests for non-

normality and each was strongly significant at the .05 level, indicating that features from these wavebands originate from non-normal distributions.

Recall that all seven electrodes record the same five wavebands for every epoch. The reason for this seemingly "redundant" data is that brain waves behave differently when emitted from different regions of the brain. For instance, alpha power is greatest in parietal regions, e.g. P_z, whereas theta power dominates the frontal region, e.g. F_z. While differences exist between topographical locations, there still remains significant correlation, or dependency, between identical wavebands collected at different electrodes; hence not every feature records unique information.

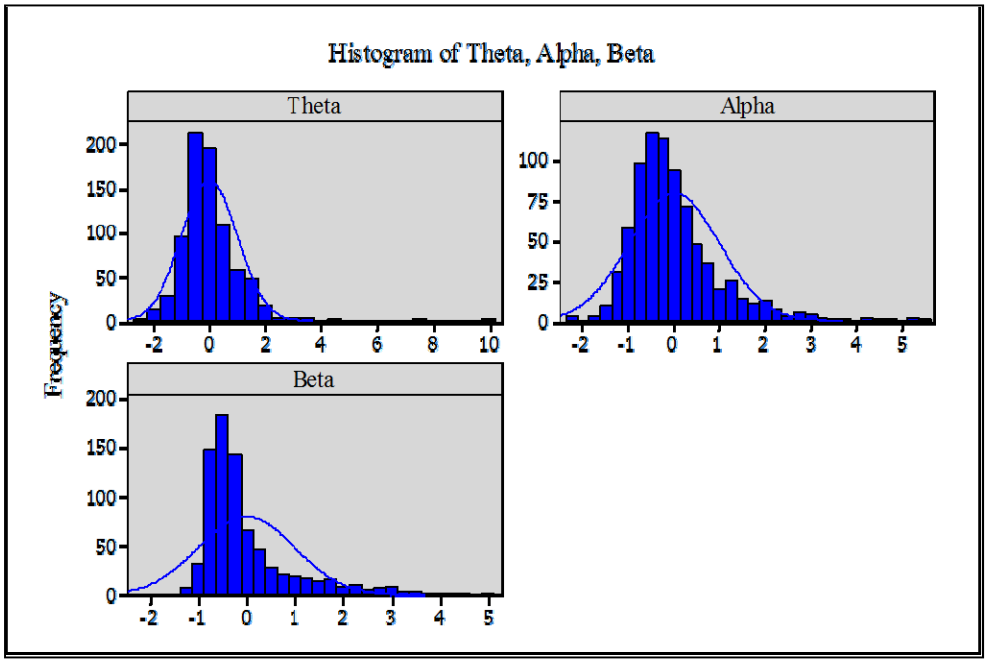

Figure 2.4 Histograms of theta, alpha, and beta wavebands collapsed across all subjects

To illustrate the correlation between features, Table 2.2 contains the correlation coefficients of features recorded at VEOG of F01. Notice that every feature was positively correlated with an average correlation coefficient of .819. In this sense, one

feature from F01's VEOG electrode could virtually represent all the information recorded there. The understanding of this phenomenon implores selecting a variety of features, from different electrodes, so they each contain unique information about the subject's cognitive state.

Many of the features exhibited another form of correlation, called autocorrelation. Autocorrelation occurs when an observation, x_{i+k}, recorded at time $i+k$, is dependent upon a previous observation, x_i, where k is the time lag. Thus, if an autocorrelated feature is observed at time i, it will contain information about its value k time moments in the future (Montgomery et al., 2008). Autocorrelation is identified by evaluating the sample autocorrelation function for a finite time series, $x_1, x_2, \dots, x_n$, defined by

$$r_k = \frac{c_k}{c_0} \quad for\ k = 0, 1, 2, \dots, K \tag{2.4}$$

where c_k and c_0 are computed from

$$c_k = \frac{1}{n}\sum_{i=1}^{n-k}(x_i - \overline{x})(x_{i+k} - \overline{x}) \quad for\ k = 0, 1, 2, \dots, K \tag{2.5}$$

Table 2.2 Intra-node correlation measured at VEOG of F01

	Delta	**Theta**	**Alpha**	**Beta**	**Gamma**
Delta	1.000	0.922	0.939	0.852	0.582
Theta	0.922	1.000	0.949	0.864	0.655
Alpha	0.939	0.949	1.000	0.914	0.663
Beta	0.852	0.864	0.914	1.000	0.853
Gamma	0.582	0.655	0.663	0.853	1.000

Beta and gamma features consistently exhibited autocorrelation. Figure 2.5 displays a sample autocorrelation function for a beta feature measured at T_5 of E01. At lag 1, the autocorrelation is the strongest and then exponentially decreases before oscillating around 0. This behavior is characteristic of a first-order, autoregressive time series (AR[1]).

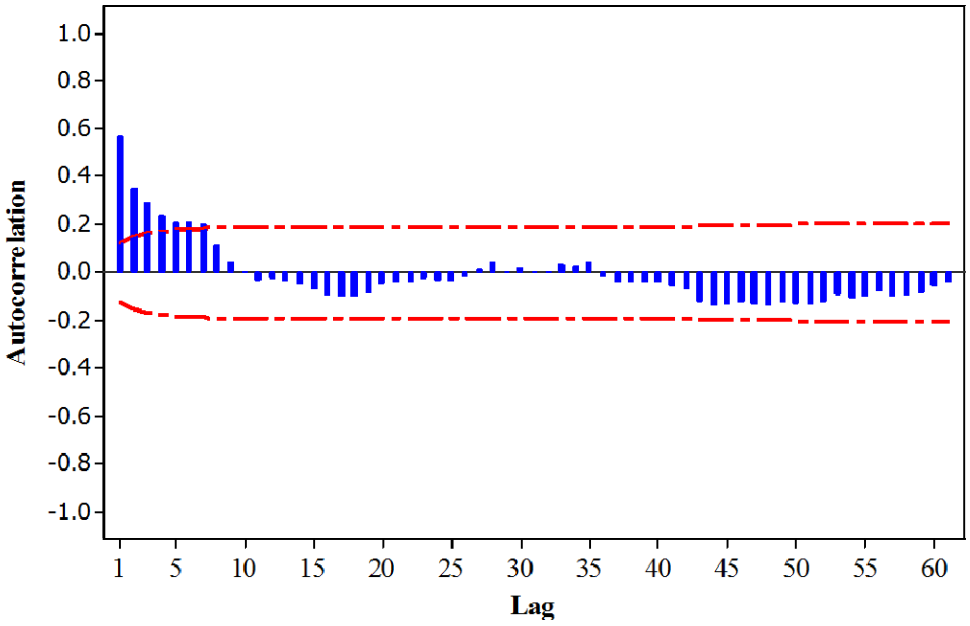

Figure 2.5 Sample autocorrelation function for T_5 beta of E01

Note: The red lines denote .05 significance levels

A survey of feature autocorrelation was conducted across subjects. The results are summarized in Table 2.3. For each waveband, a percentage is displayed representing the proportion of electrodes where the feature exhibited significant autocorrelation. For instance, the alpha waveband for A01 generated features which were autocorrelated in three of the seven electrodes, i.e., 42.86%. As shown, the wavebands which generated the least autocorrelated features across subjects were theta and alpha. In contrast, beta and

gamma wavebands gave rise to features that were autocorrelated regardless of subject or electrode.

It is also important to characterize how the properties of psychophysiological features change with time. More specifically, it is necessary to investigate if these features exhibit stationary behavior. Stationarity implies a statistical equilibrium, where the properties of a time series, such as its autocorrelation or probability distribution, are stable over time. A time series is strictly stationary when the joint probability distribution of $x_i, x_{i+1}, \dots, x_{i+j}$ is the same as the joint distribution of $x_{i+k}, x_{i+k+1}, \dots, x_{i+k+j}$. However, it is often sufficient to classify a time series as stationary as long as it varies around a fixed mean (Montgomery et al., 2008).

Table 2.3 Results of autocorrelation survey

	A01	**E01**	**F01**	**Average**
Delta	0.00%	57.14%	71.43%	42.86%
Theta	14.29%	28.57%	42.86%	28.57%
Alpha	42.86%	28.57%	42.86%	38.10%
Beta	100.00%	100.00%	85.71%	95.24%
Gamma	100.00%	100.00%	100.00%	100.00%

By the latter definition, non-stationary features will exhibit some type of trend, for instance, their mean might increase with time. These trends can be determined by plotting a feature's autocorrelation function. If the plot displays slowly-decreasing autocorrelation with increasing lag, a feature is classified as non-stationary. When this analysis was conducted on a sample of features from each waveband, none of the features exhibited

the symptoms of non-stationarity. Instead, they each varied around a fixed mean like the example shown in Figure 2.6.

Finally, an empirical analysis was conducted to characterize the relationship between features and varying task load; these results are in Appendix A. Although no casual relationship between psychophysiological features and task load are firmly established, many studies have asserted strong associations, such as increasing alpha power with decreasing task load. The features that are most valuable to OFS classifiers, are those that monotonically increase or decrease with changing task load. If this relationship is present, then these features can provide information on a subject's cognitive load. Moreover, it would be ideal if particular features had the same correlation to task load regardless of the subject; however, the results of this empirical analysis found no such "universal" feature. Instead, the features of each subject exhibited unique behavior with respect to task load.

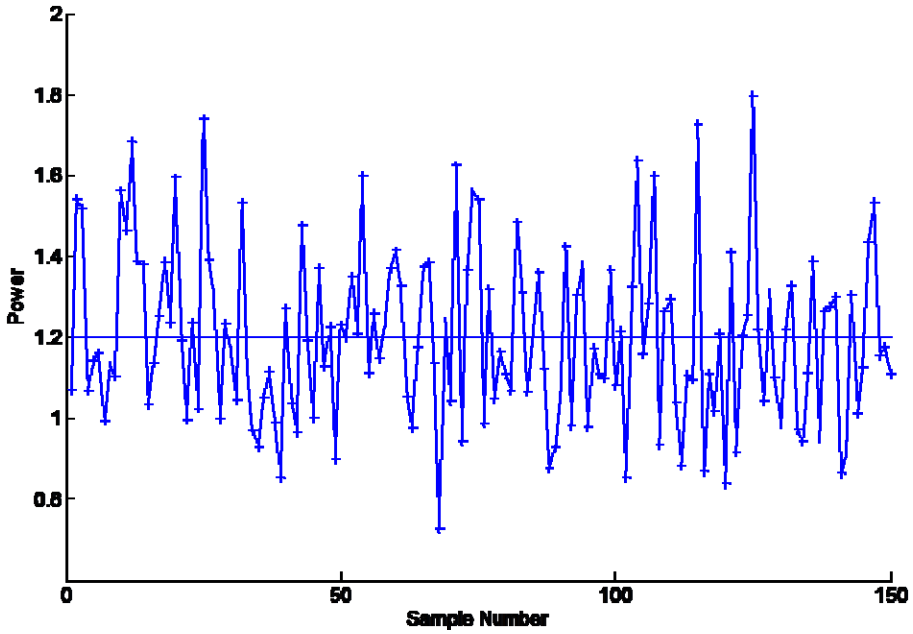

Figure 2.6 Stationary plot for F_z theta of A01

Despite the inability of the features' behavior to generalize across subjects, there was at least one feature for each subject which monotonically increased or decreased with task load. Thus by selecting features tailored to each subject, it is possible to gain reliable information on OFS. In fact, nearly all OFS classifiers adhere to this "subject-specific" approach (Wilson, Russell, 2003a; Smith et al., 2001).

CHAPTER 3
CONTROL CHARTS ON PSYCHOPHYSIOLOGICAL SIGNALS

3.1 Introduction to Control Charts

As mentioned, pattern recognition classifiers are frequently employed to measure OFS in real-time. These methods are complex and may even function like a "black box", data is input and results are output, but the process itself is not easy to interpret. In addition, classification schemes place OFS into discrete categories for every epoch, however, they do not clearly indicate *when* OFS has changed. In this chapter, control charting is presented as a simple alternative that can clearly and objectively detect real-time changes in OFS.

Control charts are frequently used in process control settings to detect when a process goes "out-of-control". They have a simple formulation, yet they remain powerful and effective in many applications. Thus, control charts are a good candidate for detecting real-time changes in OFS. If effective, they could be employed in an adaptive aiding scheme, or at the very least, serve as a visual supplement to OFS classifiers.

3.2 Control Chart Fundamentals

Control charts are graphical displays that plot observations, like those from a psychophysiological feature, over time. The goal of control charts is to detect when a process, such as a subject's cognitive load, has gone out-of-control, so that actions can be taken to address the change.

To accomplish detection, most control charts have two main components: a centerline and a set of control limits. The centerline represents the mean of the feature being plotted, where the mean is calculated from historical data that is deemed "in-control". As long as the process plotted remains in-control, the points will fall randomly around the centerline and within the control limits. If the process goes out-of-control, indicating a shift in the mean, a point or series of points will plot beyond the control

limits. In this sense, control charting is similar to hypothesis testing. For every point plotted, a hypothesis test is conducted with the null hypothesis that the point came from the in-control distribution, $\mu_o = \mu_c$, where μ_c is the centerline or in-control mean. The alternative hypothesis states $\mu_a \neq \mu_c$. The control limits are calibrated so that points that are in-control will only exceed them with a low probability, say .05. If a point does exceed the control limits, the control chart classifies the process as out-of-control and the null hypothesis is rejected in support of the alternative, $\mu_a \neq \mu_c$ (Montgomery, 2009). Control charts are designed so that there is a small probability that an out-of-control signal is a false alarm, commonly known as a type I error. A false alarm is when an out-of-control point occurs by chance, and is not the result of a change in the feature mean. Nonetheless, there are occasions when an in-control process, in this case an acceptable OFS, is deemed out-of-control due to a false alarm.

The control charts evaluated in this chapter plot psychophysiological feature(s) for every three-second epoch. The control charts are calibrated to detect when OFS changes as the result of changing task load. For example, a univariate control chart plots a single EEG feature where the centerline is the feature mean and the control limits are calibrated to be breached during a HL or ML condition. Thus, it is assumed the feature mean will change with increasing task load. In this context, the LL is assumed to facilitate a high OFS while the HL and ML are assumed to cause a low OFS, i.e., cognitive overload. Therefore, when an EEG feature signals out-of-control on a control chart, it is assumed that a change from a high to a low OFS has occurred. A proficient control chart will detect changes for every HL and ML task in a trial. Henceforth, "task" will exclusively refer to the HL and ML task loads.

The present control chart study plotted subgroups of size one, i.e., an epoch constituted one data point. It was deemed inappropriate to form subgroups any greater than size one, e.g., averaging a feature over several epochs. This is because the psychophysiological data was collected continuously, as opposed to periodic sampling, so

subgroups greater than one were not justified. Plotting size one subgroups require specially designed control charts which are discussed in the following sections (Montgomery, 2009).

3.3 Control Chart Assumptions

The psychophysiological features monitored in control charts must satisfy several key assumptions. First, the data is assumed to be normally distributed. If this assumption is violated, type I and type II errors will occur at a greater rate than is advertised by the control limits. In this context, a type II error indicates a failure to detect a true out-of-control condition. Some control charts, such as moving average charts, are very robust to violations of normality.

Control charts also assume that the features are independently distributed. Time series data, common to control charts, frequently violate this assumption in the form of autocorrelation. Control charts have been developed to mitigate autocorrelation, like those that plot the residuals of a time series model. Finally, the features are assumed to be stationary, meaning their in-control mean does not change with time (Montgomery, 2009).

The properties of the psychophysiological features under consideration were discussed in Section 2.4. As demonstrated, some features violated the assumptions of independence and normality. Fortunately, all the features evaluated were stationary, and thus, by utilizing a case-by-case approach, it is possible to use control charts which can accommodate most of the EEG\EOG data.

3.4 Univariate Control Charts

The first control charts discussed are univariate charts for individual measurements, i.e., for subgroups of size one. Different univariate charts are presented,

including those which accommodate autocorrelated or non-normal data. In later sections, these univariate charts are evaluated for their ability to detect tasks.

3.4.1 Exponentially-Weighted Moving Average and Shewart Individuals Control Charts

The exponentially-weighted moving average (EWMA) chart and the Shewart individuals chart are commonly combined to form a robust univariate control charting scheme. The EWMA chart is desirable because it is robust to non-normality and can detect small shifts in the mean. On the other hand, the Shewart individuals chart complements the EWMA with its ability to detect large shifts in the mean. The EWMA is susceptible to the "inertia" problem: it is slow to detect changes which occur in the opposite direction from points previously plotted. In contrast, the Shewart chart is impervious to the "inertia" problem and can detect sudden changes, such as outlier points. Outliers can skew the EWMA for several time periods, thus identifying these points via the Shewart individuals chart provides insight on the behavior of the EWMA.

The upper and lower control limits, denoted as UCL and LCL respectively, and the center line, denoted as CL, for the EWMA are derived by

$$\begin{aligned} UCL &= \bar{x} + L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} \\ CL &= \bar{x} \\ LCL &= \bar{x} - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} \end{aligned} \tag{3.1}$$

where $\bar{x}$ is the in-control mean of the feature, σ is the standard deviation of the feature, λ is a smoothing constant such that $0 \leq \lambda \leq 1$, i is the epoch plotted, and finally, L is a factor that determines the width of the control limits. It is customary to set L to three, so that the UCL and LCL encompass roughly 99.7% of a normal distribution.

The UCL, LCL, and CL for the Shewart individuals chart are derived by

$$UCL = \bar{x} + 3\frac{\overline{MR}}{d_2}$$

$$CL = \bar{x} \tag{3.2}$$

$$UCL = \bar{x} + 3\frac{\overline{MR}}{d_2}$$

where d_2 is a constant determined from available tables and $\overline{MR}$ is the average moving range. Moving range is calculated from two consecutive epochs as

$$MR = |x_i - x_{i-1}| \tag{3.3}$$

where x_i represents the feature's magnitude at epoch i.

On the EWMA chart, instead of plotting the actual feature observed, i.e. x_i, a smoothed version, z_i, is plotted for each epoch i by

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1} \tag{3.4}$$

It is customary to initilize z_0 to the in-control mean of the feature.

3.4.2 Fitting Time Series Models and Control Charting the Residuals

The autocorrelated psychophysiological features typically exhibited first-order autoregressive behavior, modeled as

$$\hat{x}_i = (1 - \phi)\mu + \phi x_{i-1} + \epsilon_i \tag{3.5}$$

where $\hat{x}_i$ is the forecast of the feature made from x_{i-1}, ϕ is a parameter such that $|\phi| < 1$, μ is the expected value of x_i, and ϵ_i are errors which are identically and independently distributed. This model is often recognized as the regression of x_i on x_{i-1} (Montgomery et al., 2008).

Time series models like the one in (3.5) can be used to remove the autocorrelation from affected psychophysiological features, allowing them to be monitored in traditional control charts (Montgomery, Mastrangelo, 1991). To accomplish this, an appropriate time series model is fit to the feature data. This model is then used as a one-step-ahead forecast for the feature, with the residuals (forecast errors) for each epoch plotted on the control chart. Assuming the feature data was modeled correctly, these residuals will be non-autocorrelated. In addition, residuals from the LL will be different than residuals from the ML or HL. Thus, a traditional control chart can calibrate its CL, UCL, and LCL such that, for every ML and HL, the chart detects a change in OFS.

3.4.3 Moving-Centerline Exponentially-Weighted Moving Average Control Charts

One major drawback to the time series modeling approach is that the process of removing autocorrelation is time consuming and subjective. A simpler method is to use the EWMA statistic, as computed in (3.4), in a moving-centerline exponentially-weighted moving average chart (MCEWMA). This method provides an optimal one-step-ahead forecast for a psychophysiological feature which exhibits the behavior of an integrated moving average time series with first order parameters (IMA[1,1]); this model is defined by

$$\hat{x}_i = x_{i-1} + \epsilon_i - \theta\epsilon_{i-1} \tag{3.6}$$

where $\theta = 1 - \lambda$. λ is chosen to minimize the sum of squares of the one-step-ahead forecast errors, e_i, defined by

$$e_i = x_i - \hat{x}_i \tag{3.7}$$

When IMA[1,1] can be assumed and λ is optimized as above, the forecast in (3.6) becomes equivalent to the forecast of x_i from z_{i-1}, where z_{i-1} is the EWMA statistic at epoch $i-1$. In the MCEWMA chart, the z_{i-1} forecast is plotted as the *centerline* for each

epoch *i*, thus in this sense the centerline is "moving". The feature, x_i, is also plotted on the chart for every epoch *i,* and if the forecast error ($x_i - z_{i-1}$) is large enough, it will breach the control limits. As done previously, the UCL and LCL were calibrated to be breached during the HL and ML.

The UCL and LCL for the MCEWMA chart are computed for each epoch *i* by

$$UCL_{i+1} = z_i + L\sigma(i)$$
$$LCL_{i+1} = z_i - L\sigma(i) \tag{3.8}$$

where $\sigma(i)$ is the smoothed standard deviation estimated for each epoch by

$$\hat{\sigma}^2(i) = \alpha e_i^2 + (1-\alpha)\hat{\sigma}^2(i-1) \tag{3.9}$$

where α is a smoothing constant such that $0 \leq \alpha \leq 1$ and $\hat{\sigma}^2(0)$ is estimated by dividing the sum of the squared forecast errors used to derive λ, by *n*, where *n* is the number of observations.

3.5 Multivariate Control Charts

Mastrangelo et al. claim that multivariate charts monitoring several related variables simultaneously, are superior to univariate charts monitoring those variables separately (1996). Figure 3.1 illustrates this phenomenon with an observation x, circled in red, which failed to be detected by the univariate charts monitoring the two variables separately, but was detected by the multivariate chart which monitored the variables simultaneously (the multivariate control limits are represented by the ellipse). Because multivariate control charts monitor variables in combination, rather than separately, they have improved sensitivity to change.

Multivariate control charts account for the covariance structure between features to reduce type II errors, while still conforming to an acceptable type I error level. In contrast, type I errors increase when using multiple univariate control charts. Multivariate

charts make assumptions equivalent to univariate charts, such as, multivariate normality, independence, and stationarity. Similar to univariate charts, multivariate control charts have been designed to accommodate violations of these assumptions, and a sampling of such methods is presented in the following section. Finally, these multivariate control charts will be evaluated for their ability to detect real-time changes in OFS, with the expectation of better performance than their univariate analogues.

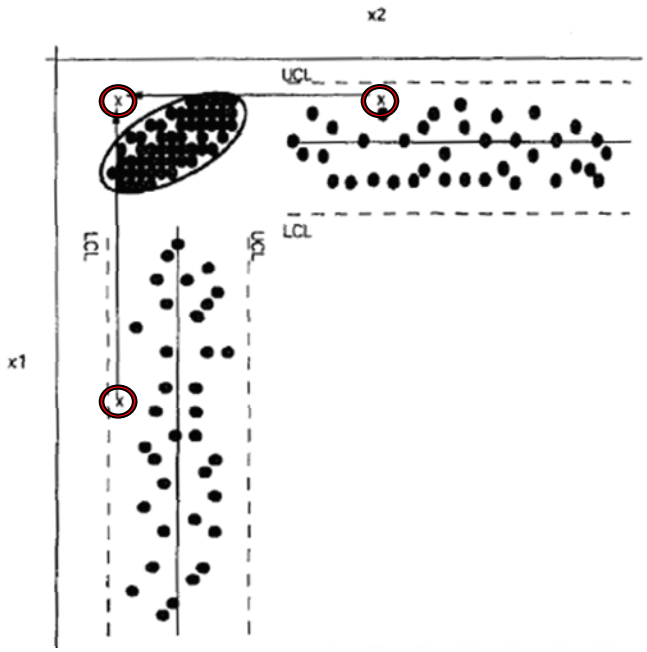

Figure 3.1 Detection illustration for multivariate control charts

Source: Scranton et al., 1996

3.5.1 Hotelling-T^2 Control Charts

The most common multivariate control chart is the Hotelling-T^2 chart, which is the multivariate analogue of the Shewart individuals chart. This chart plots a ${T_i}^2$ statistic against a single UCL for every epoch i. The ${T_i}^2$ statistic is computed by

$$T_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \tag{3.10}$$

where $\boldsymbol{x}_i$ is a p x 1 vector of a realization of p features, $\overline{\boldsymbol{x}}$ is a p x 1 vector of the features' in-control means, and $\boldsymbol{S}$ is an estimate of the features' covariance matrix. Because $\boldsymbol{S}$ is estimated from individual observations (i.e. subgroups of size one), it requires a special computation which utilizes the difference between successive epochs,

$$\boldsymbol{v}_i = \boldsymbol{x}_{i+1} - \boldsymbol{x}_i \quad for\ i = 1,2,\dots,n-1 \tag{3.11}$$

Once the vectors $\boldsymbol{v}_i$ are computed for all successive pairs of epochs, they are arranged to form a matrix $\boldsymbol{V}$, where

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1' \\ \boldsymbol{v}_2' \\ \vdots \\ \boldsymbol{v}_{n-1}' \end{bmatrix} \tag{3.12}$$

$\boldsymbol{S}$ is then computed by

$$\boldsymbol{S} = \frac{1}{2}\frac{\boldsymbol{V}'\boldsymbol{V}}{(n-1)} \tag{3.13}$$

The UCL for the Hotelling-T^2 chart utilizes the beta distribution (β) and is computed by

$$UCL = \frac{(n-1)^2}{n}\beta_{\alpha,\frac{p}{2},\frac{n-p-1}{2}} \tag{3.14}$$

where n is the number of epochs and α is chosen to define an upper quantile in the beta distribution (Montgomery, 2009).

3.5.2 Multivariate Exponentially-Weighted Moving Average Control Charts

The multivariate exponentially-weighted moving average control (MEWMA) is an extension of the univariate EWMA chart. Similar to the univariate case, MEWMA is

more sensitive to small shifts in the mean vector than the Hotelling-T^2 chart. Also, it is robust to violations of normality, frequently exhibited by psychophysiological features. The MEWMA chart monitors $\boldsymbol{z}_i$, a smoothed version of $\boldsymbol{x}_i$, computed by

$$\boldsymbol{z}_i = \lambda \boldsymbol{x}_i + (1-\lambda)\boldsymbol{z}_{i-1} \tag{3.15}$$

where $\boldsymbol{z}_0$ is initialized to zero, assuming the features are made zero mean. For each epoch, $\boldsymbol{z}_i$ is used to compute the ${T_i}^2$ statistic by

$${T_i}^2 = \boldsymbol{z}_i' \Sigma_{\boldsymbol{z}_i}^{-1} \boldsymbol{z}_i \tag{3.16}$$

where

$$\Sigma_{\boldsymbol{z}_i} = \frac{\lambda}{\lambda - 1}\left[1 - (1-\lambda)^{2i}\right]\boldsymbol{S} \tag{3.17}$$

and $\boldsymbol{S}$ is the estimate of the covariance matrix as computed in (3.13). There is a single UCL for the MEWMA chart which is set to h_4, specified by tables provided in the literature (Lowry et al., 1992).

3.5.3 Control Charting Principle Components

Jackson purports that instead of using the original p features to comprise the ${T_i}^2$ statistic, one should use a subset of their principle components (1980). *Principle component analysis* (PCA) reduces the dimensionality of multiple features into a subspace of orthogonal components. These components are chosen to represent the majority of the variation among the original features. Jackson claimed that monitoring a subset of principle components had the following advantages: (1) it greatly improved the control chart's sensitivity to change, (2) the principle components often represented physical characteristics of the process being monitored, improving the interpretation of out-of-control points, and finally, (3) it does not assume multivariate normal data (1980).

To compute the principle components, first the covariance matrix of the features is estimated by,

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' \tag{3.18}$$

where $\boldsymbol{x}_i$ is a p x 1 vector of a realization of the p original features (Montgomery, 2009). It is assumed that the features have been standardized to zero mean and unit variance

Next, $\boldsymbol{S}$ is reduced to a diagonal matrix, $\boldsymbol{L}$, via pre-multiplication and post-multiplication of an orthonormal p x p matrix $\boldsymbol{U}$, such that

$$\boldsymbol{U}'\boldsymbol{S}\boldsymbol{U} = \boldsymbol{L} \tag{3.19}$$

The diagonal elements of $\boldsymbol{L}$, denoted as $l_1, l_2, \dots, l_p$, are the eigenvalues of $\boldsymbol{S}$ and the columns of $\boldsymbol{U}$, denoted as $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_p$, are the associated eigenvectors.

The eigenvalues are found by solving the following determinant,

$$|\boldsymbol{S} - l\boldsymbol{I}| = 0 \tag{3.20}$$

where $\boldsymbol{I}$ is the p x p identity matrix. The results of this equation form a polynomial of order p in l. Solving the roots of this polynomial yields the eigenvalues, $l_1, l_2, \dots, l_p$, from which the eigenvectors forming $\boldsymbol{U}$ are determined by

$$\boldsymbol{u}_j = \frac{\boldsymbol{t}_j}{\sqrt{\boldsymbol{t}_j'\boldsymbol{t}_j}} \tag{3.21}$$

where,

$$[\boldsymbol{S} - l_j\boldsymbol{I}]\boldsymbol{t}_j = 0 \tag{3.22}$$

for $j = 1, 2, \dots, p$ (Jackson, 1980).

Finally, the principle components are computed from each $\boldsymbol{x}_i$ by

$$\boldsymbol{p}_i = \boldsymbol{U}'\boldsymbol{x}_i \tag{3.23}$$

where $\boldsymbol{p}_i$ is a p x 1 vector of principle components. The jth principle component is then, $p_j = \boldsymbol{u}_j{}'\boldsymbol{x}_i$, or in expanded form, $p_j = u_{1j}x_{1i} + u_{2j}x_{2i} + \cdots + u_{pj}x_{pi}$. In essence, PCA transforms p correlated features into p uncorrelated principle components for every epoch, such that each component is a linear transformation of the original features.

Monitoring the p principle components in a control chart is equivalent to monitoring the original p features. However, when PCA is used to dimensionally reduce the features, the resulting control chart is more sensitive to change. Dimensionality reduction is achieved by retaining the subset of the principle components which account for the most variability in the original data. Each principle component accounts for a portion of the original variability; this portion is specified by their associated eigenvalue, l_j. For instance, the proportion of variability which p_j accounts for is just

$$\frac{l_j}{l_1 + l_2 + \cdots + l_p} \tag{3.24}$$

Therefore, eigenvalues, representing their respective principle components, can be ordered by their proportion of variability computed in (3.24). The first m eigenvalues that account for more than 80% of the overall variation, specify the m principle components, $p_1, p_2, \ldots, p_m$, retained as the subset. These selected principle components are represented by an m x 1 vector, $\boldsymbol{w}_i$, which is computed for every epoch by

$$\boldsymbol{w}_i = \boldsymbol{U}_m{}'\boldsymbol{x}_i \tag{3.25}$$

where $\boldsymbol{U}_m$ is a p x m matrix, whose columns, $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m$, are the eigenvectors associated with the m selected principle components.

$\boldsymbol{w}_i$ is then used to compute the ${T_i}^2$ statistic monitored in both the Hotelling-T^2 and MEWMA charts, where ${T_i}^2$ is computed in (3.10) and (3.16), respectively. In these equations, $\boldsymbol{S}$ is just the diagonal matrix $\boldsymbol{L}_m$ of the eigenvalues $l_1, l_2, \ldots, l_m$, and $\boldsymbol{x}_i$ is substituted with $\boldsymbol{w}_i$ for every epoch (Mastrangelo et al., 1996).

There were 35 features that could possibly be monitored for each subject (5 wavebands x 7 electrodes). However, more than half of these features were autocorrelated, hence, they violated multivariate control chart assumptions. Thus, only the non-autocorrelated subset, typically 10-17 features for each subject, was submitted to PCA analysis.

One drawback to monitoring the subset of principle components is the inability to detect shifts which are orthogonal to the subspace defined by the components. The severity of this problem depends on the process, and in some cases, it can be mitigated by monitoring the residuals of predicting $\boldsymbol{x}_i$ from $\boldsymbol{w}_i$. Another drawback is when the principle components do not clearly represent a physical characteristic of the process; in such an instance, their interpretation can be ambiguous (Lowry, Montgomery, 1995; Mastrangelo et al., 1996).

3.6 Real-Time OFS Change Detection Using Control Charts

Before evaluating the various control charts, an analysis was conducted to determine how the means of common psychophysiological features changed across task loads. From a control charting perspective, it is desirable for the changes in the mean to be large, and thus easily detectable. For each subject, features from the theta, alpha, and beta wavebands, recorded from all electrodes, were analyzed. The results are in Table 3.1. The first row is the size of the shift in the feature mean from the LL to the ML. Similarly, the second row is the size of the shift in the feature mean from the LL to the HL. The data reported have been normalized by the respective waveband's data. The average shift from the LL to the ML across all subjects was .45 standard deviations; similarly, the average shift from the LL to the HL was .57 standard deviations. In a control charting context, these shifts are extremely small and their detection will prove difficult, even from the LL to the HL. Control charts optimized to detect shifts as small as .5 standard deviations have an average run length greater than 20 points; indicating that

on average, the control chart will plot 20 epochs of data before shifts this small will be detected. Thus, there will be a delay in detecting changes in OFS that result from the onset of tasks (Montgomery, 2009).

Table 3.1 Size of the shifts in feature means across task loads

	A			E			F		
	Theta	Alpha	Beta	Theta	Alpha	Beta	Theta	Alpha	Beta
LL to ML	0.183	0.845	0.064	0.332	0.307	0.506	0.244	0.551	1.021
LL to HL	0.025	1.631	0.050	0.489	1.107	0.294	0.259	0.471	0.825

3.6.1 Results of Univariate Control Charts on Non-Autocorrelated Features

The following section contains results from univariate control charts which monitored non-autocorrelated psychophysiological features. The features monitored were chosen based on research which demonstrated their responsiveness to change in OFS. For instance, Smith et al. found that theta power measured from frontal-midline locations, e.g. F_z, increased with increasing cognitive load (2001). Similarly, they found that alpha power measured from parietal locations, e.g. P_z, routinely decreased with increasing cognitive load. These and other features responsive to OFS were subsequently analyzed (Gevins et al., 1998; Wilson, Russell, 2003b).

Figure 3.2 is a segment from an EWMA-Shewart chart monitoring a theta feature recorded at the F_z electrode of A01. In this chart, as in all subsequent charts, the vertical lines denote the onset of various task loads: green for the LL, yellow for the ML, and red for the HL. The baseline LL condition always begins each trial and occurs everywhere in between the ML and HL. The EWMA chart, depicted in (a), detected one of the HL tasks, while the Shewart individuals chart detected both, as shown in (b). Notice the delay in

detection; these delays are common throughout the results and their significance was discussed in the introduction of Section 3.6.

Figure 3.2 verifies the assumption that modifying task load induced corresponding changes in cognitive load. This is because theta measured at F_z, the feature plotted on the charts, is known to increase with increasing cognitive load; thus, as theta increased for the tasks, so did cognitive load. Therefore, detecting a change in task load is synonymous with detecting a change in OFS.

Another EWMA-Shewart chart is shown in Figure 3.3, this time displaying an entire trial. This chart monitored a theta feature recorded at the VEOG electrode of E01. The EWMA chart was very effective, detecting all eight HL and ML tasks in (a). In contrast, the Shewart individuals chart was not useful in detecting tasks, as shown in (b), but it was helpful in explaining several alarms in the EWMA chart. For instance, the first out-of-control signal in the Shewart individuals chart was clearly an outlier; this point caused the EWMA chart to be out-of-control for the following six epochs. Knowledge of these outliers can improve the interpretation of EWMA alarms, more specifically, it facilitates the identification of *false* alarms. In general, the Shewart individuals chart was not effective at detecting change in OFS, mainly because it was susceptible to erratic psychophysiological behavior and was sensitive to violations of normality. Instead, it was used to discount false alarms in the EWMA chart; alarms not caused by changes in OFS, but by outliers.

Notice that VEOG theta, shown in Figure 3.3, behaved opposite to F_z theta, shown in Figure 3.2; it decreased with task load in the former and increased in the latter. This contradiction is not uncommon for identical wavebands recorded at different electrodes (Gevins et al., 1998). Therefore, the electrode location is critical when interpreting psychophysiological feature behavior.

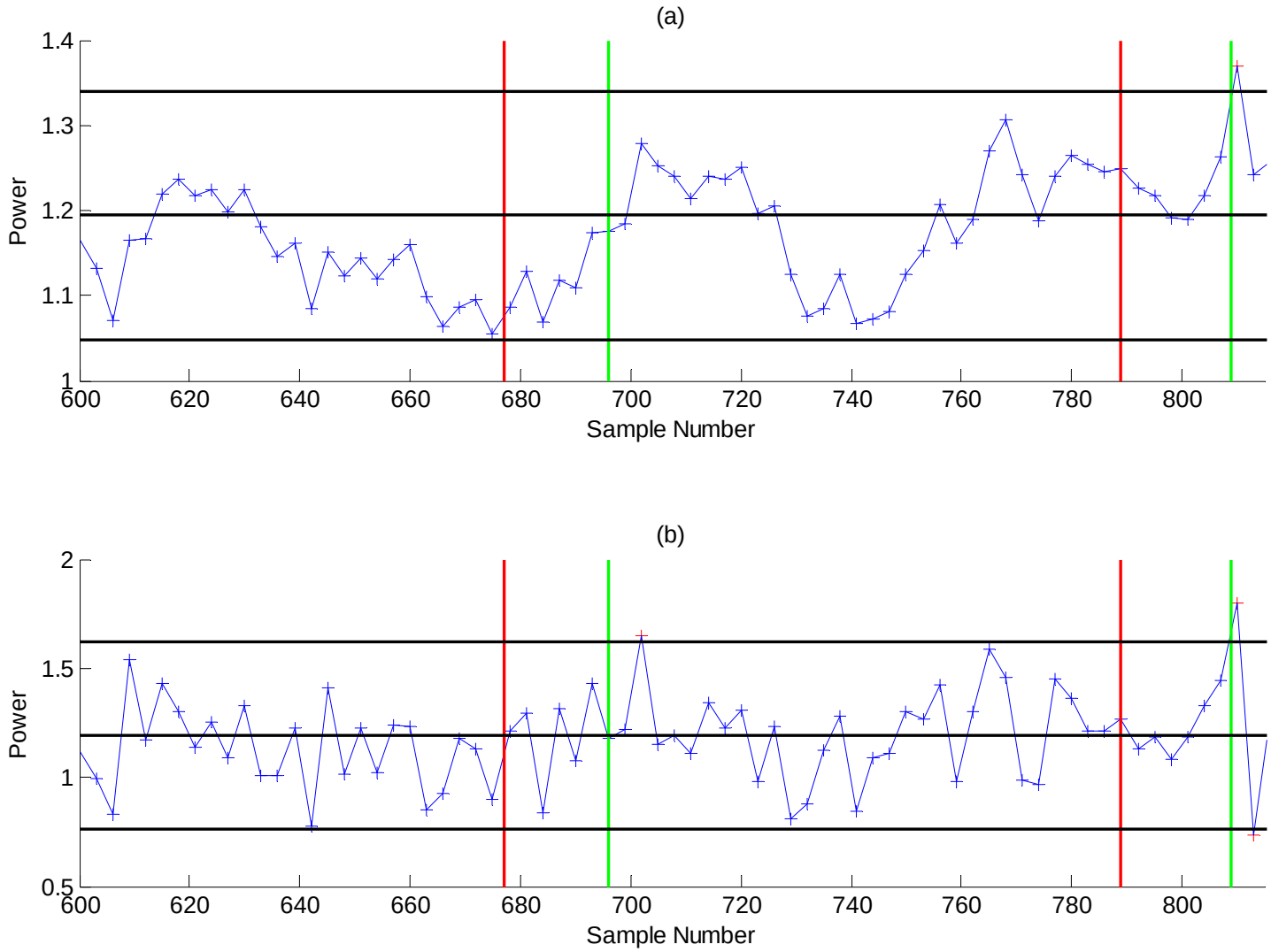

Figure 3.2 Control charts monitoring F_z theta of A01: (a) EWMA, (b) Shewart individuals

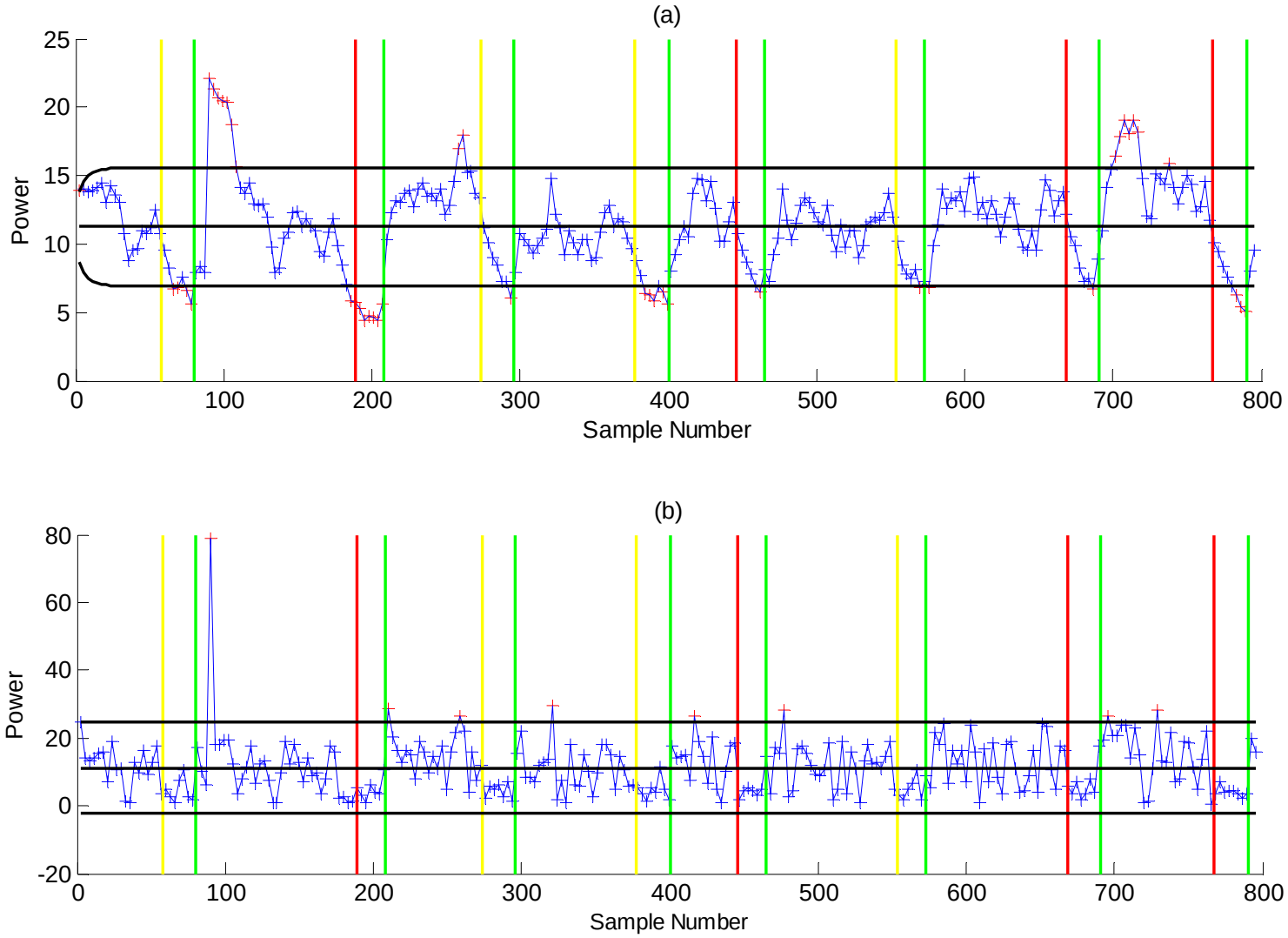

Figure 3.3 Control charts monitoring VEOG theta of E01: (a) EWMA, (b) Shewart individuals

Table 3.2 summarizes the results from the univariate EWMA-Shewart control chart analysis. The table contains the proportion of the 8 tasks which were detected for each feature and trial. Subject A was the easiest to monitor, with over 70% of the tasks detected when their results are averaged across all features. Subject F, on the other hand, proved the most difficult to monitor with only half of their tasks detected on average. The features that performed best across subjects were: VEOG theta and P_z alpha. Overall, there was a feature for each subject that could detect at least 75% of the tasks. Subjects A and E each had one or more features which could detect at least 87.5% of the tasks. Finally, the average number of false alarms for the results reported in Table 3.2 was less than one per trial. Refer to Appendix B for the complete results on false alarms of this and all subsequent control chart analyses.

Table 3.2 Results of non-autocorrelated univariate control charts

	VEOG Theta	**F_z Theta**	**F_z Alpha**	**F_7 Theta**	**0_2 Alpha**	**P_z Alpha**	**Average**
A01	0.875	0.500	0.750	0.750	0.625	0.750	**0.708**
E01	1.000	0.000	0.625	0.625	0.500	0.500	**0.542**
F01	0.375	0.500	0.375	0.500	0.500	0.750	**0.500**
Average	**0.750**	**0.333**	**0.583**	**0.625**	**0.542**	**0.667**	

It should be noted that these results differ in nature from those reported by pattern recognition classifiers. Classification methods have achieved over 80% accuracy on the UAV dataset, where accuracy is defined by correctly classifying OFS as either “high” or “low” for each epoch. By this definition, accuracy is considered strong when it is significantly greater than chance, 50%. In contrast, the goal of control charting is not to

classify OFS for each epoch, but rather, to detect when it has *changed*. In this sense, performance of control charts cannot be compared to chance, since it is possible to detect every change in OFS by simply classifying every epoch as out-of-control. Instead, a control chart is deemed proficient when it does not commit type I and II errors at rates greater than acceptable. In other words, it detects "most" of the changes in OFS and does not exhibit "many" false alarms. Defining what is acceptable depends upon the particular application.

3.6.2 Results of Univariate Control Charts on Autocorrelated Features

This section contains results on univariate control charts that monitored autocorrelated psychophysiological features. Figure 3.4 is an example of an MCEWMA chart monitoring F_7 beta of E01. It is evident that the MCEWMA chart differs greatly from the charts previously analyzed, with its moving centerline and control limits. The beta feature depicted in Figure 3.4 is weakly stationary and strongly autocorrelated; this behavior is ideal for the MCEWMA application. As shown, MCEWMA detected 5 of the 8 tasks, and the results in Table 3.3 indicate that the MCEWMA chart achieved this accuracy for every subject. Neither theta nor alpha features ever exhibited strong enough autocorrelation for the MCEWMA chart to be effective, thus only beta features were monitored.

Table 3.3 summarizes the results of monitoring various beta features in the MCEWMA chart. These features were chosen based on their ability to measure beta power and their responsiveness to changing OFS. More specifically, beta power was found to be strong in the parietal region of the brain, e.g. P_z and T_5, and the F_z feature was weighted heavily in neural network methods (Gevins et al., 1998; Wilson, Russell, 2003b). F_7 beta and P_z beta exhibited the most consistent and accurate performance, both

detecting 62.5% of the tasks for every subject. Lastly, the MCEWMA method resulted in over two false alarms per trial on average.

Table 3.3 Results of MCEWMA control charts

	F_z Beta	**P_z Beta**	**T_5 Beta**	**Average**
A01	0.625	0.625	0.500	**0.583**
E01	0.625	0.625	0.625	**0.625**
F01	0.625	0.625	0.625	**0.625**
Average	**0.625**	**0.625**	**0.583**	

The next results concern control charts which plotted the residuals from time series models. Figure 3.5 displays an EWMA-Shewart chart monitoring the residuals of an AR[1] model fitted to F_z beta of A01. The EWMA chart in (a) correctly detected 6 of the 8 tasks (see UCL), but exhibited excessive noise beyond the LCL.

Table 3.4 summarizes the results of monitoring the time series residuals of various beta features in EWMA-Shewart charts. Notice that these features are identical to those in Table 3.3, except for the inclusion of HEOG beta; this feature exhibited first-order moving average behavior (MA[1]) which was only conducive to time series modeling, not the MCEWMA chart. The results demonstrate that T_5 beta and F_z beta were the most proficient features, detecting nearly 60% of the tasks when averaged across subjects. If the best performing feature is evaluated for each subject, then 62.5% of the tasks were detected on average. In addition, there were 1.8 false alarms per trial on average. Quantitatively, the performance of control charting residuals was marginally better than the MCEWMA chart, however, fitting time series models is a very time consuming and subjective process.

Table 3.4 Results of time series residuals control charts

	HEOG Beta	**F_z Beta**	**P_z Beta**	**T_5 Beta**	**Average**
A01	0.375	0.750	0.625	0.625	**0.594**
E01	0.500	0.500	0.500	0.500	**0.500**
F01	0.375	0.500	0.500	0.625	**0.500**
Average	**0.417**	**0.583**	**0.542**	**0.583**	

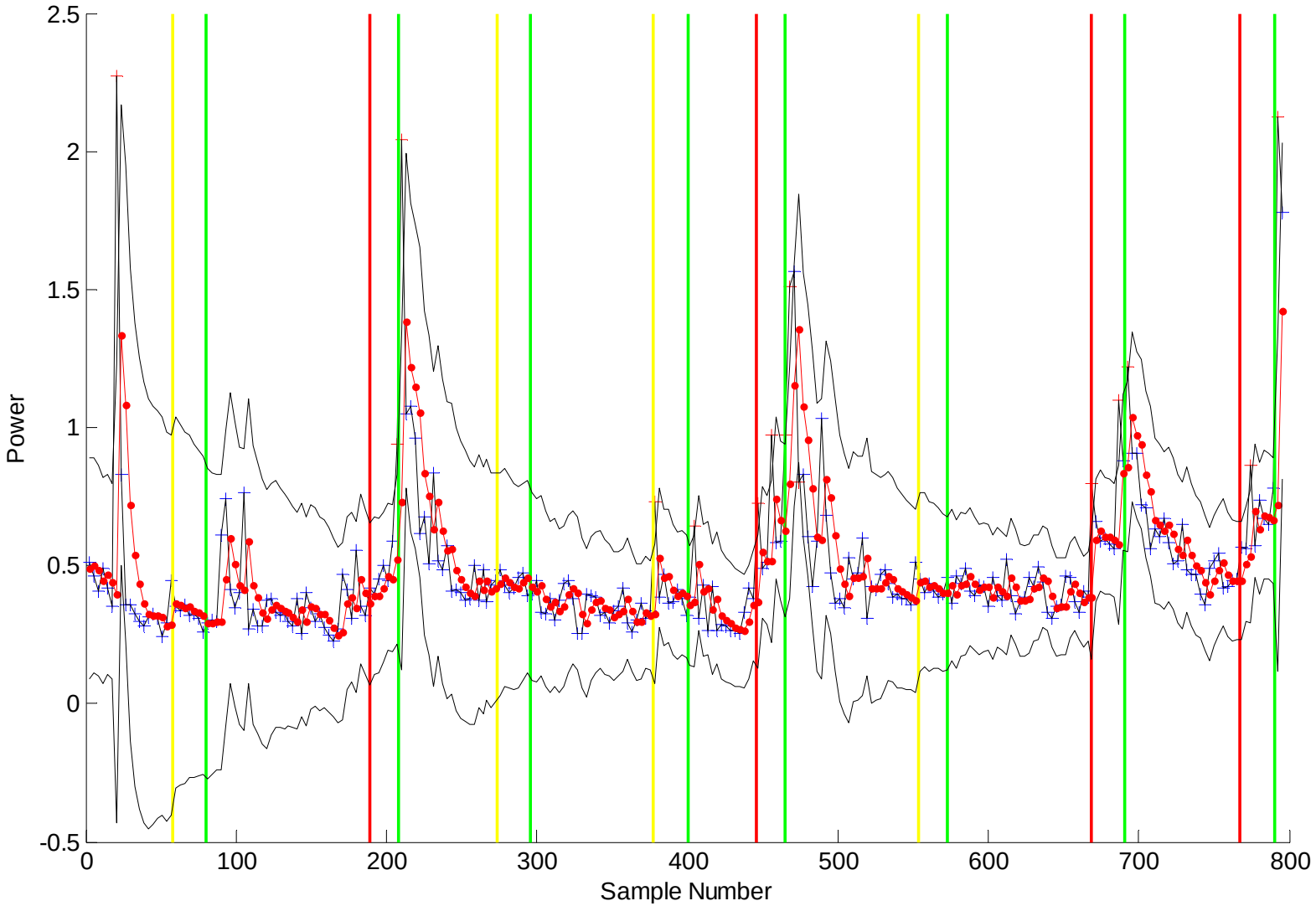

Figure 3.4 MCEWMA chart monitoring F_7 beta of E01

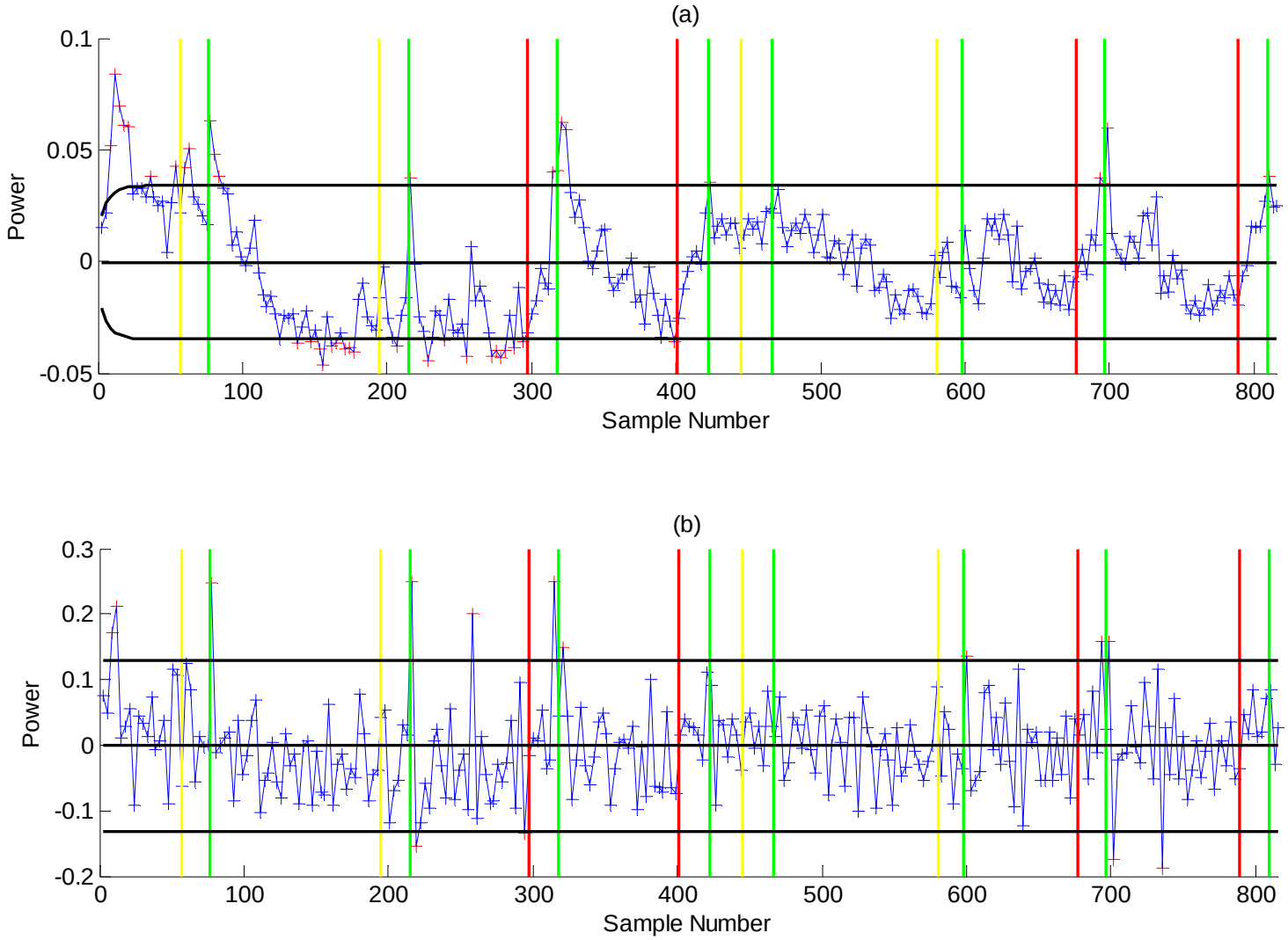

Figure 3.5 Control charts monitoring time series residuals from F_z beta of A01: (a) EWMA, (b) Shewart individuals

3.6.3 Results of Multivariate Control Charts

Multivariate charts were evaluated for their ability to detect real-time changes in OFS. Theoretically, their performance should be superior to univariate charts. While multivariate charts are capable of monitoring dozens of features simultaneously, it was observed that five or less features performed best. Consequently, the following analysis was done on two or three features which performed best for each subject in the univariate analysis (see Table 3.2). These sets of features were monitored in Hotelling-T^2 and MEWMA control charts.

Figure 3.6 is a Hotelling-T^2 chart monitoring alpha and theta recorded at O_2 and F_z, respectively, for A01. It correctly detected 5 of the 8 tasks, with a number of false alarms. Figure 3.7 is an MEWMA chart monitoring VEOG theta, F_7 theta, and P_z alpha of E01. The chart detected 6 of the 8 tasks, with a number of false alarms. Because this chart utilized an exponentially-weighted moving average, the plot was significantly smoother when compared to the Hotelling-T^2 chart in Figure 3.6. The same was true for their univariate analogues: the EWMA charts were smoother than the Shewart individuals charts.

The results for the Hotelling-T^2 and MEWMA control charts on selected features are contained in the first two columns of Table 3.5, respectively. The MEWMA chart performed better on average than the Hotelling-T^2 chart, detecting 62.5% of the tasks compared to 54.2%, respectively. Finally, the MEWMA chart incurred 2.33 false alarms on average and double that rate for the Hotelling-T^2 chart. These false alarm rates were higher than their respective univariate analogues.

Next, the subsets of principle components of the non-autocorrelated features were monitored in Hotelling-T^2 and MEWMA charts for each subject. This evaluation was done for a theoretical improvement over monitoring a few subject-specific features. Figure 3.8 is an example of a Hotelling-T^2 chart monitoring the principle component subset of A01. In this case, eight principle components comprised the subset, and

collectively, they represented the subspace of 17 non-autocorrelated features. This chart detected 5 of the 8 tasks, the same accuracy achieved when a select few features were monitored by the Hotelling-T^2 chart in Figure 3.6.

Table 3.5 Results of multivariate control charts

	Hotelling-T^2 (Select)	**MEWMA (Select)**	**Hotelling-T^2 (PCA)**	**MEWMA (PCA)**	**Average**
A01	0.625	0.750	0.625	0.750	**0.688**
E01	0.500	0.500	0.500	0.500	**0.500**
F01	0.500	0.625	0.375	0.500	**0.500**
Average	**0.542**	**0.625**	**0.500**	**0.583**	

Figure 3.9 is an MEWMA chart monitoring the subset of four principle components representing the subspace of 13 non-autocorrelated features of E01. This chart only detected 4 tasks as opposed to the 6 tasks detected when a select few features were monitored by the MEWMA chart in Figure 3.7.

The results from monitoring a subset of the principle components are contained in the last two columns of Table 3.5. The MEWMA chart performed better on average than the Hotelling-T^2 chart, detecting nearly 60% of the tasks compared to 50%, respectively. The false alarm rates were comparable to monitoring a few subject-specific features. Overall, monitoring a subset of principle components did not yield an improvement over monitoring a few subject-specific features.

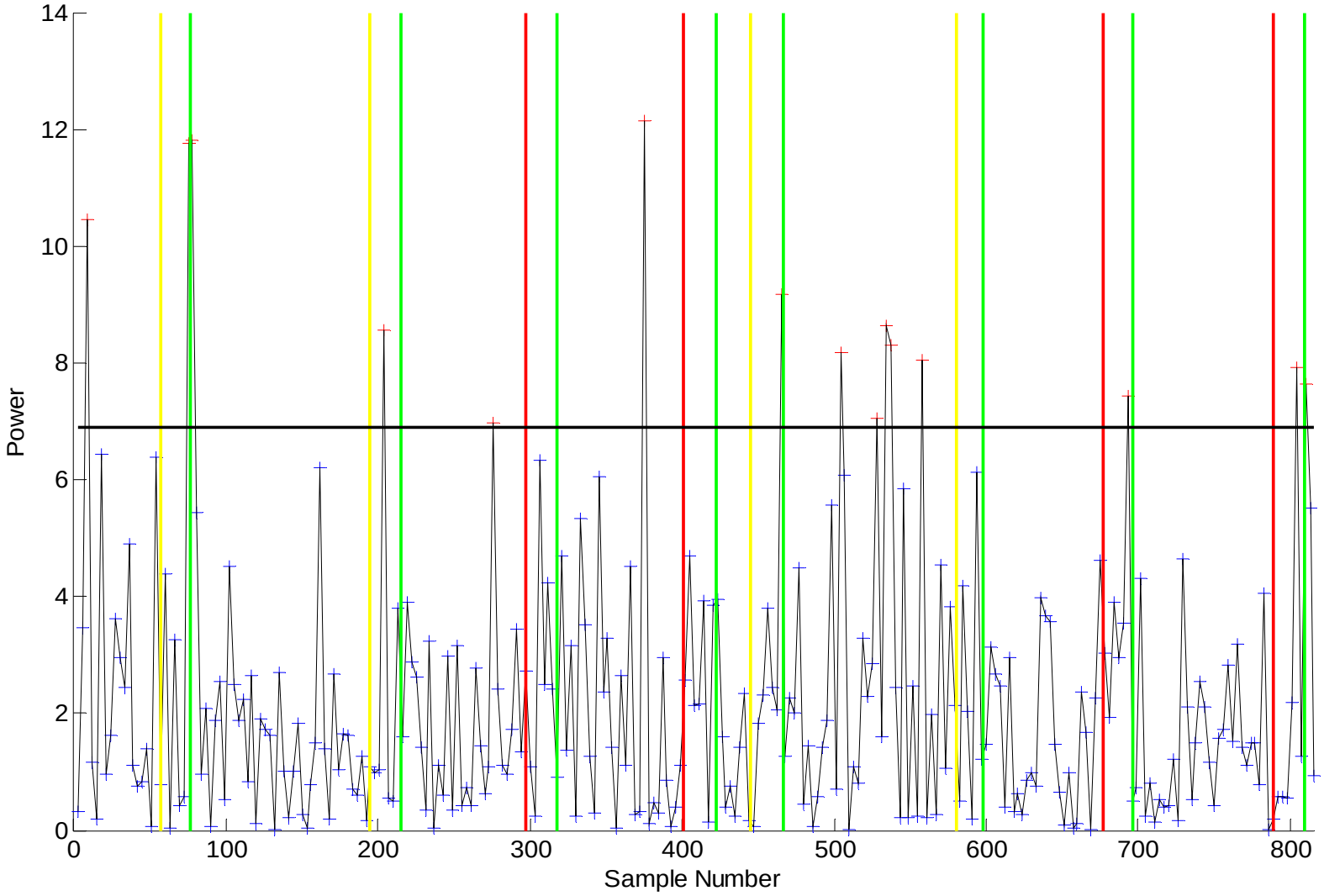

Figure 3.6 Hotelling-T^2 chart monitoring select alpha and theta features of A01

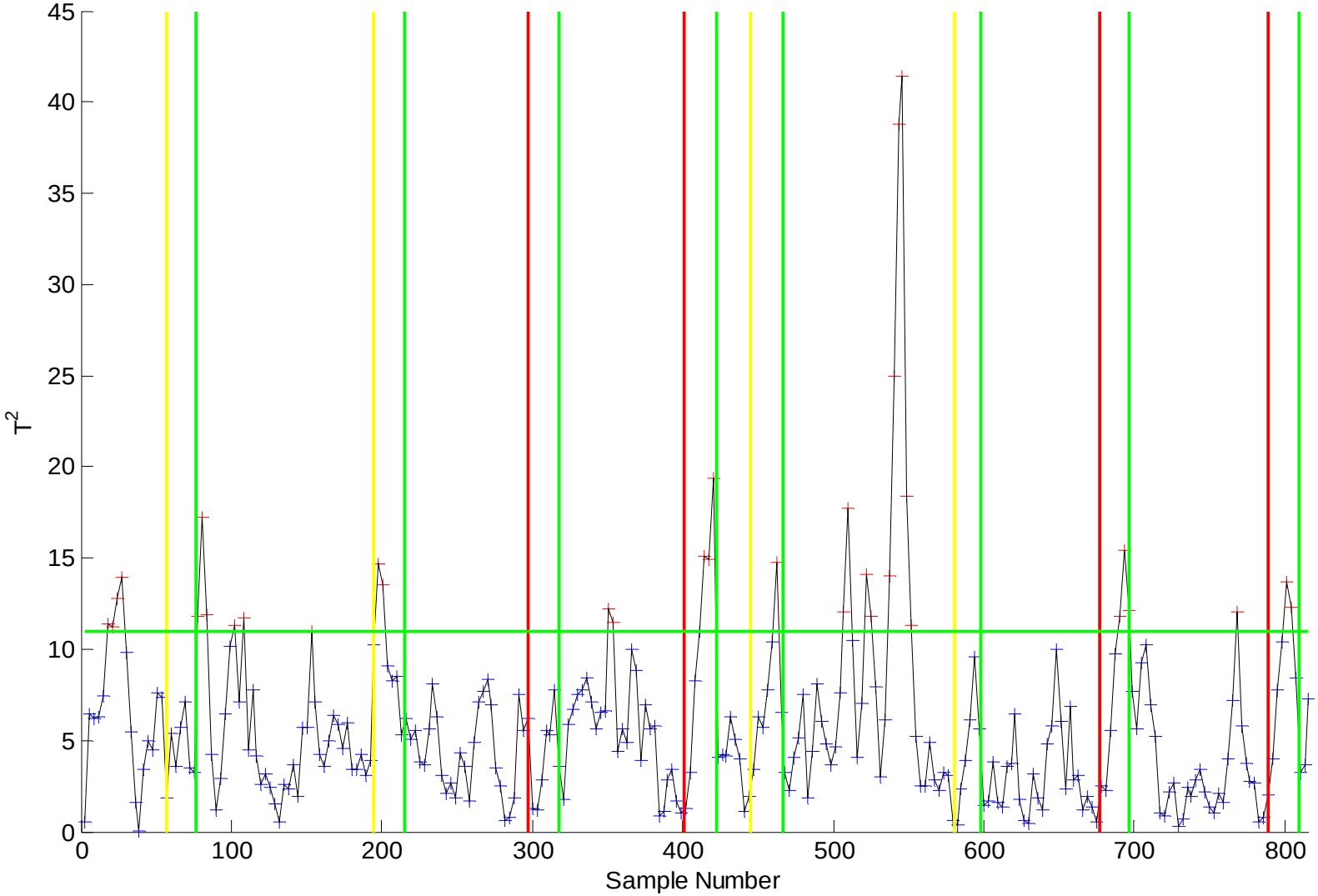

Figure 3.7 MEWMA chart monitoring select alpha and theta features of E01

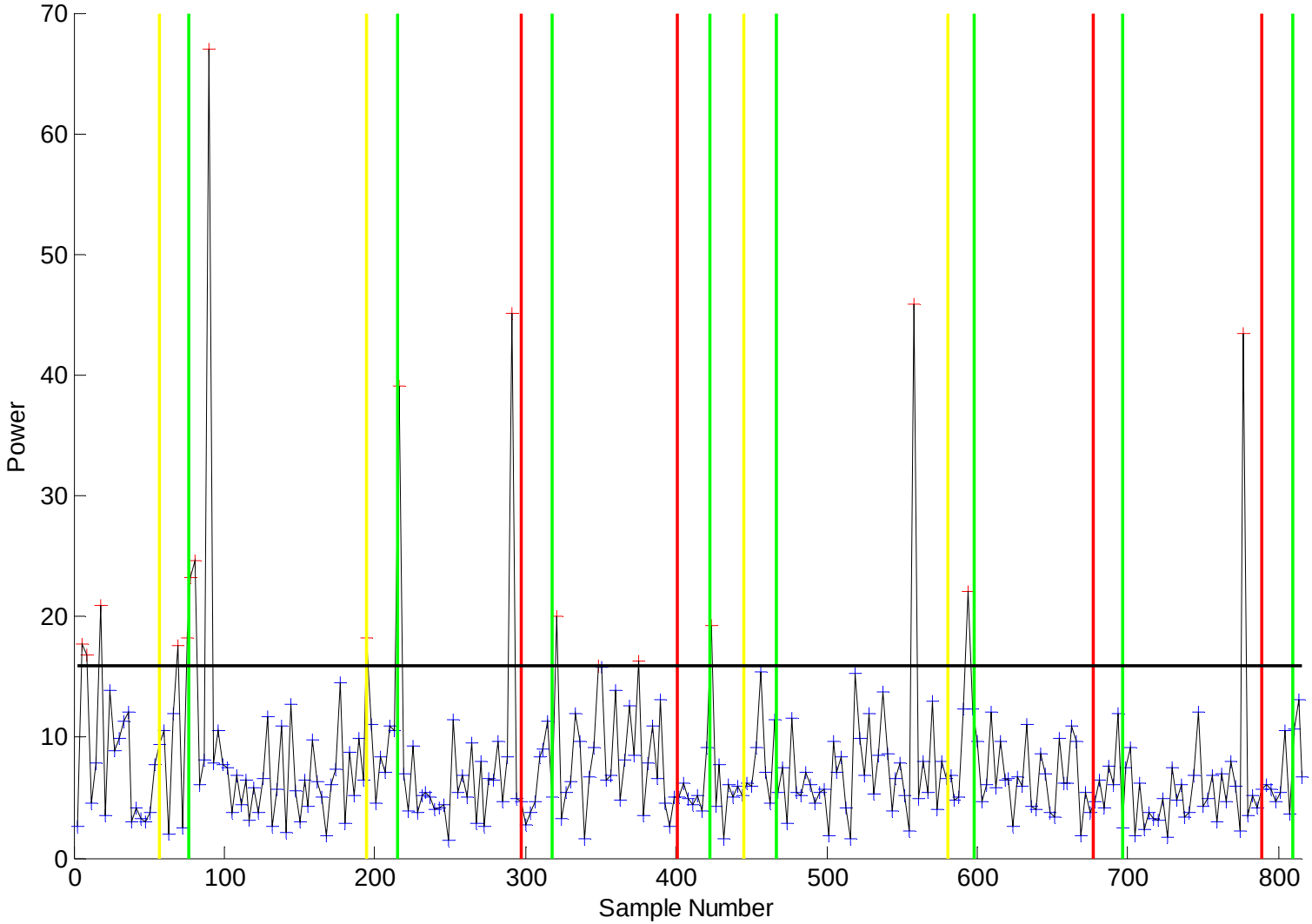

Figure 3.8 Hotelling-T^2 chart monitoring a subset of principle components of A01

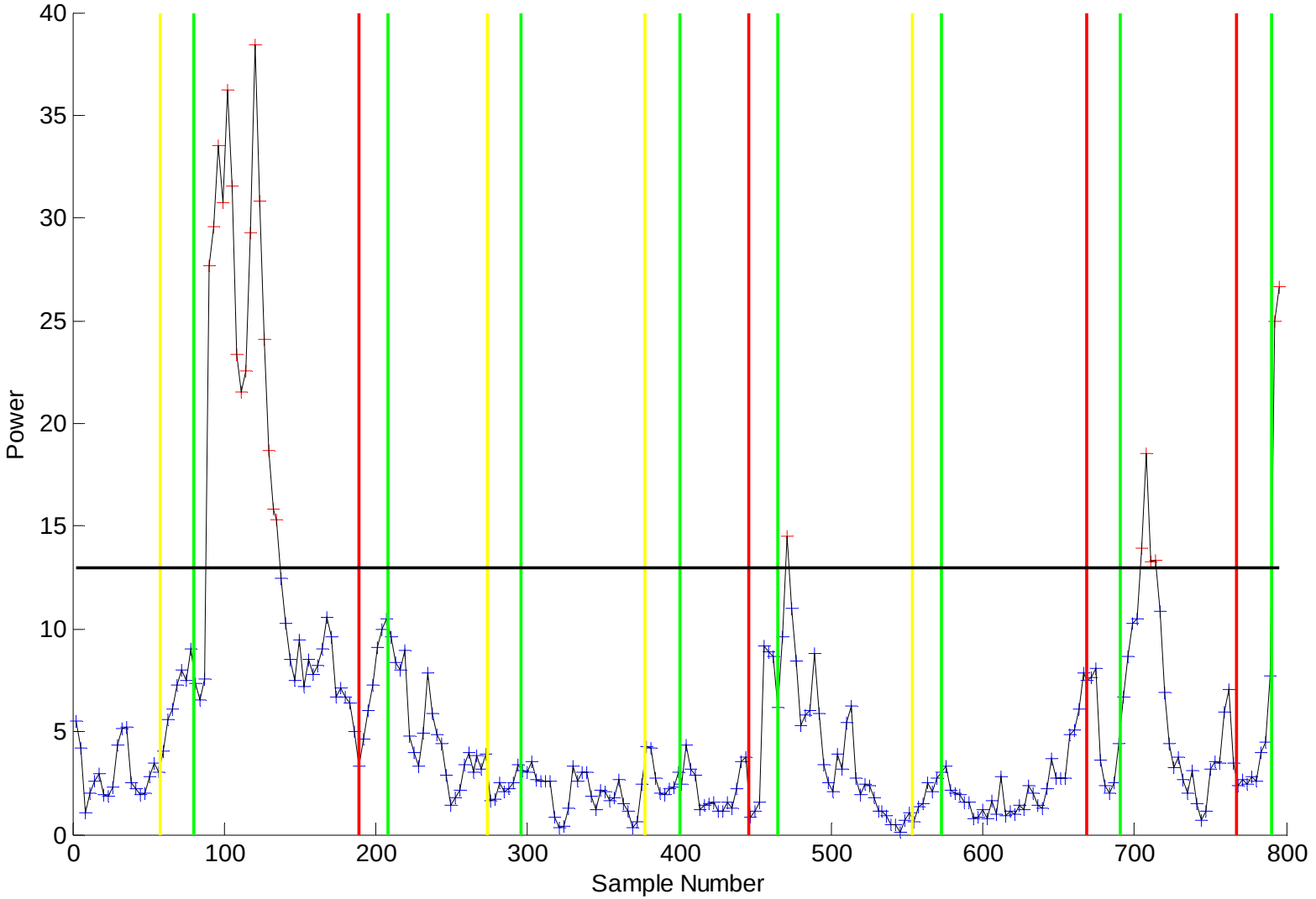

Figure 3.9 MEWMA chart monitoring a subset of principle components of E01

3.6.4 Factor Pattern Analysis on Principle Components

When utilizing principle components, it is important to uncover the relationship between the components and the original features as they often relate to physical characteristics of the underlying process (Scranton et al., 1996). This can be accomplished through *factor pattern analysis* (FPA) (Mastrangelo et al., 1996). Recall that principle components are linear combinations of the original features, e.g., $p_j = u_{1j}x_{1i} + u_{2j}x_{2i} + \cdots + u_{pj}x_{pi}$, where p_j is the *j*th principle component, u_{kj} is the factor loading, and x_{ki} is one of the *p* original features. Thus, by analyzing the factor loadings, one can determine the degree to which the original features contributed to each principle component, and subsequently, reveal patterns that potentially represent physical characteristics. In FPA, the factor loadings are not evaluated directly, rather the correlations between the principle components and their features are evaluated instead; the correlations can be considered factor loadings which have been standardized.

Previously, PCA was only done on features that were non-autocorrelated in order to satisfy multivariate control charting assumptions. The following FPA is conducted on the principle components of *all* the psychophysiological features (35 in total).

Among subjects, the first three principle components accounted for over 60% of the original features' variability, on average. These components represented the most interesting physical phenomena and subsequently, only these components are analyzed further. Table 3.6 is a summary of the FPA done on A01; the columns are the first three principle components and the rows contain the correlations between the components and the original features. For brevity, the features are lumped into groups which had similar correlations, thus for each feature group the reported correlation is an average. The legend for the feature groupings is: ALL for all seven electrodes; O for ocular electrodes including HEOG, VEOG, and F_7; NO for non-ocular electrodes (i.e. the complement of O); FH for the front hemisphere of the brain including HEOG, VEOG, F_7, and F_z; and finally, BH for the back hemisphere of the brain including P_z, T_5, and O_2. The wavebands

of the features are in this order: delta, theta, alpha, beta, and gamma. For instance, in Table 3.6 P1 is strongly correlated with beta and gamma wavebands recorded from all electrodes, and is somewhat correlated with all the remaining features. These results suggest that P1 primarily represented beta and gamma power emitted from subject A's brain. Refer to Appendix C for the complete FPAs on each subject.

Several important inferences can be drawn from the FPA results in Table 3.6. First, subject A's brain activity was dominated by beta and gamma power, since these wavebands dominated P1, and P1 is the most significant principle component. Results like these help interpret out-of-control points when monitoring principle components in control charts. For instance, if a control chart monitoring P1 for A01 went out-of-control, it would likely indicate a change in subject A's higher frequency power. Similarly, P2 and P3 also represented physical characteristic. For instance, the results suggest that P3 was the difference between non-ocular alpha power, and ocular beta and gamma power. Thus, as P3 varies in a control chart, so does the difference between alpha and beta/gamma power. Future research could investigate the implications of these physical characteristics for detecting change in OFS.

Table 3.6 Condensed results from the FPA of A01

P1		**P2**		**P3**	
ALL: Beta + Gamma	0.650	O: All	-0.542	NO: Alpha	0.465
All Else	0.293	NO: Delta thru Alpha	-0.273	O: Beta + Gamma	-0.625
		NO: Beta + Gamma	0.311		

The condensed results of the FPAs on subjects E and F are in tables Table 3.7 and Table 3.8, respectively. Several interesting trends were present across all subjects. First, beta and gamma features consistently exerted the most influence on the primary principle component, P1. This is consistent with past research, where higher frequency wavebands were the most heavily weighted features in PCA and ANN schemes (Wilson, Russell, 1995; Wilson, Russell, 1999; Freeman et al., 1999). Second, the principle components often represented the difference of waveband powers measured from distinct regions of the brain, such as ocular sites vs. non-ocular sites, and front vs. back hemispherical sites. Lastly, features from HEOG and VEOG were consistently among the most correlated with the principle components. Again, this is consistent with past research which has demonstrated the strong relationship between eye activity and changes in OFS (Wilson, Russell, 2003b; Yu, 2009).

Table 3.7 Condensed results from the FPA of E01

P1		**P2**		**P3**	
ALL: Beta + Gamma	-0.823	FH: Delta thru Alpha	-0.470	HEOG: All Bands	0.438
All Else	-0.546	NO: Beta + Gamma	0.273	NO: Theta + Alpha	-0.356

Table 3.8 Condensed results from the FPA of F01

P1		**P2**		**P3**	
ALL: Beta + Gamma	0.764	O: All	0.338	VEOG: All	-0.369
All Else	0.560	NO: Delta	0.249	BH: Delta thru Alpha	0.482
		NO: Beta + Gamma	-0.403		

3.7 Discussion of Control Charts

The results demonstrate that control charts can detect real-time change in OFS with varying degrees of accuracy. The most accurate control charting method was to monitor a single subject-specific feature in a univariate EWMA-Shewart chart. Averaged across all subjects, this method detected 75% of the changes in OFS. In general, control charts displayed several advantages. Primarily, they are simple, only requiring a rudimentary knowledge of statistics; they are online, capable of detecting real-time change; they can be tailored to a specific subject; and finally, they are visual in nature and readily interpretable.

The optimal method, the EWMA-Shewart chart on a subject-specific psychophysiological feature, has several drawbacks. Primarily, this method cannot accommodate autocorrelated features, such as beta and gamma wavebands, and it is unable to monitor multiple features simultaneously. The first drawback can be overcome by control charts such as the MCEWMA. But, despite MCEWMA's superiority over other autocorrelated methods, it only detected 62.5% of the changes in OFS on average.

The evaluation of multivariate control charts investigated their benefits over univariate control charts. Two different approaches were analyzed: monitoring a few subject-specific features and monitoring a subset of the principle components from all non-autocorrelated features. The latter approach was expected to perform the best among all control charts, univariate and multivariate alike, but this was not the case. The best multivariate control chart approach was to monitor several subject-specific features in an MEWMA chart. The Hotelling-T^2 chart monitoring the same subject-specific features was less accurate and had twice the false alarms. When averaged across subjects, the optimal multivariate control charting method detected 62.5% of the changes in OFS, much less than the 75% accuracy of the optimal univariate method.

The control charts which performed the best shared several common characteristics. First, they smoothed the features, in this case, via an EWMA. It is well

known that smoothing time series data can help separate the signal from the noise, reveal important trends, and render non-normal data more normal (Montgomery, 2008). With respect to psychophysiological features, there is an abundance of noise which creates a highly variable and often difficult signal to interpret. Consequently, smoothing filters are common in psychophysiological research, most notably, the simple moving average with overlap and the Hanning window (Freeman et al., 1999; Wilson, Russell, 2007). A second characteristic common among the best performing control charts was that they monitored subject-specific features. Each subject responded differently to varying task load, thus no feature or subset of features were optimal with respect to all subjects. This result is substantiated in research, where developing generalizable OFS classification methods has proven difficult.

Contrary to expectations, monitoring the principle components was not the optimal control charting method. In the present analysis, principle components were generated from the non-autocorrelated features. This was done to prevent the principle components from being autocorrelated, and thus, incapable of being monitored in traditional multivariate control charts. This constraint, however, routinely excluded beta and gamma features which were consistently autocorrelated. Recall that the FPA results revealed that these higher frequencies were the most prominent features in a subject's brain activity; hence, excluding these features may have been detrimental to the control chart's performance.

Another unexpected result was that multivariate control charts monitoring either a few subject-specific features or a subset of principle components, were inferior to the best univariate control chart. In the multivariate charts, psychophysiological features were simultaneously monitored for a theoretical improvement in detection accuracy. However, combining the features actually resulted in *decreased* accuracy. As mentioned, psychophysiological data is highly variable, and perhaps combining their features via the T^2 statistic resulted in an unpredictable, ambiguous signal. An alternate to the T^2 statistic

may perform better, in particular, one that doesn't oversimplify the higher order, non-linear behavior of psychophysiological signals.

Finally, several control chart deficiencies need to be addressed for them to reliably detect real-time changes in OFS. First, they need to be made more sensitive to changes in the mean. Psychophysiological features changed from one OFS to another, but their change was too subtle for existing control charts. As such, the onset of HL and ML tasks were detected over 10 epochs late on average, i.e., a 30 second delay. This delay is far too long for adaptive aiding in operational settings. Second, the control charts exhibited false alarms at unacceptable rates. In an operational environment, it is imperative to withhold adaptive aiding unless it is absolutely necessary, otherwise, it will serve to actually distract the operator. Third, control charts do not indicate to what degree a subject is in a low OFS; it is only classified as "out-of-control". Lastly, current control charts are simply too rigid to fully capture behavior common to psychophysiological features. They rely on traditional statistical methods which are only valid when a host of assumptions are satisfied, most of which were violated by the features. Moreover, control charts are linear methods, in other words, they neglect the non-linear interactions and higher order nature of psychophysiological signals. Wilson and Russell purport that non-linear methods, such as ANN, "may be advantageous in complex task situations that involve multiple cognitive components, especially if multiple physiological features are used" (2003a). If this is true, then linear-based control charts will be sub-optimal.

CHAPTER 4
INDEPENDENT COMPONENT ANALYSIS OF PSYCHOPHYSIOLOGICAL SIGNALS

4.1 Introduction to Independent Component Analysis

Independent component analysis (ICA) is a method used to extract the underlying components of signals. ICA assumes that physical processes, like the brain, are comprised of distinct operators which emit signals independent of each other. When these signals are recorded by sensors, they become "mixed" and indistinguishable. ICA is a method for separating these mixed signals into their underlying components (Stone, 2002).

For instance, two EEG electrodes record brain activity at different locations on the scalp. The signals recorded from these electrodes, denoted as $x_1(i)$ and $x_2(i)$, are assumed to be a mixture of components, $s_1(i)$ and $s_2(i)$, originating from independent operators in the brain. $x_1(i)$ and $x_2(i)$ are expressed as linear combinations of the source components,

$$\begin{aligned} x_1(i) &= a_{11}s_1(i) + a_{21}s_2(i) \\ x_2(i) &= a_{12}s_1(i) + a_{22}s_2(i) \end{aligned} \tag{4.1}$$

The objective of ICA is to reveal the source components using only the recorded EEG signals, $x_1(i)$ and $x_2(i)$. As shown in (4.1), this requires determining the coefficients a_{kj}. This is a difficult task, however, it is made feasible by assuming the source components are independent of each other, hence "independent components" (Hyvärinen, Oja, 2000).

ICA is more generally defined by considering the system of equations in (4.1) in vector-matrix notation,

$$\boldsymbol{x}(i) = \boldsymbol{A}\boldsymbol{s}(i) \tag{4.2}$$

where $\boldsymbol{A}$ is the mixing matrix, the recorded psychophysiological signals are in vector $\boldsymbol{x}(i)$, and the independent components are in vector $\boldsymbol{s}(i)$, for each epoch i. Once $\boldsymbol{A}$ is determined, its inverse, $\boldsymbol{W}$, is used to compute the independent components for every realization, $\boldsymbol{x}(i)$, by

$$\boldsymbol{s}(i) = \boldsymbol{W}\boldsymbol{x}(i) \tag{4.3}$$

Notice that $\boldsymbol{x}(i)$ is what is observed, but $\boldsymbol{s}(i)$ is what needs to be computed. Thus, similar to PCA, the independent components are determined indirectly and called latent variables (Hyvärinen, Oja, 2000). Moreover, ICA is the most common means of blind source separation (BSS), where underlying factors are determined "blindly". An assumption of BSS, which ICA necessarily makes, is that the number of independent components is equal to the number of recorded signals. For instance, in the present data there were seven EEG and EOG signals recorded, thus, ICA assumes each were a mixture of seven independent components (Stone, 2002).

Some fundamental assumptions of ICA have been mentioned, but more assumptions are necessary to make ICA feasible. In order to achieve components which are independent, it is necessary to assume the components are non-Gaussian distributed (i.e. non-normal). In Section 4.3, independent components are established by constructing a demixing matrix which maximizes non-Gaussianity. Next, the source signals are assumed to propagate through a medium, in this case brain tissue, instantaneously, before being *linearly* mixed at the electrodes. Finally, the source signals are assumed to be stationary.

ICA is made robust to moderate violations of these assumptions through the FastICA algorithm, detailed in Section 4.3. The major assumption which cannot be compromised, however, is the number of source signals assumed to exist. For the present purpose, this number is less important than finding components which are meaningful to OFS change detection (Makeig et al., 1996; Vigário et al., 2000).

Despite their similarities, ICA differs fundamentally from PCA. They are both multivariate analyses, whose latent variables are linear combinations of observed variables. However, the primary goal of PCA is dimensionality reduction, whereas ICA's is to create a generative model to reveal the underlying factors of a process. Additionally, PCA merely de-correlates its latent variables. Under the assumption of normality, de-correlation is sufficient to achieve independence. However, psychophysiological signals have consistently displayed non-normal behavior, implying that PCA's de-correlation does *not* result in truly independent components. In contrast, ICA imposes stricter constraints which achieve un-rotatable, de-correlated, and *independent* components (Stone, 2002).

4.2 Pre-Processing the Data for Independent Component Analysis

Before conducting ICA, the data must be pre-processed. The data are made to be zero mean, resulting in independent components which are also zero mean; this is done to facilitate the estimation of $\boldsymbol{A}$. After it is estimated, the mean is added back to the independent components.

Next, the data are whitened to make $\boldsymbol{A}$ orthogonal which significantly reduces the number of parameters ICA must estimate. Whitening linearly transforms $\boldsymbol{x}(i)$ into a new vector, $\tilde{\boldsymbol{x}}(i)$, that is uncorrelated and has unit variance. Whitening can be done through eigenvalue decomposition (EVD), the mechanics of which were covered in Section 3.5.3. Using the same notation, $\tilde{\boldsymbol{x}}(i)$ is computed by

$$\tilde{\boldsymbol{x}}(i) = \boldsymbol{U}\boldsymbol{L}^{-1/2}\boldsymbol{U}'\boldsymbol{x}(i) \tag{4.4}$$

This step can also be used to dimensionally reduce the independent components, which serves to decrease noise and prevent overfitting (Hyvärinen, Oja, 2000).

4.3 The FastICA Algorithm

Once the data are pre-processed, an efficient algorithm called FastICA is used to estimate the demixing matrix, $\boldsymbol{W}$. FastICA determines the columns of $\boldsymbol{W}$, denoted as $\boldsymbol{w}_j$, one-by-one, by maximizing the non-Gaussiantity of the projection, $\boldsymbol{w}_j{}'\boldsymbol{x}(i)$. There are several measures of non-Gaussianity, among the most common is negentropy. Negentropy is an entropic metric that captures the "randomness" of a variable. Because Gaussian variables are the most random, they have the highest entropy among all random variables. Therefore, FastICA indirectly maximizes non-Gaussianity by pursuing the least entropic variables as measured by negentropy. As mentioned, the non-Gaussianity of components is essential to achieve their independence. An approximation of negentropy, $J(y)$, is computed by

$$J(y) = (E[G(y)] - E[G(v)])^2 \tag{4.5}$$

where G is some non-quadratic function, y and v are both zero mean and unit variance variables, and v is Gaussian distributed. In this metric, if y is also Gaussian then negentropy is zero; any other distribution of y results in positive negentropy, with magnitude proportional to y's deviation from Gaussianity. In the present research, the non-quadratic function was specified as, $G(y) = y^4$, rendering (4.5) a kurtosis-based approximation of negentropy.

The FastICA algorithm is presented in Table 4.1. Steps 2.3 through 2.5 iteratively refine a randomly chosen vector, $\boldsymbol{w}_j$, until it converges. Convergence occurs when a new $\boldsymbol{w}_j$ points in the same direction as the vector from the previous cycle of steps 2.3 and 2.4. Once convergence occurs, step 2.8 employs a deflation procedure to prevent different the $\boldsymbol{w}_j$ vectors from reaching the same maximum; deflation de-correlates the outputs of $\boldsymbol{w}_1'\boldsymbol{x}(i), \boldsymbol{w}_2'\boldsymbol{x}(i), \ldots, \boldsymbol{w}_p{}'\boldsymbol{x}(i)$. Finally, step 2.9 renormalizes $\boldsymbol{w}_j$ after deflation. The algorithm iterates until $\boldsymbol{W}$ is fully defined.

Table 4.1 FastICA algorithm

1. **Initialize:**
   1.1. $j = 1$
   1.2. p is the number of independent components

2. **Run Algorithm:**
   2.1. **for** $j < p$
   2.2. Choose a random vector $\boldsymbol{w}_j$
   2.3.
   $$\boldsymbol{w}_j{}^{+} = E\left[\boldsymbol{x}(i)G'\left(\boldsymbol{w}_j{}'\boldsymbol{x}(i)\right)\right] - E\left[G''\left(\boldsymbol{w}_j{}'\boldsymbol{x}(i)\right)\right]\boldsymbol{w}_j$$
   Note: expectations are estimated by sample means.
   2.4.
   $$\boldsymbol{w}_j = \frac{\boldsymbol{w}_j{}^{+}}{\left\|\boldsymbol{w}_j{}^{+}\right\|}$$
   2.5. **if** not converged, go back to 2.3
   2.6. **else if** $j == 1$ go back to 2.1
   2.7. **end if/else**
   2.8.
   $$\boldsymbol{w}_j = \boldsymbol{w}_j - \sum_{k=1}^{j-1} \boldsymbol{w}_j{}'\boldsymbol{w}_k\boldsymbol{w}_k$$
   2.9.
   $$\boldsymbol{w}_j = \frac{\boldsymbol{w}_j}{\sqrt{\boldsymbol{w}_j{}'\boldsymbol{w}_j}}$$
   2.10. $j = j + 1$
   2.11. **end for**

3. **Output:** the demixing matrix, $\boldsymbol{W}$

4.4 Independent Component Analysis to Facilitate Real-Time OFS Change Detection

In past research, ICA has been conducted on physiological signals for two purposes: to identify and extract artifacts, such as eye blinks, and to facilitate the analysis of event-related potentials (ERP). There are many artifacts in psychophysiological signals, such as neck muscle activity, eye blinks, heart rate, and line noise originating from recording equipment. In some cases, the artifacts' amplitude exceeds the brain activity, the very thing intended to be measured. Therefore, it is critical to identify and extract these artifacts before using psychophysiological features to detect change in OFS. However, the methods which eliminate artifacts in real-time, such as the adaptive filtering algorithm in Section 2.2, are computationally expensive and are only tailored to eliminate one artifact source. ICA, on the other hand, automatically extracts most of the prominent artifacts, producing less complex and less noisy source signals (Vigário et al., 2000).

ICA is also commonly used to analyze ERPs in EEG signals. ERPs are spikes in brain activity that result from the onset of a stimulating event, like hearing a "beep". It is believed that ERPs originate from different sensory systems in the brain, corresponding to the sense being stimulated, like sight, sound, and touch. To identify the location of these sensory systems, experiments are conducted which simultaneously stimulate multiple senses of a subject while recording their EEG. The ERPs are contained in the EEG, but are mixed upon recording. ICA is then applied to "de-mix" these signals and generate independent components corresponding to the ERPs which originated from their respective sensory systems (Vigário et al., 2000; Makeig et al., 1996).

To the best of our knowledge, ICA has yet to be employed to facilitate the online change detection of OFS. Nevertheless, ICA is a promising candidate for this application since it has been effective at artifact extraction and generating independent components

that correspond to external stimuli. The objective is to use ICA to extract source signals which are free of artifacts and responsive to varying task load.

4.5 Analyzing the Independent Components

For each subject, all seven EEG and EOG signals were submitted to ICA and the resulting independent components were analyzed. In every case, at least two independent components correlated directly with eye activity measured at VEOG and HEOG; this is consistent with the visual demands of the UAV tasks. This phenomenon repeated itself when only EEG signals were submitted to the ICA, hence providing strong evidence that eye activity contaminated the EEG data. Figure 4.1 depicts two independent components extracted from the five EEG signals of E01; it clearly displays their correlation to the VEOG and HEOG signals that were not included in the analysis.

The independent components that were highly correlated with VEOG and HEOG signals were deemed "artifact" signals and were excluded from further analysis. The remaining components were empirically analyzed to determine what brain process they represented. This was accomplished by constructing a frequency spectrum for each component, to reveal the frequencies that dominated them. Again, the DFT, as shown in (2.3), was used to convert the components from the time domain to the frequency domain.

Figure 4.2 depicts the frequency spectrums of four independent components from A01. Each component was partitioned in HL, ML, and LL sets and submitted to the DFT in three-second epochs. The results from the DFT were averaged across all epochs to form three frequency spectrums, one for each task load. As shown, each independent component exhibited unique frequency spectrums, some of which varied by task load. For instance, plot (c) displays the strong influence 10-12 Hz waves had on the third independent component. More importantly, the power of these waves varied by task load.

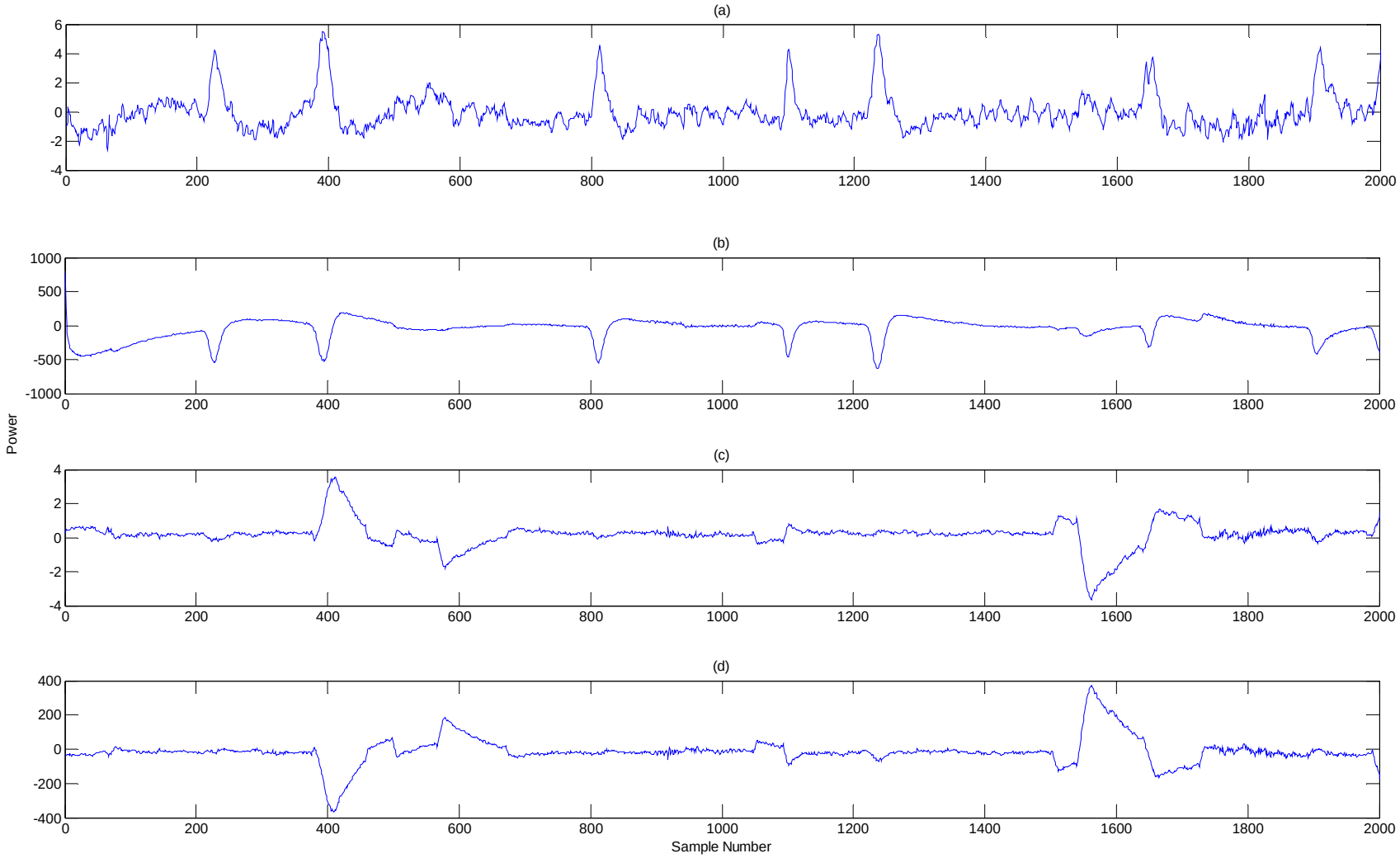

Figure 4.1 Independent components vs. EOG signals of E01: (a) IC1, (b) VEOG, (c) IC2, (d) HEOG

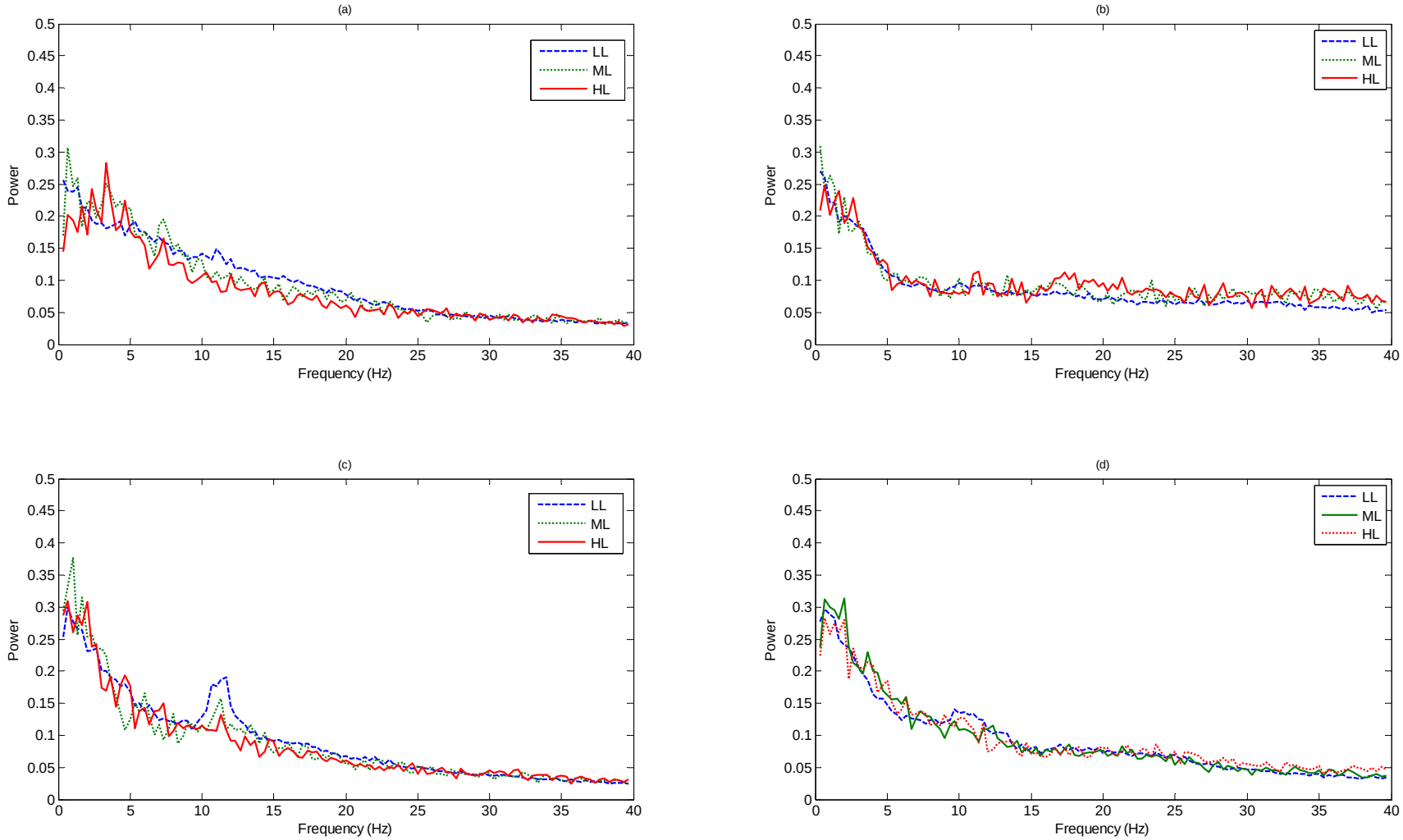

Figure 4.2 Frequency spectrum of independent components of A01: (a) IC1, (b) IC2, (c) IC3, (d) IC4

In fact, every subject had at least one component that was strongly influenced by 8-12 Hz waves, corresponding to the alpha waveband. Recall that this waveband correlates to relaxation; as a subject becomes more relaxed, their alpha power increases. This is consistent with plot (c) of Figure 4.2, where alpha power is highest in the LL and lowest in the HL.

Finding alpha's presence in the psychophysiological signals of cognitive tasks is not surprising, but revealing its prominence with such clarity is rare. Alpha power is frequently obscured by noise when signals are mixed and recorded at the scalp. ICA segregated these noisy signals into source signals that clearly displayed alpha's contribution. Makeig et al. confirmed this result, reporting that ICA revealed "alpha activity (near 10 Hz) not obvious in the EEG data" (1996). This highlights another result, that alpha power most often peaked around 10 Hz in the independent components. This particular frequency's significance has been established in other research, where it has been shown to correlate with performance on cognitive tasks (Makeig, Inlow, 1993). Research has also shown 10 Hz to be associated with spatial tasks, more so than verbal ones (Gevins et al., 1997). This result is consistent with the experiment, as the UAV tasks were largely spatial in nature.

Analyzing the frequency spectrum of a subject's independent components may prove useful in detecting changes of OFS. An abundance of research has employed the frequency domain of EEG signals for classification, most commonly to analyze the traditional wavebands of delta, theta, alpha, beta and gamma. Perhaps a more effective strategy would be to use independent components to derive subject-specific and task-specific wavebands. An analysis similar to the one depicted in Figure 4.2 could be done for each subject to identify frequency bands which are sensitive to changes in task load. Wilson and Fisher report using subject-specific wavebands derived from latent variables to increase classification accuracy by 29% over traditional wavebands (1995). In another

case, estimating the error rates on task performance was improved significantly by using a subject's entire EEG spectrum over predefined, narrow wavebands (Jung et al., 1997).

However, one shortcoming of this method is the subjective process of selecting frequency bands from the independent components. It requires expert knowledge to identify which components represent artifacts, to understand how brain activity responds to different tasks, and finally, to decide which frequency bands are sensitive to task load. These ambiguities can be mitigated through dimensionality reduction, where multiple psychophysiological signals are reduced to result in a pair of independent components. From this reduction, an independent component may emerge with properties which are easily identifiable and robust across subjects. The next section discusses this approach.

4.6 Dimensionally Reduced Independent Components

During the pre-processing for ICA, EVD was used to whiten the data. EVD is equivalent to PCA, and like PCA, it can dimensionally reduce the psychophysiological signals to a subset of important components. Dimensionality reduction prevents overfitting of the independent components and decreases noise, not to mention, reduces the vast number of features into a small, yet crucial subspace. Most importantly, dimensionality reduction reveals the independent components responsive to task load with less ambiguity.

EVD was used to reduce the seven EEG and EOG signals to a pair of signals for submission to ICA. Reducing the signals from seven to two was justified by a Pareto chart analysis which, for every subject, displayed a clear break at two when plotting the eigenvalues in descending order. The resulting pair of signals was submitted to ICA which generated two independent components. These components were then plotted and empirically assessed for any correlation they had with task load.

This method yielded consistent results when conducted on each trial. In every case, one independent component exhibited clear variation with changing task load, while

the other was mostly noise. Figure 4.3 displays the results of A01, where (a) is the component responsive to task load and (b) is the noise component. Notice that the independent component in (a) became attenuated during the HL and ML, compared to its variation in the LL. The second component in (b) exhibited similar behavior, but was not consistent for every task and was obscured by noise. The utility of the first component for real-time OFS change detection is promising since: one exists for each subject, its behavior is distinguishable by task load, and finally, the ICA method is automatic and leaves little error-prone subjectivity.

Further investigation of the dimensionally reduced independent components disclosed strong correlations with the VEOG and HEOG signals. Previously, the independent components correlated with VEOG and HEOG were excluded from further analysis since they were deemed "artifacts". However, in the present analysis these components were extracted as source signals for all the EEG\EOG data, hence they were retained.

Table 4.2 and Table 4.3 contain the correlations between the original psychophysiological signals and the first and second independent components, respectively. The correlations were averaged across trails by subject. Subject F is considered separately since, when his components were plotted, they exhibited unique behavior. Subject F has consistently been the most difficult to accommodate in a variety of methods, and the correlation analysis provided clues as to why.

Table 4.2 Correlation analysis of dimensionally reduced first independent component

	VEOG	**HEOG**	**F_z**	**F_7**	**P_z**	**T_5**	**O_2**
Subjects A, E	0.903	0.107	0.835	0.385	0.389	0.297	0.179
Subject F	0.977	0.075	0.380	0.137	0.261	0.228	0.174

Table 4.3 Correlation analysis of dimensionally reduced second independent component

	VEOG	**HEOG**	**F_z**	**F_7**	**P_z**	**T_5**	**O_2**
Subjects A, E	0.373	0.992	0.067	0.860	0.281	0.036	0.475
Subject F	0.213	0.991	0.104	0.870	0.045	0.159	0.227

In Table 4.2, notice the strong correlation the first independent component had with the VEOG signal across subjects. Another notable correlation for subjects A and E, is with the EEG signal recorded at F_z. However, this is not true for subject F, who averaged less than half the correlation at F_z than subjects A and E.

The second independent component, in Table 4.3, correlated with one EOG and one EEG signal, similar to the first component. But in this case, the correlation was with HEOG and F_7, respectively. Notice that subject F's behavior was not appreciably different than the other subjects.

This analysis elucidates several important findings, principally, that the source signals of the EEG\EOG were strongly correlated to eye activity. This is useful knowledge, as otherwise, VEOG and HEOG contaminated signals may be discarded as artifacts rather than treated as important. As mentioned, one of the two independent components always varied with task load and the other was mostly noise. For subjects A and E, the independent component which best varied with task load was the first. The correlations in Table 4.2 indicate that this component was a mixture of signals from VEOG and F_z. Recall that the F_z electrode mostly records theta waves which are sensitive to changes in cognitive load. This was true except for subject F, who did not have a strong F_z presence in the first component. In fact, subject F's first component was noisy and it was his second component, the one dominated by HEOG and F_7, that best varied with task load. Two things can be inferred from this: first, ICA can reveal the source of

between-subject differences, and second, those differences can be overcome with at least one independent component that is sensitive to task load. Henceforth, the independent component that best varied with task load, post-dimensionality reduction, is referred to as the TVIC (task-varying independent component).

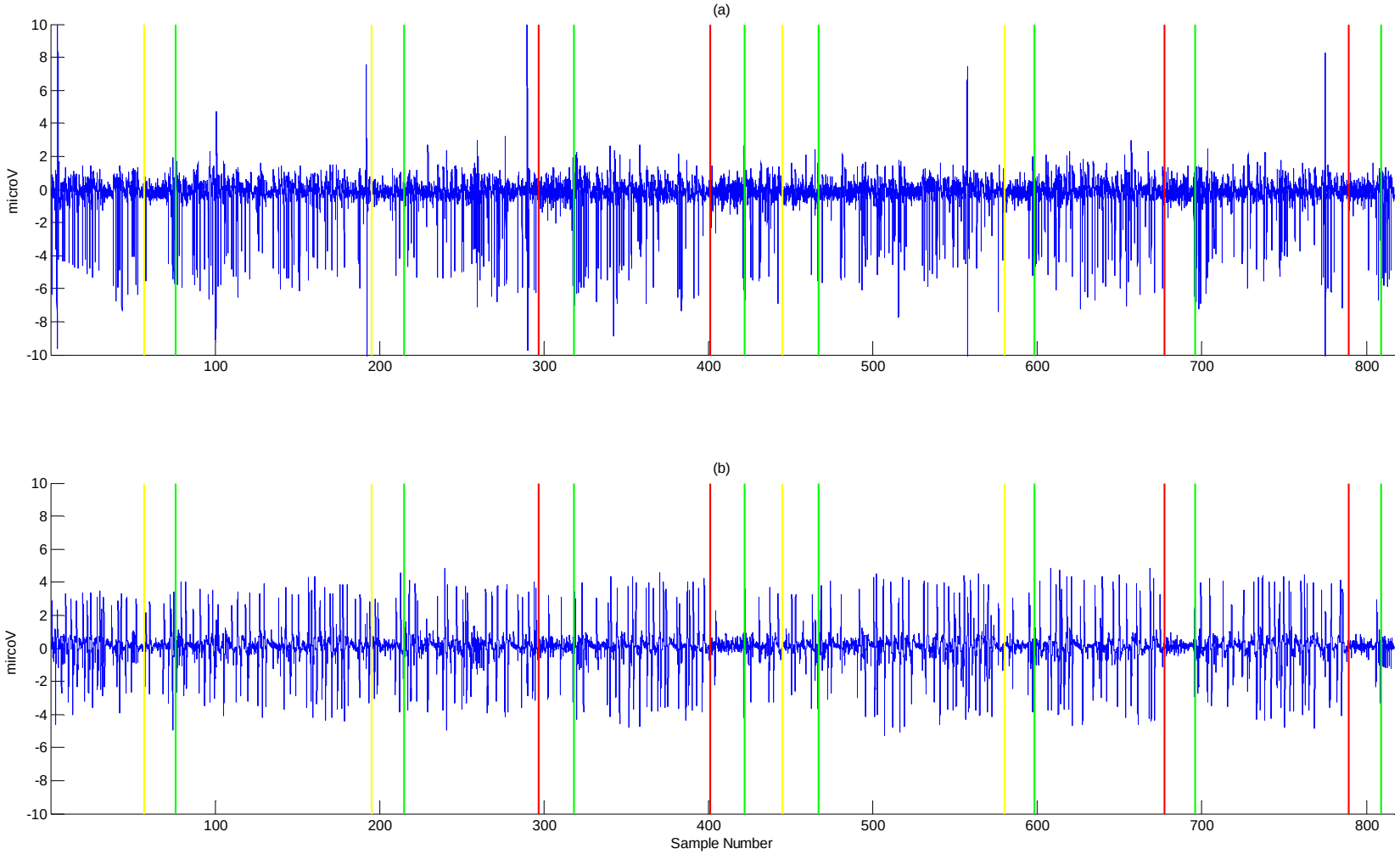

Figure 4.3 Dimensionally reduced independent components of A01: (a) IC1, (b) IC2

CHAPTER 5
REAL-TIME METRICS FOR PSYCHOPHYSIOLOGICAL SIGNALS AND ALGORITHMS FOR CHANGE DETECTION

5.1 The Peak Detection Method

Now that a robust signal, the TVIC, can be identified for each subject, a quantitative metric is necessary to characterize the signal for each epoch; this metric will be used to facilitate the online change detection of OFS. The TVIC plotted in (a) of Figure 4.3 suggests that a metric of variance might best characterize the signal's response to task load, since variance noticeably increased during the LL and drastically reduced during the ML and HL. Ideally, monitoring a metric like variance, computed for each epoch, would result in a signal with less noise and better discrimination between task loads than the raw TVIC. In essence, the magnitude of the metric would "peak" when the tasks are incurred, thus signifying changes to higher cognitive loads.

Although variance is an obvious choice, there are other quantitative metrics to characterize a signal in real-time, such as the third moment (skewness), the fourth moment (kurtosis), and finally, entropy-based metrics. Of these metrics, one may prove the most sensitive to the onset of tasks and result in the best change detection.

An objective method was needed to evaluate and compare each metric's ability to produce consistent task-induced peaks. One such technique is called the *peak detection method* (PDM). The PDM compares a measure x_i, for each epoch i, with two thresholds, an absolute threshold defined by parameter γ and a relative threshold defined by parameter δ; the comparisons are made to determine if a peak has occurred.

For the PDM analysis, each metric was computed by a sliding window to smooth irregularities and to induce peaks during the tasks. The PDM was then used to objectively determine which metric most consistently peaked for all 8 HL and ML tasks, while committing the fewest false alarms. Hence, the results of the PDM analysis indicated which metric for the TVIC best facilitated OFS change detection. The algorithm for the PDM is presented in Table 5.1 (Yu, 2009).

5.2 Review of Metrics for Characterizing Psychophysiological Signals

The simplest metric for characterizing the TVIC is the standard deviation. The standard deviation is the positive square root of a signal's second moment, defined as

$$\sigma = \sqrt{M_2} \tag{5.1}$$

where M_j is

$$M_j = \frac{\sum_{i=1}^{n}(x_i - \mu)^j}{n} \tag{5.2}$$

μ is the mean of the signal, generally estimated by the sample mean, $\bar{x}$, computed from a sample of size n. For the standard deviation calculation, n is traditionally reduced by one in the denominator of (5.2) (Montgomery, 2009).

A related metric is the third moment of the TVIC, termed skewness when normalized, which measures the asymmetric nature of a distribution. The TVIC in plot (a) of Figure 4.3 displays clear skewness in the distribution of its amplitudes during the LL, but much less skewness during the tasks. Skewness is defined as

$$\gamma = \frac{\mathrm{M}_3}{M_2^{3/2}} \tag{5.3}$$

(Montgomery, 2009).

The fourth moment may also prove a valuable metric as it characterizes the shape of a distribution, discriminating distributions that are tall and skinny from those that are short and stout. When standardized, the fourth moment is termed kurtosis, and because it is fourth order, it is always positive. Kurtosis is defined as

$$k = \frac{M_4}{M_2^2} \tag{5.4}$$

(Montgomery, 2009).

Table 5.1 PDM algorithm

1. **Initialize:**
   - 1.1. Testing set, $\boldsymbol{S} = \{x_1, x_2, \ldots, x_n\}$
   - 1.2. δ, γ are user specified
   - 1.3. $L_{max} = \infty^-$, $I_{max} = 0$
   - 1.4. $L_{min} = \infty^+$, $I_{min} = 1$
   - 1.5. $B = 1$
   - 1.6. $\mathbf{P} = \emptyset$, $\mathbf{I} = \emptyset$
   - 1.7. $i = 1$
2. **Run Algorithm:**
   - 2.1. **while** $i < n$
   - 2.2. **if** $x_i > L_{max}$
   - 2.3. $L_{max} = x_i$, $I_{max} = i$
   - 2.4. **end if**
   - 2.5. **if** $x_i < L_{min}$
   - 2.6. $L_{min} = x_i$
   - 2.7. **end if**
   - 2.8. **if** $B == 1$
   - 2.9. **if** $x_i < L_{max} - \delta$ AND $L_{max} > \gamma$
   - 2.10. $\mathbf{P} \leftarrow \mathbf{P} \cup L_{max}$
   - 2.11. $\mathbf{I} \leftarrow \mathbf{I} \cup I_{max}$
   - 2.12. $B = 0$
   - 2.13. $L_{min} = x_i$
   - 2.14. **end if**
   - 2.15. **else if** $x_i > L_{min} + \delta$
   - 2.16. $L_{max} = x_i$
   - 2.17. $I_{max} = i$
   - 2.18. $B = 1$
   - 2.19. **end if/else**
   - 2.20. $i = i + 1$
   - 2.21. **end while**
3. **Output:** peaks and their indices are contained in sets **P** and **I**, respectively

The next two metrics are both measures of entropy, where entropy quantifies the "randomness" of the TVIC. The more random the data, the higher entropy it will have, conversely, the more ordered the data, the less entropy it will have; data in perfect order has zero entropy. In recent research, entropy measures of psychophysiological data have been found to correlate with changing states of vigilance, in particular, entropy increased with increasing vigilance (Bruzzo et al., 2008; Zhang et al., 2009). This result suggests entropy metrics may be able to detect changes in OFS.

The first entropy metric considered, called sample entropy (SampEn), is a non-linear metric that measures the regularity of a time series (Richman et al., 2004). SampEn was chosen because it is computationally efficient and can be applied to short, noisy data typical of the TVIC. In addition, SampEn is an improvement over the more traditional time series entropy measure, approximate entropy (ApEn); SampEn exhibits less bias towards lower entropy, improves the measure's relative consistency between datasets, and is more robust with regards to varying record lengths (Richman, Moorman, 2000).

SampEn measures the regularity of a time series by computing the conditional probability that two arbitrarily similar epochs of size *m*, remain arbitrarily similar for the next point in the series. "Arbitrary" is used here to emphasize a degree of error allowed when classifying two epochs as similar. This is controlled by the parameter *r*, typically a factor between .1 and .25 of the record's standard deviation. By SampEn's definition, a perfectly ordered series will have a conditional probability of one, corresponding to entropy of zero. Time series with less order will have positive entropies and conditional probabilities less than one (Richman et al., 2004). The algorithm to compute SampEn is in Table 5.2 (Alcaraz, Rieta, 2008).

The second entropy metric, termed Kullback-Liebler divergence (KLD), measures the relative entropy between two probability distributions, $p = \{p_k\}$ and $q = \{q_k\}$. In the present study, frequency distributions were analyzed instead of probability distributions, where *k* is a frequency and p_k is the normalized density at *k*,

Table 5.2 SampEn algorithm

1. **Initialize:**
   - 1.1. n is the length of the record
   - 1.2. m and r are user defined parameters
   - 1.3. Form vectors of length m, $\boldsymbol{x}_m(1),\ \boldsymbol{x}_m(2), \ldots, \boldsymbol{x}_m(n-m+1)$
     - 1.3.1. $\boldsymbol{x}_m(i) = [x(i), x(i+1), \ldots, x(i+m-1)]\ \forall\ i$
   - 1.4. B_i, A_i, B^m, A^m, D are initialized to zero
   - 1.5. $i = 1, j = 1$
2. **Run Algorithm:**
   - 2.1. **while** $i \leq n - m + 1$
   - 2.2. $b = 0$
   - 2.3. **while** $j \leq n - m + 1$ AND $j \neq i$
   - 2.4. Select two vectors $\boldsymbol{x}_m(i)$ and $\boldsymbol{x}_m(j)$
   - 2.5. $$D = \max_{k=0,\ldots,m-1} |x(i+k) - x(j+k)|$$
   - 2.6. **if** $D < r$
   - 2.7. $b = b + 1$
   - 2.8. **end if**
   - 2.9. **end inner while**
   - 2.10. $$B_i = \frac{1}{n-m-1} b$$
   - 2.11. **end outer while**
   - 2.12. $$B^m = \frac{1}{n-m} \sum_{i=1}^{n-m} B_i$$
   - 2.13. Repeat steps 1-3, except for vectors of length $m + 1$. j ranges from 1 to $n - m$ and replace b with a, and B with A
3. **Output:**

$$SampEn = -ln\left[\frac{A^m}{B^m}\right]$$

similarly for q_k (Quiroga et al., 2000). The KLD computed the relative entropy between a static frequency distribution constructed from the LL data, and a frequency distribution that changed with a sliding window over the TVIC. The hypothesis was that the frequency distribution from the HL and ML differed markedly from the LL distribution, thus KLD would peak during the tasks.

KLD is always positive and its magnitude is proportional to the difference in entropy between distributions; it is zero for identical distributions. KLD is computed by

$$KLD(p|q) = \sum_k p_k ln\left(\frac{p_k}{q_k}\right) \tag{5.5}$$

5.2.1 Results of the PDM on Various Metrics

Several different metrics were presented, each with the ability to characterize the task-varying nature of the TVIC. The PDM was used to objectively evaluate which of the metrics most consistently peaked for tasks. This section presents those results, and identifies the metric which is best suited for online OFS change detection.

The standard deviation, skewness, and kurtosis metrics were all computed by a sliding window over each subject's TVIC. The span of the window was 18 seconds, providing an update of the metric every second. The length of the span was chosen for optimal peak production, as 18 seconds corresponds to the duration of the tasks; hence, at one point the span of the window precisely overlaps an entire task, resulting in the most extreme realization of the metric, i.e., the largest peak. This span may seem extensive, but it was necessary to smooth the metrics and to resolve any task-induced trends.

For the following analyses, the δ and γ parameters of the PDM were optimized with respect to each subject and metric under evaluation. This strategy is consistent with operational requirements, as between-subjects variation is too profound to generalize parameter settings. An example of the standard deviation of the TVIC from E02 is plotted

in Figure 5.1; the standard deviation is negated for interpretability and the peaks detected via the PDM are indicated by cross-hairs. All 8 tasks were detected.

SampEn was also computed by a sliding window, in this case, with a span of 10 seconds. The span was shorter because a second window was applied in addition to the first. SampEn exhibited brief, erratic outliers whose false peaks were not sufficiently smoothed by a single window. To compensate, a second window was applied over the first-windowed TVIC, and for each set of 10 points, it computed the area under the curve, i.e., approximated the integral. The double-windowing smoothed false peaks, as their short duration was not associated with large area under the curve. Figure 5.2 displays an example of this for A02. In (a), the first sliding window of SampEn was applied to the TVIC. Notice the presence of several short, narrow outlier peaks in contrast to the wider peaks exhibited during each task. Plot (b) is the integral from the second sliding window over the first windowed signal in (a); it demonstrates the dampening of outlier peaks and the correct detection of all 8 tasks with no false alarms. The parameters from SampEn were set as follows: r to .2 times the standard deviation of the record, m to 2, and the record length, n, to 2000 data points (Richman et al., 2004).

The last metric analyzed was the KLD which computed the relative entropy between two frequency distributions. The static frequency distribution was derived by averaging the frequencies of every 10-second epoch from the LL of an entire trial's TVIC; as before, the frequencies were computed via the DFT. The second distribution was computed in real-time from a sliding window over the subject's TVIC. The density of each frequency distribution was normalized, so the area under the distribution was one. After normalization, both the static baseline distribution and the real-time distribution were submitted to the KLD, for every second, to form a relative entropy signal.

The span of the sliding window for the KLD was again set to 10 seconds, because as with SampEn, the KLD necessitated a second window to dampen outlier peaks. As before, the second window slid over the initial windowed signal and computed the area

under the curve for each set of 10 points. Figure 5.3 is an example for E02 where plot (a) is the KLD windowed over the TVIC and plot (b) is the integral from the second sliding window over the first windowed signal in (a). Similar to Figure 5.2, the double windowing smoothed outliers while amplifying the task-induced peaks to detect 7 of the 8 tasks.

The overall results of the PDM analysis are in Table 5.3. The proportion of correctly detected tasks is listed for each metric by trial. The average accuracy and false alarm rate is also reported in bold for each metric. According to the results, the standard deviation was the most proficient at producing peaks for each task, as the PDM detected 93.8% of the tasks with only .5 false alarms per trial on average. The standard deviation dominated all other metrics, i.e., for no trial did another metric perform better. SampEn also had strong performance, detecting 81.3% of tasks on average, but it failed on subject F. The next most consistent metric was kurtosis, with 68.8% of the tasks correctly detected on average. Recall that kurtosis quantifies the shape of a distribution, meaning that with some success, task loads can be distinguished by the shape of the TVIC's amplitude distribution. KLD also utilized distributions, in this case, to quantify differences in entropy. The KLD metric detected 64.6% of tasks on average in the PDM.

Table 5.3 Results of the PDM on various metrics

	Std. Deviation	**Skewness**	**Kurtosis**	**SampEn**	**KLD**
A01	1.000	1.000	1.000	1.000	1.000
A02	1.000	0.250	0.750	1.000	0.750
E01	1.000	0.375	0.625	1.000	0.750
E02	1.000	0.750	0.625	1.000	0.875
F01	0.875	0.625	0.750	0.500	0.250
F02	0.750	0.625	0.375	0.375	0.250
Average	**0.938**	**0.604**	**0.688**	**0.813**	**0.646**
FA/Trial	**0.500**	**0.000**	**1.250**	**0.750**	**0.750**

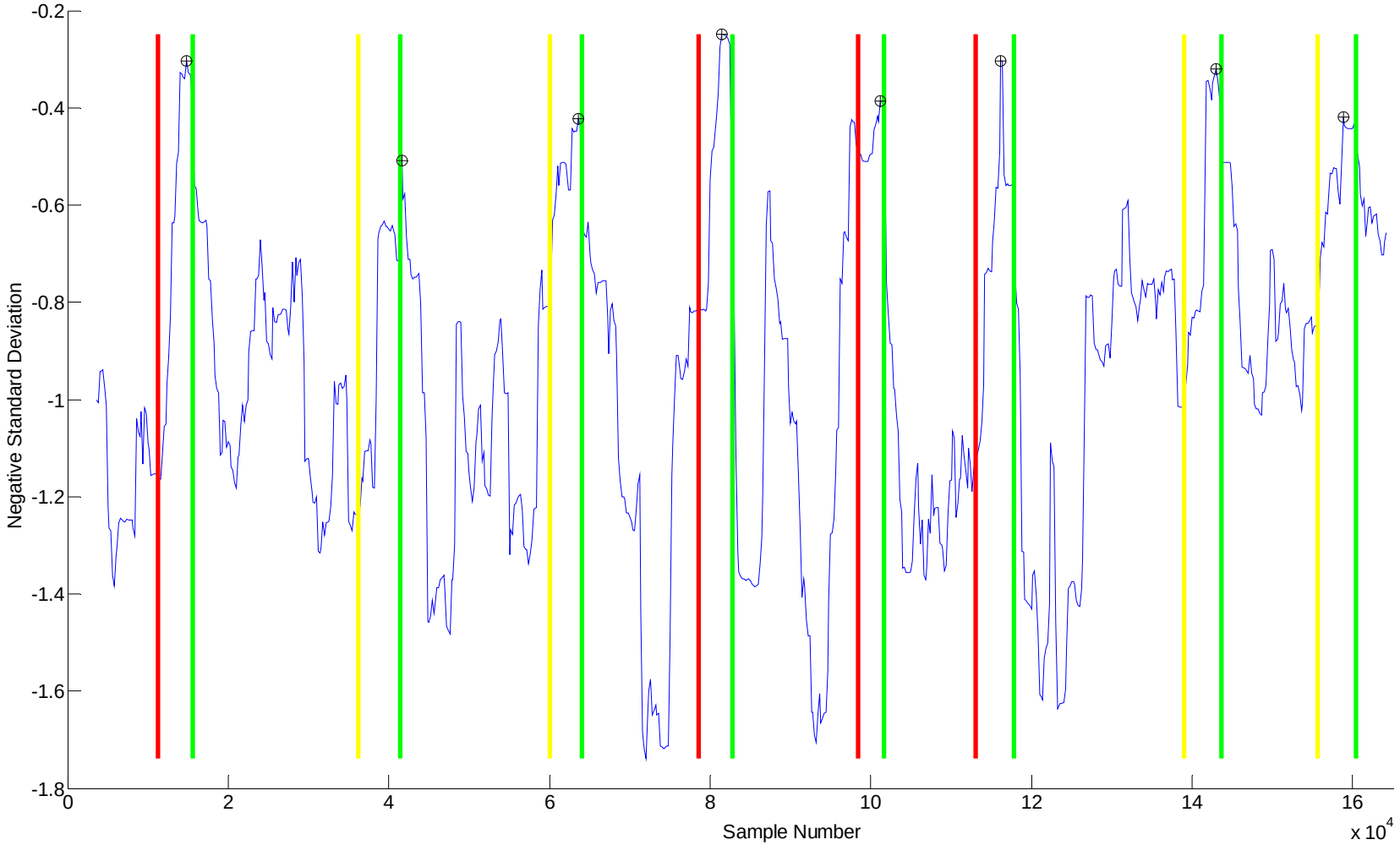

Figure 5.1 PDM on the standard deviation of the TVIC of E02

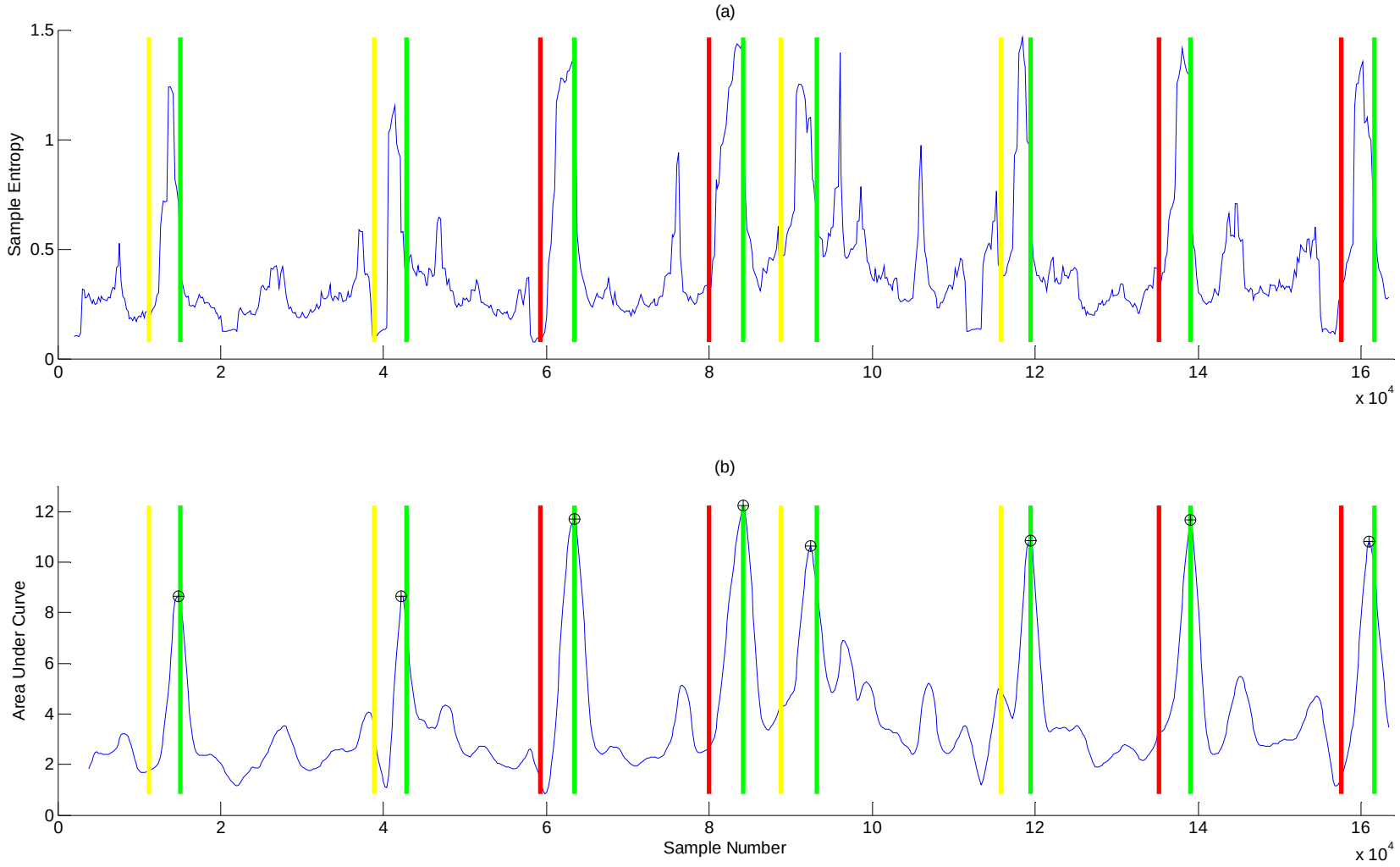

Figure 5.2 PDM on the SampEn of the TVIC of A02: (a) first sliding window, (b) second window of integral on (a)

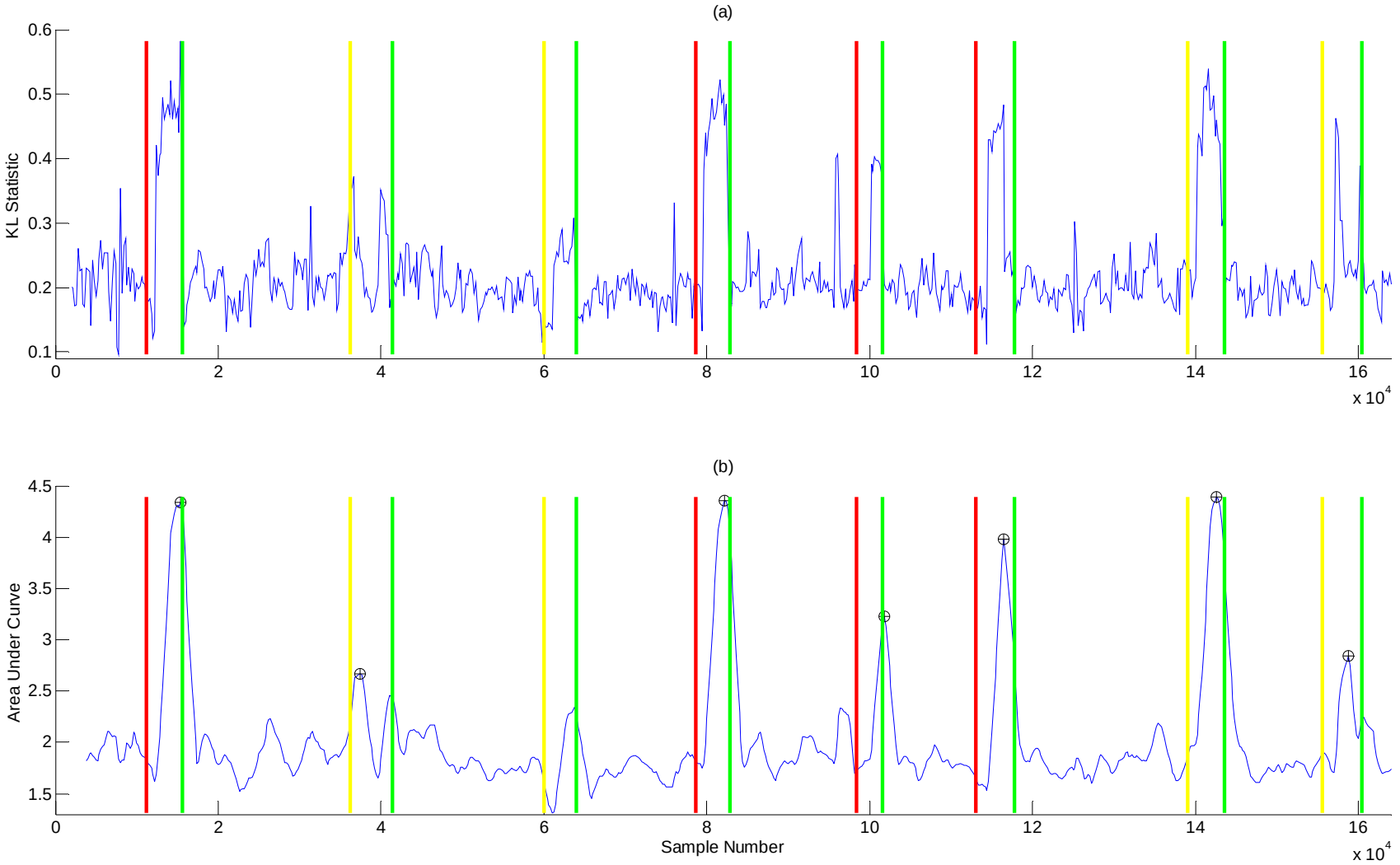

Figure 5.3 PDM on the KLD of the TVIC of E02: (a) first sliding window, (b) second window of integral on (a)

5.2.2 Discussion of Metrics and the PDM

The PDM was used as a tool to objectively evaluate how well each metric could detect change in task loads. The PDM analysis yielded that the windowed standard deviation of the TVIC was the best. By monitoring this signal and metric for each subject, a change detection scheme could accurately identify when the subject changes state.

The entropy metrics were also competitive in detecting tasks, but only for subjects A and E. Once again, subject F's psychophysiological signals proved difficult to accommodate, as they did not exhibit a recognizable entropy pattern. With data on more subjects, the entropy metrics may prove excellent in characterizing task-varying signals, on the other hand, it may be substantiated that entropies are largely subject-specific.

Yu (2009) first introduced the PDM's application to the UAV dataset. He presented task detection results from the PDM on the windowed standard deviation of each subject's F_z signal. The span of the sliding window was set to 18 seconds and the PDM parameters, δ and γ, were fixed. In order to make a fair comparison with these findings, the PDM analysis was repeated, this time with fixed parameters on the standard deviation of a subject's TVIC. The outcome of the comparison is in Table 5.4.

The average accuracy using a subject's TVIC was 2% higher than using their F_z signal. While this alone is not a significant improvement, the PDM on the TVIC also yielded less than half the false alarms on average when compared to the F_z signal.

The results support the notion of using independent components instead of the raw psychophysiological signals for OFS change detection. Independent components are less complex and less noisy, and thus result in fewer false alarms. Not only do these results confirm ICA's ability to extract artifacts, they directly demonstrate ICA's ability to facilitate the detection of change in a subject's cognitive state.

The PDM is a valuable tool, but its role should not be extrapolated to that of an online change detector. In the context of the present study, a peak is only defined once a

metric reverses direction, and begins descending back to the LL baseline. In other words, a peak is only detected once the cognitive load returns to normal, long after a task has ended. Therefore, the PDM cannot be a true online change detector, as it can only detect tasks in retrospect.

Table 5.4 Results of the PDM on the standard deviation of F_z and TVIC

	F_z	**TVIC**
A01	1.000	1.000
A02	0.875	0.750
E01	1.000	1.000
E02	0.750	0.875
F01	1.000	0.875
F02	0.500	0.750
Average	**0.854**	**0.875**
FA/Trial	**2.000**	**0.833**

The PDM does provide an interesting model, one that could be modified into a real-time scheme that detects a metric's task-induced *trends*. After identifying the standard deviation as the optimal metric to characterize a subject's TVIC, the next step is to employ the metric in a real-time OFS change detector. The following section introduces a novel algorithm, loosely based on the PDM, which monitors and detects task-induced trends instead of peaks.

5.3 The Trend Detection Method

A true online classifier cannot benefit from hindsight, meaning the information used to classify an epoch of data must come from that epoch and/or the information that preceded it. The PDM does not qualify as a true online change detector because it requires information about what transpires *after* an epoch in order to retroactively detect peaks. An online classifier must also be instantaneous, and most importantly, it must be accurate.

The *trend detection method* (TDM) introduced in this section was developed to meet these requirements and to address the deficiencies of the PDM. The TDM assumes that the ML and HL will cause a windowed metric to monotonically increase (or decrease) from the LL baseline. This behavior was observed when monitoring the standard deviation of a subject's TVIC in Section 5.2.1. The TDM identifies and detects these task-induced trends in order to detect real-time changes in OFS.

5.3.1 The TDM Algorithm

The keys to trend detection are to identify significant trends through noise and to do so for non-stationary signals, where trends can begin at different magnitudes over time. The TDM addresses these challenges via an algorithm, presented in Table 5.5, which has two main components: an adaptive threshold and a trend detector. The adaptive threshold adjusts to non-stationary behavior and the trend detector identifies task-induced trends. Detecting trends accomplishes two things: first, coupled with information on the signal's magnitude, a trend helps discriminate between true task-induced increases and irrelevant transient spikes. Second, depending on the length of the trend, tasks occurring below the threshold can be identified and adapt the threshold accordingly.

The TDM judges a signal, x_i, at time i against two criteria: its magnitude relative to the adaptive threshold and how well the previous points, $x_{i-1}, x_{i-2}, \ldots, x_{i-k}$, have

trended. Trends are identified by counting the number of positive "slopes" between consecutive points over a period of time, where τ is the count. A positive slope is defined when $x_i > x_{i-1}$. 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 of time, no point falls above the trend's last point, say, x_i. This time period is defined by the parameter R_{lim}. If, however, R_{lim} is not exceeded and some point, $x_{i+k} > x_i$, then $\tau = \tau + 1$ and the trend detection continues. When $\geq T_{low}$, the second phase of the algorithm is commenced.

Once an emerging trend is identified, the last value in the trend, x_i, is compared to the threshold, *Z*, to determine if a task-induced cognitive load has occurred. If $x_i > Z$, the trend is classified as task-induced. In contrast, if $x_i < Z$, the emerging trend remains unclassified but continues to be monitored until $\tau > T_{up}$, where T_{up} 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*, and the second, is when the trend is occurring below *Z*, but is sustained long enough to be deemed task-induced. T_{up} quantifies "long enough" by accounting for the distance x_i falls below *Z*; 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*, the threshold updates by an EWMA 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 These mechanisms of the TDM are all detailed in Table 5.5.

.

Table 5.5 TDM algorithm

1. **Initialize:**
   1.1. Testing set, **S** = $\{x_1, x_2, \ldots, x_n\}$
   1.2. T_{low}, R_{lim}, λ, β, and L are all defined during training
   1.3. $\tau = 0$, $v = 0$
   1.4. $T_{up} = \infty^+$
   1.5. $L_{max} = \mu_{\boldsymbol{X}}$, where $\mu_{\boldsymbol{X}}$ is the mean of **S** or a training set $\boldsymbol{X}$
   1.6. $Z_o = \mu_{\boldsymbol{X}} + L\sigma_{\boldsymbol{X}}$, where $\sigma_{\boldsymbol{X}}$ is the standard deviation **S** or of a training set $\boldsymbol{X}$
   1.7. **V** = ∅, **I** = ∅
   1.8. $i = 1$
2. **Run Algorithm:**
   2.1. **do**
   2.2. $i = i+1$
   2.3. **if** $x_i > x_{i-1}$
   2.4. **if** $x_i > L_{max}$
   2.5. $L_{max} = x_i$
   2.6. $\tau = \tau + 1$
   2.7. $v = 0$
   2.8. **else** $v = v + 1$ **end if/else**
   2.9. **else** $v = v + 1$ **end if/else**
   2.10. **while (** $v < R_{lim}$ AND $\tau < T_{low}$ AND $i < t$ **)**
   2.11. **if** $v \geq R_{lim}$
   2.12. $\tau = 0$
   2.13. $L_{max} = x_i$
   2.14. **else if** $x_i \geq Z$ OR $\tau \geq T_{up}$
   2.15. **V** ← **V** ∪ x_i
   2.16. **I** ← **I** ∪ i
   2.17. $\tau = 0$, $v = 0$
   2.18. $T_{up} = \infty^+$
   2.19. $L_{max} = \mu_{\boldsymbol{X}}$
   2.20. $Z = min\{ Z_o, Z(1-\lambda) + \lambda L_{max}\}$
   2.21. **else**
   2.22. **if** $\tau = T_{low}$
   2.23. $T_{up} = \left\lceil \frac{Z - x_i}{\sigma_{\boldsymbol{X}}} \beta \right\rceil + T_{low} + 1$
   2.24. $\tau = \tau + 1$
   2.25. **end if**
   2.26. **end if/else**
   2.27. **if** $i < n$ **then** repeat **do/while** loop
3. **Output**: task-induced trends and their indices are in sets **V** and **I**, respectively

5.3.2 Results of the TDM on the Standard Deviation of TVIC

The TDM was conducted on the standard deviation of an 18 second sliding window over each subject's TVIC; this is the same setup as the PDM analysis. Normally, the parameters of the TDM are calibrated during training. However, for the purpose of comparing results to the PDM, the parameters were calibrated on the testing data. In general, the parameters were set within the following ranges: T_{low} [4,6], L [1.2,1.4], β [8,9]. The α and R_{lim} parameters were both held constant at .3 and 5, respectively. Figure 5.4 is a plot of the TDM of E01; 8 task-induced trends were detected, with one false alarm. Notice the instances a trend was identified more than once for the same task; this only indicates the TDM detected the task early and continued to detect further trending as the task proceeded.

The results of the TDM analysis for each trial are in Table 5.6. The accuracy is defined as the proportion of tasks that were correctly detected. The average accuracy of the TDM was 85.4%, with over one false alarm per trial on average. Lastly, the tasks were detected, on average, 13 seconds after their onset, a comparatively quick detection when compared to the half minute delays in control charting.

Table 5.6 Results of the TDM analysis

Trial	Proportion Detected
A01	1.00
A02	.750
E01	1.00
E02	.875
F01	.750
F02	.750
Average	**.854**
FA/Trial	**1.166**

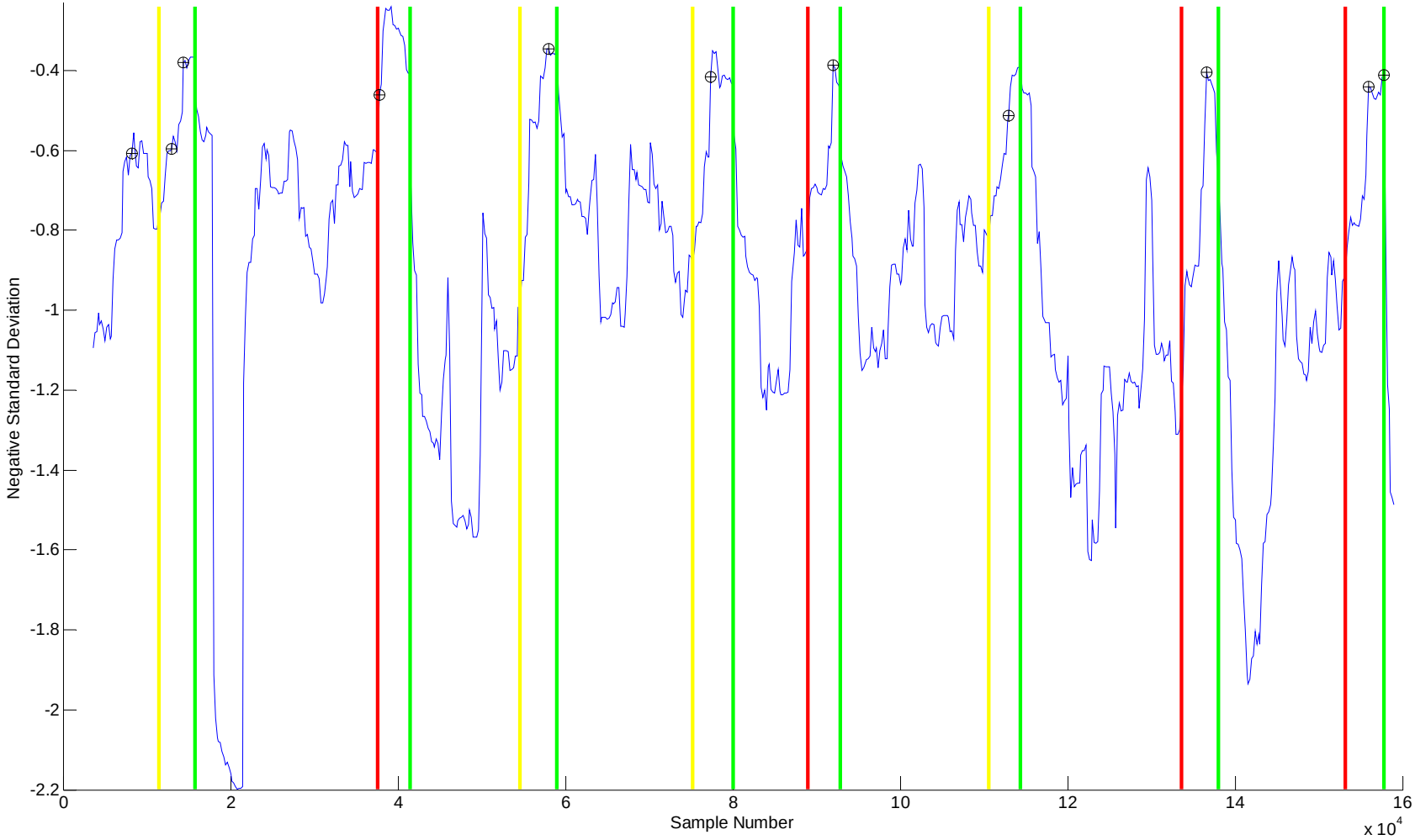

Figure 5.4 TDM on the standard deviation of the TVIC of E01

5.3.3 Discussion of the TDM

Overall, the results indicate that the TDM detected changes in cognitive load accurately, quickly, and most importantly, in real-time. Pattern recognition methods that classify epochs into discrete categories of cognition are not conducive to real-time change detection; it is unclear when a subject changes state. In contrast, the TDM, in combination with a subject's TVIC, can clearly indicate when a change has occurred.

The principle shortcomings of the TDM are twofold: its inability to signal when a task has ended and the number of parameters that must be calibrated. In adaptive aiding, it is necessary for the TDM to not only identify the beginning of a task, but also to indicate the task's conclusion so aiding can be withdrawn. This capability can be added to the algorithm by monitoring trends that reverse to the baseline once a task completes. The second drawback is the number of parameters that need to be set for the TDM, five in all. In the present analysis, only three parameters required adjustment across trials, but even setting these parameters was not straightforward. This process could be formalized by conducting a design of experiments on training data, systematically determining the optimal parameter settings for each subject.

These suggestions to remedy the shortcomings of the TDM were not pursued further, since the scope of this study is a proof-of-concept rather than an optimization of any single method. The goal of employing adaptive aiding from psychophysiological signals will require more than a single method. The TDM and the TVIC have been developed to augment the arsenal of schemes which identify changing cognitive load, and it is hoped that in concert with these established methods, future adaptive aiding schemes can be realized. The next chapter introduces a metric that is not only conducive to online change detection, but can also *measure* OFS in real-time.

CHAPTER 6
REAL-TIME MEASUREMENT AND CHANGE DETECTION OF OFS USING PSYCHOPHYSIOLOGICAL SIGNALS

6.1 Introduction to the Subject-Specific Index

In this chapter, an index of cognitive load is formulated from EEG\EOG signals. In this context, an index is a continuous measure, and unlike methods such as ANN and SWDA, it is not limited to discrete categories of OFS. Hence, it can indicate the *degree* to which a subject is engaged, not merely classify their OFS as "high" or "low". Most importantly, a continuous index facilitates online OFS change detection by exhibiting trends over time. Unlike the output of discrete classifiers, the time variation of the index can be utilized to indicate when a subject is changing from one OFS to another. Thus, after formulating an index which accurately measures OFS, the second objective of this chapter is to demonstrate the ability of the index to detect real-time changes in OFS. In this sense, the index can achieve the change detection results of the TVIC, presented in the previous chapter, in addition to *measuring* cognitive load in real-time.

The application of an index to facilitate real-time change detection is well established. One such index, 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. first created the EI to be "maximally sensitive to changes in task demand", so that when utilized in a change detection scheme, a subject's OFS could be stabilized to an optimal level (1995). The relative trending of the EI was used to detect change: depending on whether a subject's engagement was increasing or decreasing, a negative feedback control loop would allocate a task to stabilize their engagement. In this setup, Pope et al. purported that an effective index would exhibit stable, short oscillations that would result in frequent task allocations (1995). The EI which best met this criterion is computed by EEG wavebands alpha, beta, and theta as

$$\frac{\beta}{(\alpha + \theta)} \tag{6.1}$$

However, Pope et al. never indicated whether the changes detected by the EI were significant enough to warrant the modification of task load; perhaps the demands of the single task were well within a subject's cognitive capacity when viewed on an absolute scale (1995). This issue was addressed by Freeman et al. who advocated using the absolute value of the EI to drive task allocation, rather than its relative changes (1999). In other words, they deemed it necessary for an index to measure OFS before it could detect its changes. On this recommendation, Prinzel et al. investigated if the EI, in its present form, was capable of measuring engagement (2000). They augmented the single task condition with a multiple task condition in order to induce two levels of cognitive load, a low and high level, respectively. Then, they tested for a difference in the mean of the EI across the two conditions, but concluded that it failed to discriminate between levels of task load. Thus, the EI was unable to measure OFS, and therefore, any conclusions drawn about the changes it detects remain questionable.

The features which comprise the EI are not tailored to each subject, instead, they are chosen for their sensitivity to engagement when viewed across all subjects. This generalization neglects between-subject differences which can dramatically alter how well certain features correlate to engagement. In addition, recent research has shown that theta power in the parietal lobe, where it is measured for the EI, has the opposite relationship to engagement than the one EI assumes (Brookings et al., 1996; Fairclough et al., 2005; Fournier et al., 1999; Gevins, Smith, 2003; Hankins, Wilson, 1998). Therefore, EI's features are not optimized per subject, and their assumed correlation with engagement is incorrect.

This chapter derives an index with the objective to both measure OFS *and* detect its changes in real-time. To achieve this, a *Subject-Specific Index* (SSI) of cognitive load is pursued. This index was first introduced by Smith et al. (2001). The SSI's

responsiveness to task load is optimized by selecting features that perform best in a subject's training session. In addition, the SSI is computed by a multivariate distance function which enables it to further discriminate between task loads and more accurately gauge OFS. The output of this function is scaled to vary over a range from 0 to 1, with 1 representing the highest cognitive load. Collectively, these properties amount to an index that is easily interpretable, customized, and most importantly, one that can measure OFS and detect its changes in real-time.

6.2 Derivation of the Subject-Specific Index

6.2.1 Selection of Psychophysiological Features

Psychophysiological features were selected from a candidate set, shown in Table 6.1, to comprise the SSI. 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 shown to vary with task load in previous research (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 shown to vary with task load in similar UAV simulations, and has successfully been used in OFS classifiers (Wilson, Russell, 2003b; Yu, 2009). Lastly, two beta features were included to further augment the candidate set. The significance of beta was neglected in Smith et al. (2001), despite research which has consistently demonstrated its sensitivity to task load (Freeman et al., 1999; Wilson, Russell, 1995; Wilson, Russell, 1999). Some research has suggested that beta does not originate from electric potential in the brain, but from neck muscles contracting during a task (Wilson, Russell, 1995; Makeig et al., 1996). Regardless of its true source, beta was included since it may still prove effective at discriminating task load.

Training data was utilized to select a subset of features from the candidate set that were optimized with respect to each subject. In the present study, each subject had two

trials containing identical tasks; therefore one trial served to train the SSI and the other to test it.

Table 6.1 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)
T_5	alpha (8-10, 10-12 Hz), beta (14-30 Hz)
O_2	alpha (8-10, 10-12 Hz), beta (14-30 Hz)

The training trial was divided into three datasets depending on if the psychophysiological data arose from the HL, ML, or LL. For instance, data that resulted from the four HL tasks were consolidated into a single HL training set, similarly for the ML and LL data. Once the three training sets were created, the features from the candidate set were computed via the DFT in four-second epochs. These features then replaced the psychophysiological data in the three training sets.

For each subject, features which exhibited the greatest statistical distance between the HL and LL training sets were chosen from the candidate set to comprise the SSI. This resulted in an index whose average magnitude in the HL was maximally different from its average magnitude in the LL. Notice that this criterion does not concern distinguishing the ML from the HL or the LL. In reality, subjects must perform tasks on a continuum of difficulty, thus, a robust index cannot be tailored to a particular "intermediate" level task. Rather, it is reasonable to define the SSI on the extremes of task load, and then assume that tasks of intermediate difficulty, like the ML, will fall within the index.

The "distance" criterion between the HL and LL training sets was defined by the divergence. Divergence is the total average information for discriminating between two classes of data, e.g. HL and LL sets, which consist of the same features (Tou, Gonzalez, 1974). Let $\boldsymbol{x}$ be a realization of p psychophysiological features, $x_1, x_2, \ldots, x_p$. Then let the probability that $\boldsymbol{x}$ occurs given that it comes from the HL set, denoted as ω_h, be $p_h(\boldsymbol{x}) = p(\boldsymbol{x}|\omega_h)$; similarly, define $p_l(\boldsymbol{x})$ for the LL set. The divergence or "distance" between the HL and LL training sets is then

$$J_{hl} = \int_{\boldsymbol{x}} [p_h(\boldsymbol{x}) - p_l(\boldsymbol{x})] \ln \frac{p_h(\boldsymbol{x})}{p_l(\boldsymbol{x})} \, d\boldsymbol{x} \tag{6.2}$$

If it can be assumed that the distribution of $\boldsymbol{x}$ is p-variate normal, $N(\boldsymbol{m}_j, \boldsymbol{C}_j)$, where $\boldsymbol{m}_j$ is the mean vector and $\boldsymbol{C}_j$ is the p x p covariance matrix for class j, then divergence between the sets is computed by

$$J_{hl} = \frac{1}{2} \, tr[(\boldsymbol{C}_h - \boldsymbol{C}_l)(\boldsymbol{C}_l^{-1} - \boldsymbol{C}_h^{-1})] + \frac{1}{2} \, tr[(\boldsymbol{C}_h^{-1} + \boldsymbol{C}_l^{-1})(\boldsymbol{m}_h - \boldsymbol{m}_l)(\boldsymbol{m}_h - \boldsymbol{m}_l)'] \tag{6.3}$$

Divergence is additive for independent features; this important property states:

$$J_{hl}(x_1, x_2, \ldots, x_p) = \sum_{k=1}^{p} J_{hl}(x_k) \tag{6.4}$$

This indicates that the divergence computed for a number of psychophysiological features combined, will be equivalent to the sum of each feature's divergence when computed separately. Therefore, if each feature in the candidate set is rank-ordered with respect to their individual divergence, then selecting the top features will result in a subset with the most discriminatory information between the HL and LL training sets.

In fact, the property in (6.4) does not fully hold for the training data since the features are not independent. However, preliminary analysis using (6.3) to rank-order features, and selecting the top features from this list, yielded a subset superior to one

whose features exhibited the highest divergence in combination. Subsequently, the former method was utilized to select the features of the SSI.

Results from a preliminary analysis indicated that the SSI performed best when comprised of four or five features. Thus, the size of the subset was chosen to be four features. In general, this size promotes more robust and stable performance and is consistent with previous index methods (Smith et al., 2001; Pope et al., 1995).

6.2.2 Maximizing Divergence through a Multivariate Distance Function

Once the features were identified for the SSI, a multivariate distance function was utilized to maximize the divergence between the HL and LL training sets. The multivariate distance function was introduced in Smith et al. (2001) to accomplish two things: first, to further "distance" the distributions of the HL and LL conditions, thus making them easier to distinguish, and second, to transform each realization of $\boldsymbol{x}$ into a singular value.

Regarding the latter point, many methods which assess OFS utilize some form of dimensionality reduction. Psychophysiological signals constitute a multivariate time-series containing dozens, even hundreds of features which, if not reduced, can be very difficult to accommodate. For the SSI, the dimension is reduced to one in order to form an interpretable index which is conducive to trend detection methods. Although this reduction caused a loss of information, the results indicate the loss was not critical as the performance of the index remained strong.

The multivariate distance function transforms the vector $\boldsymbol{x}$ into y, termed an image of $\boldsymbol{x}$, by

$$y = \boldsymbol{a}'\boldsymbol{x} \tag{6.5}$$

where $\boldsymbol{a}$ is an p-vector of coefficients and $\boldsymbol{x}$ is the realization of p features. The transformation results in an image distribution assumed to be normally distributed, $N(m_j^*, C_j^*)$ for class j, with a mean and variance derived from its untransformed distribution by

$$m_j^* = \boldsymbol{a}'\boldsymbol{m}_j \tag{6.6}$$

$$C_j^* = \boldsymbol{a}'\boldsymbol{C}_j\boldsymbol{a} \tag{6.7}$$

The divergence between the HL and LL image distributions is then computed by

$$J_{hl}^* = \frac{1}{2}tr\left[\left(C_l^{*-1}\right)\left(C_h^*\right) + \left(C_h^{*-1}\right)\left(C_l^*\right)\right] - 1 + \frac{1}{2}tr\left[\left(\left(C_h^{*-1}\right) + \left(C_l^{*-1}\right)\right)\delta^*\delta^{*\prime}\right] \tag{6.8}$$

where $\delta^* = m_h^* - m_l^*$.

In order for the multivariate distance function in (6.5) to maximize the divergence as computed in (6.8), it is necessary that $\boldsymbol{a}$ be optimal. Optimizing $\boldsymbol{a}$ is a non-trivial problem, but it is made considerably easier if the covariances of the class distributions are assumed equal. If it can be assumed that $\boldsymbol{C}_l$=$\boldsymbol{C}_h$= $\boldsymbol{C}$, then (6.8) is maximized when $\boldsymbol{a}$ is set to the non-zero eigenvector of $\boldsymbol{C}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}'$, where $\boldsymbol{\delta} = \boldsymbol{m}_h - \boldsymbol{m}_l$ (Tou, Gonzalez, 1974). Consequently, this method was used to select an optimal $\boldsymbol{a}$ for each subject's multivariate distance function.

6.2.3 Smoothing and Scaling

The output of the multivariate distance function, derived for each subject, forms a sort of "raw" index which maximally discriminates between the HL and LL training sets. However, the index scores varied greatly by subject and were often different scales of magnitude. Therefore, a common scaling was applied as well as smoothing, in order to dampen erratic behavior inherent to psychophysiological signals (Freeman et al., 1999; Wilson, Russell, 2007).

The EWMA was chosen to smooth the SSI. Let y_i represent the output of the raw index at epoch i, then the EWMA is applied by

$$y_i^* = \lambda y_i + (1 - \lambda) y_{i-1}^* \tag{6.9}$$

where y_i^* is the smoothed index and λ is defined as $0 \leq \lambda \leq 1$. A suitable λ for this analysis was chosen to be 0.35, which aptly balanced the tradeoff of smoothing irregularities and remaining responsive to change. y_0^* was initialized to the mean of y from the training data.

Once smoothed, the index was scaled to range from 0, the least cognitive load, to 1, the most cognitive load. This scaling was achieved through normalization, defined as

$$\mathcal{s}_i = \frac{y_i^* - \ell}{\mathcal{r}} \tag{6.10}$$

where $\mathcal{s}_i$ is the SSI, $\mathcal{r}$ is the range computed by $\mathcal{r} = \mathcal{u} - \ell$, and $\mathcal{u}$ and ℓ are the upper and lower limits, respectively. It is customary to define $\mathcal{u}$ and ℓ on the upper and lower extremes of y^*, respectively. However, this definition yields a scale susceptible to outlier points. For instance, if an unusually large y_i^* occurred, possibly an outlier with an assignable cause, then defining $\mathcal{u}$ on this value would artificially skew the index towards 0. The variable nature of psychophysiological signals virtually assures the presence of abnormally large and/or small y_i^*, thus the normalization method must be defined more robustly.

To prevent extreme data from skewing the scale, range restriction was employed prior to normalization. Range restriction defines $\mathcal{u}$ and ℓ not on the extremes of y^*, but rather to encompass the bulk of the y^* distribution. In our analysis, y^* was permitted to vary two standard deviations above the y^* mean of the HL training set, and two standard deviations below the y^* mean of the LL training set; any y^* which occurred beyond this range was truncated.

To formalize the definition of r, consider the means of y^* from the HL and LL training sets to be defined as $\bar{y}^*_{HL}$and $\bar{y}^*_{LL}$, respectively. Similarly define the variances of y^* to be s^2_{HL} and s^2_{LL}. The variances are averaged to derive a common standard deviation, $\bar{s}_{y^*}$, defined as

$$\bar{s}_{y^*} = \sqrt{\frac{(s^2_{HL} + s^2_{LL})}{2}} \tag{6.11}$$

Finally, the limits of r are defined:

$$\begin{aligned} u &= \bar{y}^*_{HL} + 2\bar{s}_{y^*} \\ \ell &= \bar{y}^*_{LL} - 2\bar{s}_{y^*} \end{aligned} \tag{6.12}$$

These limits were used to range restrict and normalize the index in (6.10), resulting in an SSI which varied between 0 and 1.

Finally, it should be reemphasized 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. Therefore, it was assumed that a subject's variation between training and testing trails would not be so great as to warrant new scaling.

6.3 Measuring Cognitive Load with the Subject-Specific Index

The objective of this section is to demonstrate the ability of SSI to distinguish between levels of a UAV task. If the SSI can delineate task load, and thus cognitive load, it can serve as a proxy measure of OFS and be integrated into a real-time OFS change detector.

The following analysis was done "a posteriori", i.e. it was not done in real-time. The goal was to evaluate how SSI discriminated between the HL, ML, and LL *on average*. To accomplish this, the time aspect of the data was lost. Regardless, in Section

6.4 it is confirmed that an SSI which accurately measures OFS on average and a posteriori, will also perform well with respect to the real-time detection of changing OFS.

6.3.1 Methods

As aforementioned, each subject had two trials of data; one trial was assigned as training data and the other as test data. The training data was used to define three key components of the SSI: a subset of psychophysiological features which best discriminated between the HL and LL, a multivariate distance function to dimensionally reduce the subset and to maximize the divergence between task loads, and finally, to define the parameters of normalization and range restriction which scale the index between 0 and 1. Once the SSI was defined, its performance was evaluated on the test data. As performed for the training data, the test data was parsed into three testing sets, HL, ML, and LL, corresponding to the task load in which the data arose.

Two sets of results were generated for each subject by rotating a trial's role between train and test. This is a cross-validation mechanism which served to account for a subject's between-trial variability and to reveal the degree of dependence the SSI had on the training.

6.3.2 Results

The four features which most frequently comprised the SSI 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 wavebands 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 "untraditional" features in the candidate set. Interestingly, they were the most consistently divergent features. VEOG (5-8 Hz) was especially prominent as it was included in the SSI of every trial. This feature is primarily a measure of eye activity and its importance is a reflection of the UAV tasks' visual nature. This implores future

research to construct candidate sets based on a careful review of the tasks involved and how they affect cognition.

Table 6.2 displays the means of the SSI across the three task loads for each trial. Greater index scores correspond to higher cognitive loads while lower index scores indicate lower cognitive loads. The average SSI monotonically increased from the LL to the HL for subjects A and E, but not for subject F.

Table 6.2 SSI means across task loads

Testing Trial	HL	ML	LL
A01	0.807	0.683	0.265
A02	0.725	0.462	0.185
E01	0.762	0.407	0.150
E02	0.901	0.482	0.247
F01	0.626	0.634	0.591
F02	0.546	0.639	0.319

ANOVAs were conducted to determine which results in Table 6.2, if any, were statistically significant. For each subject, a mixed-model was formulated as

$$\mathcal{s} = L_k + T_j + L_k T_j + \epsilon_{kj} \tag{6.13}$$

where $\mathcal{s}$ is the index score predicted by: L_k, the task load of level k, T_j the trial j, $L_k T_j$, the interaction between task load and trial, and finally, an error term, ϵ_{kj}. The sole fixed-effect, task load, was expected to explain the variability in SSI. 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 6.3 contains the ANOVA results for each subject. For subjects A and E, task load was a significant predictor of the SSI, with *p*-values well below the .05 threshold. In contrast, there is no evidence that subject F's SSI varied with task load.

Table 6.3 Results of type III tests on 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

Post-hoc analyses were conducted to reveal which SSI task load means significantly differed from each other. Table 6.4 contains the *p*-values of 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 HL and LL SSI means were significant at the .05 threshold. Again, subject F's SSI was not distinguishable by task load.

These results support SSI's ability to discriminate between the HL and LL for subjects A and E. For these subjects, the SSI also discriminated the ML with significance or near significance in three of the four pair-wise tests; this despite the index not being optimized to delineate intermediate level tasks. Section 6.4 will address whether these statistically significant results prove *practically* significant in real-time OFS change detection. It is expected that an SSI which significantly measured task load will also be proficient in detecting task load changes in real-time; in contrast, the SSI which measured task load poorly, like that of subject F, is predicted to be incapable of detecting its changes in real-time.

Table 6.4 Results of pair-wise tests for the difference of task load means

Subject	HL vs. LL	HL vs. ML	LL vs. ML
A	0.030	0.237	0.072
E	0.009	0.033	0.064
F	0.928	1.000	0.620

6.4 Real-Time OFS Change Detection Using the Subject-Specific Index

In the "a posteriori" analysis, the SSIs of some subjects successfully distinguished between levels of task load on average; these SSIs are considered accurate measures of cognitive load, and thus OFS. In this section, each SSI is evaluated to determine if it can facilitate real-time detection of changing OFS. If an SSI proves successful at both OFS measurement and detection, it could be used to initiate dynamic task allocation in a future adaptive aiding scheme.

6.4.1 Methods

The evaluation was conducted on the UAV dataset which consists of four HL and four ML tasks for each trial. If effective, the SSI will detect an increased cognitive load for all eight tasks, when compared to the baseline LL. As before, this criterion hinged on several assumptions: first, varying task load directly affected cognitive load, second, other factors potentially affecting cognitive load were controlled, and lastly, it was assumed that the subject had ample time to return to the baseline cognitive state between the HL and ML tasks.

To assess SSI's change detection ability, the TDM was again employed as an objective evaluator (see Section 5.3). The index responds to the abrupt changes in tasks,

not by sudden spikes, but by exhibiting steady trends; this is because the SSI is the product of a smoothing filter. The TDM was used to detect these task-induced trends.

The parameters of the TDM were calibrated through training and subsequently used during testing. Cross-validation was again performed, producing two sets of results per subject by swapping training and testing trials.

6.4.2 Results

The TDM evaluated SSI's ability to detect real-time changes in cognitive load. The metric used to judge the SSI was the proportion of tasks correctly identified by the TDM. Tasks were successfully identified when the TDM signaled a task-induced trend directly following a task's onset. Occasionally the TDM signaled a trend when there was no associated task load; these false alarms were tabulated as another metric of performance.

Figure 6.1 and Figure 6.2 present examples of the TDM on the SSIs of subjects A and F, respectively. Trends that were classified as task-induced are denoted by cross-hairs. As shown in Figure 6.1, the TDM correctly identified 7 of the 8 tasks. In Figure 6.2, on the other hand, every task eluded detection and there was one false alarm.

Table 6.5 displays the results of evaluating the SSI via the TDM for each trial. The SSI of subjects A and E performed with an average task detection rate of 81.25%. The SSI of subject F proved wholly ineffective at detecting task-induced changes in cognitive load, only detecting 12.50% of the tasks on average. Potential reasons for this failure are offered in the following section. Including the results of subject F, the average task detection rate of the SSI was 58.3%, averaging less than one false alarm per trial. Notice the false alarm rate is comparable to the other methods.

Table 6.5 Results of the TDM on the SSI

Testing Trial	Proportion Detected	False Alarms
A01	.750	1
A02	.875	0
E01	.750	0
E02	.875	2
F01	.000	1
F02	.250	1
Average	**.583**	**0.83**

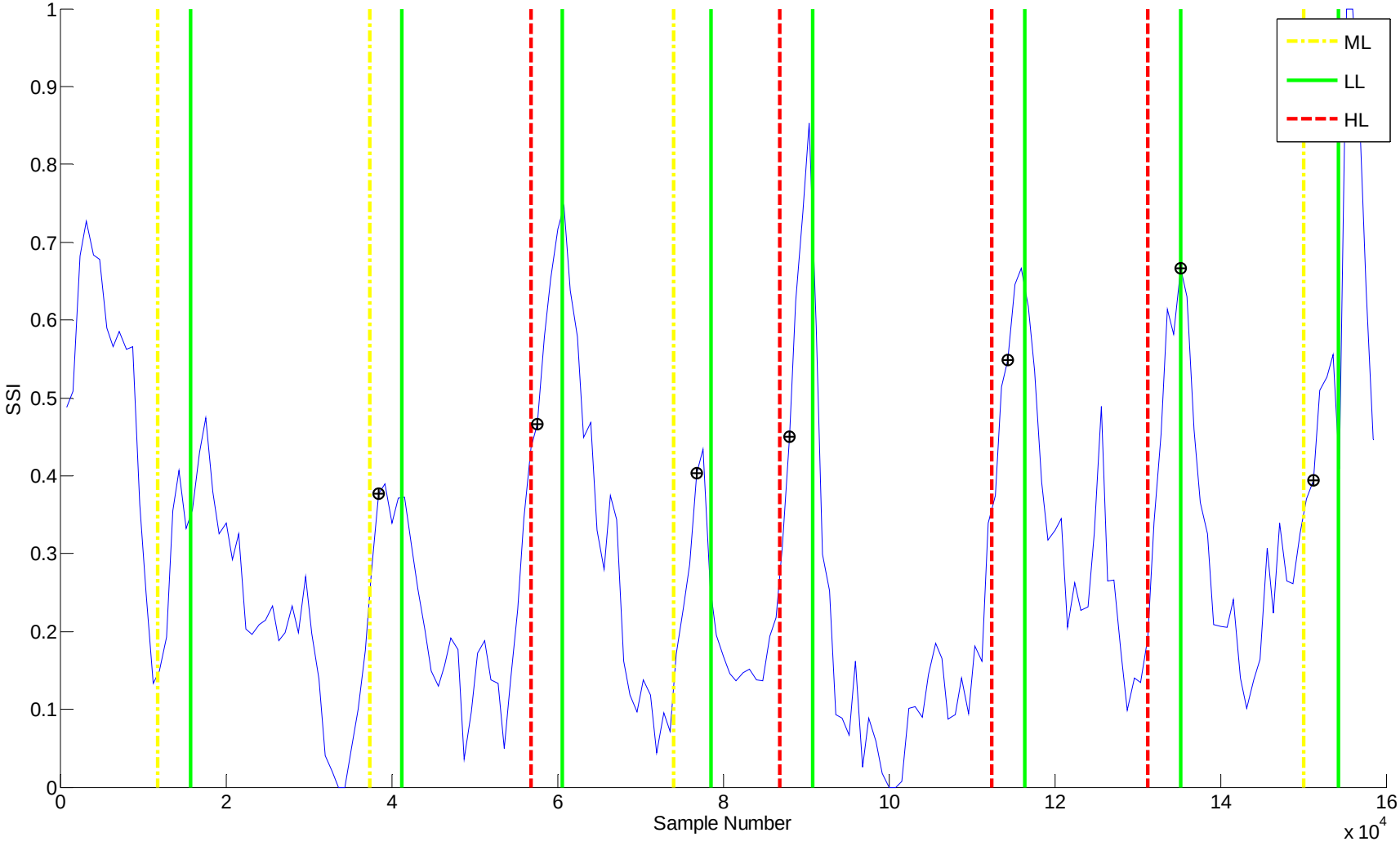

Figure 6.1 TDM on the SSI of A02

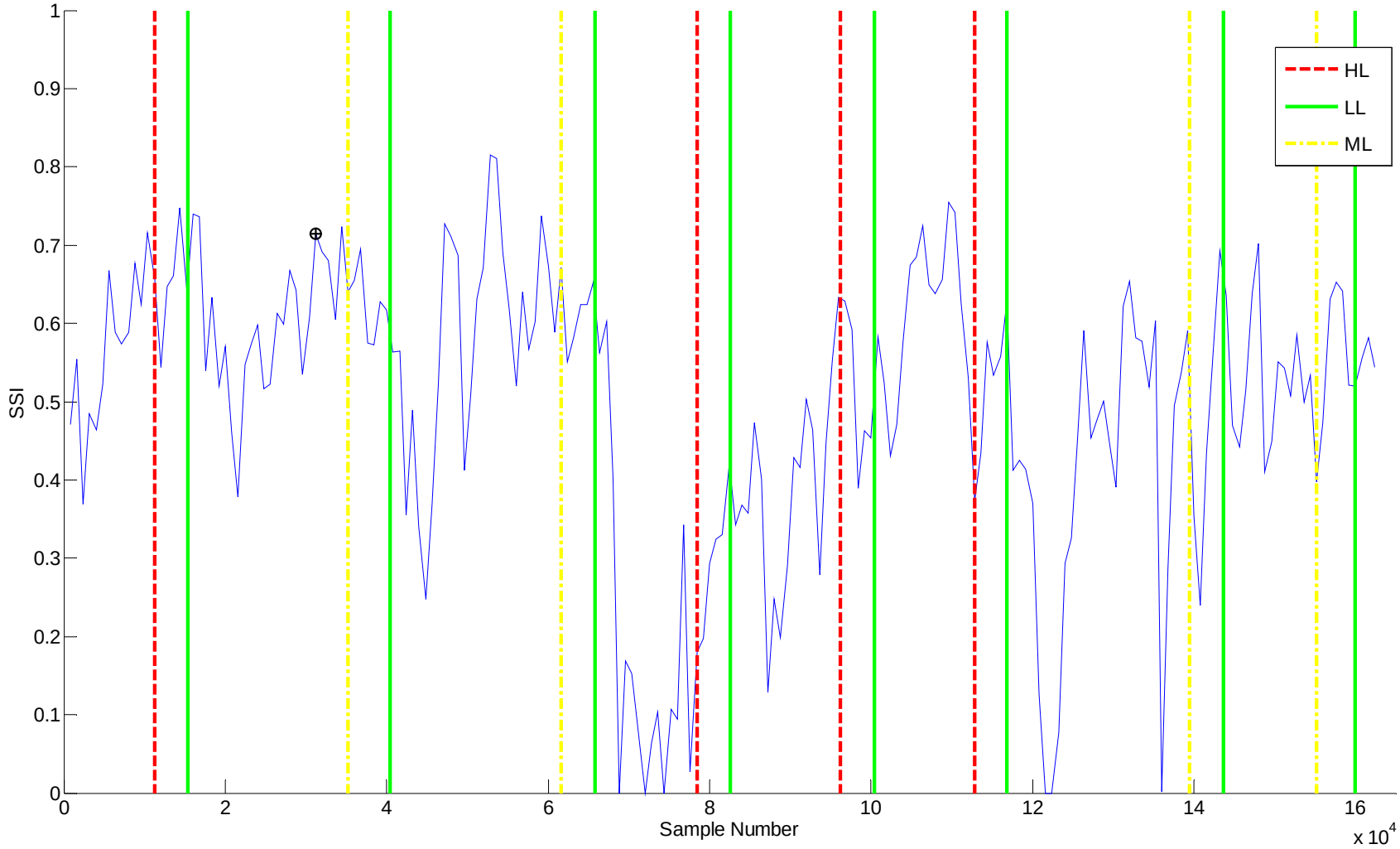

Figure 6.2 TDM on the SSI of F01

6.5 Discussion of the Subject-Specific Index

For subjects A and E, the SSI demonstrated that it could accurately detect real-time changes in OFS. However, this was not the case for the SSI of subject F, which failed to respond to changing task load.

Recall that in the average-case analysis done a posteriori, Subject F's SSI failed to statistically discriminate between levels of task load. The SSIs of subjects A and E, on the other hand, were proficient in discriminating levels of task load, and consequently, they enjoyed vastly superior online detection results. In other words, the index which *measured* OFS accurately, most consistently detected its changes in real-time. This was the index sought: one that could detect online changes of OFS and moreover, be able to measure those changes on an absolute scale.

In this study, the SSI was not effective for every subject. This was also observed by Smith et al., who argued that poor performance of the SSI does not necessarily indicate a deficiency, but may be the result of subject differences (2001). These differences can be physiological, such as, the EEG\EOG response to task load for some subjects may be too subtle for detection. Or perhaps the difference is behavioral, like when a subject's effort does not increase with increasing task demands. For instance, subject F may have been unwilling to exert the effort necessary in the HL but did try in the LL. Such a strategy would have violated the implicit assumption that subjects increased their effort proportional to task demands. When this assumption is violated, any attempt to correlate task load to cognitive load is futile.

It was also assumed that a subject's cognitive load was not influenced by factors other than task load, however, this too was possibly violated. An analysis of the SSI's behavior revealed that indeed other factors were likely present, principally, factors internal to a person's cognition. For example, the SSI was often uncharacteristically high at the beginning of a trial, when no task was present, and it often spiked after a task ended. These phenomena may be explained by cognitive constructs such as anxiety and

self-reflection, respectively. Factors such as these were not controlled in the experiment, consequently, their effects on cognitive load might be the cause of false alarms which were not "false" at all; they were valid signals not caused by task load.

The TDM was also accountable for a portion of the inaccuracy attributed to the SSI. On several occasions the SSI clearly signaled a change in task load, but since it did not trend "long enough", the TDM did not classify the change as task-induced. An example of this is in Figure 6.1, where the SSI clearly increases for the first task, but the TDM fails to indentify it.

Lastly, although the SSI has its merits, it is not a perfect index. In fact, several alternate versions of the SSI were attempted, but none were an improvement. In one version, the use of traditional wavebands as features (e.g. alpha, beta, theta, etc.) were abandoned, and instead, a candidate set was constructed based on subject-specific bands. These features were derived from analyses on a subject's full EEG and EOG frequency spectrums and how they varied with task load. For example, Figure 6.3 displays the frequency spectrum from 0 to 40 Hz of an EEG signal from E01 that was parsed by task load; analyzing this plot would suggest a waveband from 7-10 Hz to be a feature in the candidate set of subject E. Despite this concept yielding positive results in previous research, it did not improve on the SSI (Wilson, Fisher, 1995; Jung et al., 1997).

Another version of the SSI attempted to optimize the coefficient vector $\boldsymbol{a}$ of the multivariate distance function through a steepest-ascent approach (Tou, Gonzalez, 1974). At present, $\boldsymbol{a}$ is determined by assuming that the class distributions are multivariate normal and their covariances equal. Since these assumptions are often violated, it follows that $\boldsymbol{a}$ is not optimal. However, attempts to use the steepest-ascent approach to optimize $\boldsymbol{a}$ failed to offer any improvement on SSI's ability to measure or detect OFS.

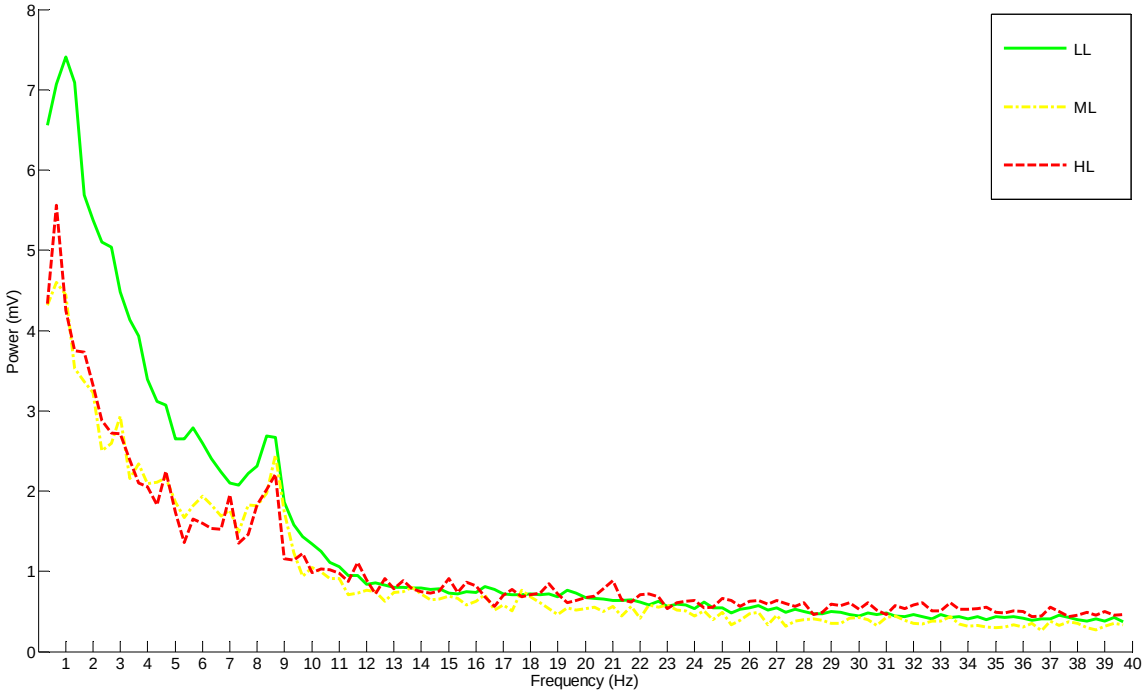

Figure 6.3 Frequency spectrum of E01 across task loads

CHAPTER 7
CONCLUSION

The following study was conducted on psychophysiological data collected during an experiment on subjects performing UAV tasks at three levels. The experiment assumed that changing the levels of task load induced corresponding changes in cognitive load. Moreover, the psychophysiological signals were assumed to respond to the changes in cognitive load. The objective of this study was to conduct a statistical analysis of these signals, and to develop algorithms which utilize the signals to detect real-time change in cognitive load. In this study, cognitive load was assumed to be a proxy measure for an operator's functional state (OFS); thus detecting a change in task load, is synonymous with detecting a change in OFS.

In Chapter 2, statistical analysis was conducted to characterize several properties of EEG/EOG features, including: normality, inter-correlation, autocorrelation, and stationarity. All of the features exhibited some degree of inter-correlation, i.e., correlation with other features, and they were all non-normally distributed. Some of the features, such as beta and gamma, also exhibited autocorrelation. Finally, despite the erratic behavior inherent in psychophysiological signals, all the features were at least moderately stationary, signifying their means did not vary with time.

In Chapter 3, univariate and multivariate control charts were evaluated for their ability to detect real-time changes in OFS using psychophysiological signals. Control charts were presented as a simple, yet objective change detector, and a possible alternative to popular pattern recognition classifiers. Multivariate control charts were expected to perform better than univariate charts, since they monitor multiple features simultaneously. Subsets of principle components were also monitored in multivariate charts, to realize an even greater theoretical improvement in detection accuracy. However, the best performing control chart was the univariate EWMA chart monitoring a single subject-specific feature. This method detected 75% of the tasks on average,

compared to only 62.5% for the best multivariate control chart. More surprisingly, monitoring a subset of principle components yielded the worst performance among all charts evaluated. Overall, control charting was deemed effective in online OFS change detection, however, it displayed several shortcomings. First, the control charts exhibited too many false alarms; these alarms need to be minimized in operational settings. Second, the control charts were not sensitive enough to detect the shifts in the means of features across task loads. This frequently caused the charts to be delayed when detecting tasks. Lastly, and most importantly, control charts were simply too rigid and required too many assumptions to accommodate the bulk of the psychophysiological data. It was evident that more flexible and robust methods were necessary to detect change in psychophysiological signals.

In Chapter 4, independent component analysis was employed to extract a task-varying independent component (TVIC) for each subject. The TVIC was a compilation of a subject's eye activity and frontal lobe power; features which consistently responded to changing task load in prior research. Compared to the raw psychophysiological signals, the TVIC was less noisy, and less complex, as it represented a "pure" source signal.

In Chapter 5, a variety of metrics were evaluated for their ability to characterize the TVIC's task-variation with time. These metrics included: standard deviation, skewness, kurtosis, sample entropy, and the Kullback-Liebler divergence. Each metric was computed by a sliding window over the TVIC, and submitted to the peak detection method which evaluated how well the metrics "peaked" during tasks. The results indicated that the standard deviation best characterized the TVIC's task-variation, and subsequently, it was chosen as the metric to further facilitate online OFS change detection.

Section 5.3 of Chapter 5 introduced the trend detection method (TDM) as a true online change detector. For each trial, the windowed standard deviation of the TVIC was submitted to the TDM to detect real-time changes in OFS. This method detected 84.5%

of the tasks on average, an improvement over control charts. Moreover, the TVIC in combination with the TDM averaged fewer false alarms than control charts.

In Chapter 6, the Subject-Specific Index (SSI) of cognitive load was presented. This measure was designed to accurately measure OFS *and* facilitate the detection of its changes in real-time. In essence, the SSI combined the strengths of pattern recognition classifiers, which measure cognitive load, and the strengths of the TDM/TVIC combination, which detects its changes in real-time.

The SSI achieved this goal through the following derivation. First, features of the SSI were selected from a candidate set based on a divergence criterion. Second, a multivariate distance function was used to further maximize SSI's response to task load. Lastly, the index was smoothed and scaled to range between 0 and 1.

The SSI results included mixed-model ANOVAs conducted for each subject. For subjects A and E, the SSI proved to be a statistically significant indicator of task load on average; thus, at least in an "a posteriori" analysis, SSI accurately measured cognitive load. For these subjects, the SSI went on to facilitate real-time change detection in the TDM, with 81.25% of the tasks detected on average.

However, the SSI was not effective for subject F. In fact, every method evaluated including those from previous research, displayed the worst performance on subject F's psychophysiological data. This phenomenon may have been caused by subject F's physiological differences, such as less powerful brain waves, or more likely, his/her behavioral differences, where subject F may not have increased his/her motivation proportional to task load. Regardless, this highlights the difficulty of adaptive aiding methods to generalize across subjects and tasks.

In conclusion, this study has introduced several algorithms which detect changing OFS in psychophysiological signals, and in the case of the SSI, measure those changes on an absolute scale. All the algorithms displayed some degree of proficiency when evaluated on the UAV dataset. However, further research is necessary to experiment and

test each method's efficacy in other applications and on other datasets. It is hoped that at least one of these methods will emerge as robust, and will contribute to the promising future of adaptive aiding.

APPENDIX A
TASK LOAD MEANS ANALYSIS

Table A.1 Means of features across task loads for A01

	Theta	**Alpha**	**Beta**
LL	1.182	0.637	0.988
ML	1.246	0.609	0.809
HL	1.213	0.572	0.614

Table A.2 Means of features across task loads for E01

	Theta	**Alpha**	**Beta**
LL	1.685	0.569	1.023
ML	1.564	0.447	0.654
HL	1.493	0.635	0.870

Table A.3 Means of features across task loads for F01

	Theta	**Alpha**	**Beta**
LL	0.987	0.332	0.585
ML	0.925	0.277	0.414
HL	1.026	0.263	0.439

Table A.4 Means of features for A01, E01, and F01 averaged across task loads

	Theta	**Alpha**	**Beta**
LL	1.283	0.514	0.866
ML	1.237	0.435	0.614
HL	1.241	0.487	0.639

APPENDIX B
FALSE ALARM DATA FROM CONTROL CHARTS

Table B.1 False alarms of EWMA-Shewart control charts

	VEOG Theta	F_z Theta	F_z Alpha	F_7 Theta	0_2 Alpha	P_z Alpha	Average
A01	1	2	1	2	0	1	**1.167**
E01	0	0	1	1	0	0	**0.333**
F01	1	0	1	0	1	0	**0.500**
Average	**0.667**	**0.667**	**1.000**	**1.000**	**0.333**	**0.333**	

Table B.2 False alarms of time series residuals control charts

	HEOG Beta	F_z Beta	P_z Beta	T_5 Beta	Average
A01	1	1	2	1	**1.250**
E01	2	1	1	1	**1.250**
F01	4	4	3	1	**3.000**
Average	**2.333**	**2.000**	**2.000**	**1.000**	

Table B.3 False alarms of MCEMWA control charts

	F_z Beta	P_z Beta	T_5 Beta	Average
A01	3	3	3	**3.000**
E01	2	2	1	**1.667**
F01	2	4	2	**2.667**
Average	**2.333**	**3.000**	**2.000**	

Table B.4 False alarms of various multivariate control charts

	Hotelling-T^2 (Select)	**MEWMA (Select)**	**Hotelling-T^2 (PCA)**	**MEWMA (PCA)**	**Average**
A01	4	2	4	3	**3.250**
E01	4	1	2	2	**2.250**
F01	6	4	4	2	**4.000**
Average	**4.667**	**2.333**	**3.333**	**2.333**	

APPENDIX C
PRINCIPLE COMPONENT FACTOR PATTERN ANALYSIS

Table C.1 Principle component factor pattern of A01

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
VEOG	**delta**	0.301	-0.696	0.295	-0.036	0.053	-0.460	0.098	-0.016	0.017	-0.026
	theta	0.353	-0.634	0.308	-0.039	0.029	-0.463	0.236	0.037	0.109	-0.095
	alpha	0.466	-0.583	0.024	-0.029	-0.039	-0.368	0.359	0.003	0.074	-0.045
	beta	0.429	-0.093	-0.667	-0.106	0.136	-0.049	0.454	-0.017	-0.197	0.045
	gamma	0.338	0.168	-0.680	-0.083	0.118	-0.002	0.426	-0.034	-0.237	0.101
HEOG	**delta**	0.030	-0.786	0.129	-0.129	-0.074	-0.082	-0.314	-0.024	0.044	0.042
	theta	0.159	-0.826	-0.147	-0.162	0.019	0.041	-0.296	-0.007	0.080	-0.017
	alpha	0.191	-0.695	-0.512	-0.191	0.171	0.202	-0.161	0.039	0.067	-0.066
	beta	0.116	-0.645	-0.581	-0.195	0.271	0.175	-0.116	0.069	0.075	-0.081
	gamma	0.148	-0.632	-0.573	-0.167	0.290	0.157	-0.105	0.076	0.071	-0.071
F_z	**delta**	0.355	-0.207	-0.005	0.497	0.140	0.086	-0.157	-0.027	0.036	0.604
	theta	0.230	-0.235	-0.149	0.014	-0.491	0.021	-0.288	-0.324	-0.568	-0.028
	alpha	0.106	-0.237	0.250	-0.191	-0.120	0.437	0.300	-0.560	0.235	0.080
	beta	0.804	0.275	0.027	-0.103	0.036	0.071	-0.025	-0.018	0.063	0.030
	gamma	0.808	0.432	-0.031	-0.115	0.034	-0.039	-0.157	-0.003	0.058	-0.006
F_7	**delta**	0.429	-0.471	0.114	0.324	0.138	-0.236	-0.020	-0.089	-0.063	0.360
	theta	0.398	-0.637	0.009	-0.013	-0.198	-0.157	-0.168	-0.161	-0.153	-0.049
	alpha	0.450	-0.356	0.070	-0.017	-0.312	0.256	0.198	-0.305	0.303	-0.017
	beta	0.745	0.306	-0.172	-0.017	-0.013	0.095	0.022	-0.024	0.052	0.094
	gamma	0.776	0.337	-0.096	-0.040	0.010	-0.017	-0.084	-0.019	0.025	0.135

Table C.1—Continued

P_z	**delta**	0.378	-0.155	0.017	0.759	0.167	0.182	-0.027	-0.048	-0.043	-0.104
	theta	0.287	-0.341	0.042	-0.022	-0.632	0.140	0.177	0.180	-0.235	-0.013
	alpha	0.174	-0.250	0.590	-0.313	0.320	0.347	0.115	0.042	-0.257	-0.035
	beta	0.852	0.090	0.213	-0.110	0.084	-0.004	-0.034	0.062	-0.025	-0.072
	gamma	0.790	0.421	-0.036	-0.149	-0.041	-0.064	-0.194	-0.030	0.070	-0.065
T_5	**delta**	0.338	-0.120	0.002	0.760	0.183	0.089	0.038	-0.013	-0.073	-0.238
	theta	0.328	-0.288	-0.074	0.223	-0.470	0.160	-0.024	0.178	0.171	-0.181
	alpha	0.382	-0.146	0.504	-0.350	0.234	0.298	0.057	0.020	-0.152	-0.019
	beta	0.845	0.329	0.020	-0.047	0.011	-0.079	-0.010	0.067	0.019	-0.095
	gamma	0.768	0.509	-0.116	-0.125	-0.053	-0.063	-0.061	-0.015	0.003	-0.068
O_2	**delta**	0.301	-0.087	0.114	0.761	0.222	0.151	0.066	-0.055	-0.038	-0.256
	theta	0.216	-0.197	-0.090	0.195	-0.517	0.251	0.143	0.504	0.125	0.179
	alpha	0.291	-0.370	0.517	-0.198	0.214	0.262	0.056	0.335	-0.155	0.088
	beta	0.852	0.021	0.213	-0.040	0.038	-0.041	-0.024	0.080	0.007	0.001
	gamma	0.827	0.395	-0.011	-0.118	-0.028	-0.053	-0.153	-0.020	0.067	-0.049

Table C.2 Principle component factor pattern of E01

		P1	P2	P3	P4	P5	P6
VEOG	**delta**	-0.5189	-0.6534	0.144	-0.2419	0.0522	0.2724
	theta	-0.4243	-0.7641	0.0467	-0.159	0.1439	0.2716
	alpha	-0.433	-0.6961	0.1683	-0.0844	0.0845	0.3606
	beta	-0.6895	-0.3628	0.0832	-0.3572	0.1927	0.0261
	gamma	-0.7851	0.0523	0.2799	-0.2339	0.272	-0.0508
HEOG	**delta**	-0.336	-0.4305	0.355	0.5134	0.0103	0.2619
	theta	-0.3309	-0.384	0.4095	0.6438	-0.0397	0.0569
	alpha	-0.3289	-0.2771	0.4425	0.6549	0.0255	-0.2625
	beta	-0.5119	-0.1063	0.4952	0.3784	0.2077	-0.3599
	gamma	-0.6706	0.1094	0.4881	0.0588	0.2694	-0.2786
F_z	**delta**	-0.6035	-0.4614	-0.1269	-0.2895	-0.1911	-0.2703
	theta	-0.3918	-0.5648	-0.4679	-0.114	-0.076	-0.3507
	alpha	-0.4978	-0.1078	-0.5945	0.3151	0.2686	-0.0735
	beta	-0.8873	0.2272	-0.1378	0.0452	0.1014	0.0789
	gamma	-0.9107	0.2738	0.1111	-0.1049	0.0996	0.0166
F_7	**delta**	-0.7248	-0.4215	0.0103	-0.2809	0.0081	-0.1126
	theta	-0.4696	-0.6763	-0.3183	-0.1975	0.0708	-0.2488
	alpha	-0.7304	-0.2067	-0.4707	0.1129	0.1365	0.0058
	beta	-0.9062	0.2418	-0.0412	0.0034	0.0692	0.104
	gamma	-0.8862	0.2518	0.1427	-0.0794	0.1388	0.0524

Table C.2—Continued

P_z	**delta**	-0.6402	-0.08	0.1507	-0.1932	-0.5275	-0.0622
	theta	-0.4799	-0.2524	-0.2622	0.3264	-0.4346	-0.0246
	alpha	-0.4693	-0.0026	-0.5046	0.4817	0.1778	0.0789
	beta	-0.9078	0.2354	-0.1689	0.0688	0.0452	0.0739
	gamma	-0.911	0.3084	0.0667	-0.1007	0.0427	0.0329
T_5	**delta**	-0.6846	0.1718	0.2425	-0.1408	-0.3044	-0.051
	theta	-0.7532	0.1347	-0.1896	0.0684	-0.1755	-0.1948
	alpha	-0.7807	0.2635	-0.3097	0.2122	0.0648	0.0046
	beta	-0.9077	0.3354	0.0736	-0.097	0.0588	0.045
	gamma	-0.8565	0.3032	0.189	-0.1725	0.0366	-0.0144
O_2	**delta**	-0.6391	-0.0012	0.2664	-0.0898	-0.5272	0.0437
	theta	-0.5422	0.0021	-0.1065	0.3426	-0.4705	0.1319
	alpha	-0.6947	0.2539	-0.3355	0.2498	-0.0538	0.1979
	beta	-0.8691	0.267	-0.0338	-0.0628	-0.0304	0.0687
	gamma	-0.8201	0.2839	0.0569	-0.1676	-0.008	0.0435

Table C.3 Principle component factor pattern of F01

		P1	P2	P3	P4	P5	P6
VEOG	**delta**	0.6481	0.4212	-0.3956	0.2711	-0.2277	0.1072
	theta	0.6769	0.4747	-0.4191	0.156	-0.093	0.166
	alpha	0.6629	0.4442	-0.4431	0.1988	-0.1474	0.1474
	beta	0.7153	0.4356	-0.3604	0.0943	-0.1592	-0.0201
	gamma	0.7047	0.3298	-0.2268	-0.0884	-0.1227	-0.2113
HEOG	**delta**	0.4053	0.4785	-0.007	-0.2073	0.5887	0.0549
	theta	0.3171	0.4229	-0.0119	-0.2102	0.7706	0.1341
	alpha	0.327	0.401	0.0166	-0.2174	0.7216	0.0349
	beta	0.7429	0.0939	-0.0473	-0.2071	0.2727	-0.3319
	gamma	0.8429	-0.1172	-0.0888	-0.1152	0.1073	-0.3839
F_z	**delta**	0.6776	0.2758	0.1412	-0.0529	-0.2137	0.1642
	theta	0.7061	0.1645	0.0022	0.0995	-0.1612	0.3521
	alpha	0.5967	0.0709	0.2467	0.6371	-0.0346	-0.0227
	beta	0.9084	-0.1756	-0.129	-0.0609	-0.1039	-0.1901
	gamma	0.8487	-0.3597	-0.1385	-0.1057	-0.0291	-0.2219
F_7	**delta**	0.792	0.0362	0.0809	-0.0665	-0.0444	-0.198
	theta	0.8477	0.1565	-0.18	0.1385	-0.0355	0.0885
	alpha	0.7383	0.1694	-0.0532	0.4833	0.0014	-0.0765
	beta	0.8639	-0.11	-0.2735	-0.0184	-0.0179	-0.0425
	gamma	0.8661	-0.2598	-0.2619	-0.0743	-0.0073	-0.078

Table C.3—Continued

P_z	**delta**	0.5817	0.3189	0.4502	-0.3692	-0.3161	-0.0487
	theta	0.5812	-0.005	0.3404	-0.0409	0.0119	0.2805
	alpha	0.3621	-0.1135	0.4305	0.6667	0.2255	-0.1346
	beta	0.8372	-0.4329	-0.0042	-0.0682	0.1052	-0.1091
	gamma	0.7565	-0.5424	-0.0852	-0.0934	0.0703	-0.0969
T_5	**delta**	0.5121	0.3098	0.5481	-0.4078	-0.2885	-0.1013
	theta	0.6887	-0.0558	0.4622	-0.1293	-0.0936	0.0457
	alpha	0.4858	-0.068	0.5637	0.483	0.1444	-0.1009
	beta	0.6636	-0.559	-0.0096	-0.1271	0.1269	0.1912
	gamma	0.5974	-0.6202	-0.1568	-0.1055	0.1261	0.2266
O_2	**delta**	0.4669	0.3024	0.5679	-0.383	-0.2818	-0.1089
	theta	0.3816	0.222	0.4075	-0.2627	-0.013	0.3271
	alpha	0.2958	-0.0041	0.5662	0.5619	0.1806	0.0296
	beta	0.7242	-0.4314	0.0924	-0.0718	0.0529	0.3149
	gamma	0.6222	-0.5409	-0.0505	-0.0899	-0.0604	0.3376

REFERENCES

Alcaraz, R., & Rieta, J. J. (2008). Wavelet bidomain sample entropy analysis to predict spontaneous termination of atrial fibrillation. *Physiological Measurement, 29*(1), 65-80.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology, 42*(3), 361-377.

Bruzzo, A. A., Gesierich, B., Santi, M., Tassinari, C. A., Birbaumer, N., & Rubboli, G. (2008). Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients. A preliminary study. *Neurological Sciences, 29*(1), 3-9.

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

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology, 56*(2), 171-184.

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology, 31*(2), 129-145.

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & 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. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4*, 18.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (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., & 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.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation Space and Environmental Medicine, 69*(4), 360-367.

He, P., Wilson, G., Russell, C., & 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 & Biological Engineering & Computing, 45*(5), 495-503.

Hyvarinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks, 13*(4-5), 411-430.

Inouye, T., Shinosaki, K., Iyama, A., Matsumoto, Y., Toi, S., & 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., Yoshimine, T., et al. (1999). Medial prefrontal cortex generates frontal midline theta rhythm. *Neuroreport, 10*(4), 675-679.

Jackson, J. E. (1980). PRINCIPAL COMPONENTS AND FACTOR-ANALYSIS .1. PRINCIPAL COMPONENTS. *Journal of Quality Technology, 12*(4), 201-213.

Jung, T. P., Makeig, S., Stensmo, M., & Sejnowski, T. J. (1997). Estimating alertness from the EEG power spectrum. *Ieee Transactions on Biomedical Engineering, 44*(1), 60-69.

Klimesch, W., Schmike, H., & Pfurtscheller, G. (1993). Alpha frequency, cognitive load, and memory performance. *Brain Topography, 5*, 10.

Lowry, C. A., & Montgomery, D. C. (1995). A Review of Multivariate Control Charts. *IIE Transactions, 27*, 10.

Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A MULTIVARIATE EXPONENTIALLY WEIGHTED MOVING AVERAGE CONTROL CHART. *Technometrics, 34*(1), 46-53.

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. (1996). Independent Component Analysis of Electroencaphalographic Data. *Advances in Neural Information Processing Systems 8*, 6.

Makeig, S., & Inlow, M. (1993). LAPSES IN ALERTNESS - COHERENCE OF FLUCTUATIONS IN PERFORMANCE AND EEG SPECTRUM. *Electroencephalography and Clinical Neurophysiology, 86*(1), 23-35.

Mastrangelo, C. M., Runger, G. C., & Montgomery, D. C. (1996). Statistical process monitoring with principal components. *Quality and Reliability Engineering International, 12*(3), 203-210.

Montgomery, D. (2009). *Statistical Quality Control* (6th edition ed.): John Wiley & Sons, Inc.

Montgomery, D. C., Jennings, C. L., & Kulachi, M. (2008). *Introduction to Time Series Analysis and Forecasting*. Hoboken: John Wiley & Sons.

Montgomery, D. C., & Mastrangelo, C. M. (1991). SOME STATISTICAL PROCESS-CONTROL METHODS FOR AUTOCORRELATED DATA. *Journal of Quality Technology, 23*(3), 179-193.

Phillips, C., Parr, J., & Riskin, E. (2008). *Signals, Systems, and Transforms* (4 ed.). Upper Saddle River: Prentice Hall.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). BIOCYBERNETIC SYSTEM EVALUATES INDEXES OF OPERATOR ENGAGEMENT IN AUTOMATED TASK. *Biological Psychology, 40*(1-2), 187-195.

Prinzel, L. J., Freeman, F. C., Scerbo, M. W., Mikulka, P. J., & 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.

Quiroga, R. Q., Arnhold, J., Lehnertz, K., & Grassberger, P. (2000). Kulback-Leibler and renormalized entropies: Applications to electroencephalograms of epilepsy patients. *Physical Review E, 62*(6), 8380-8386.

Richman, J. S., Lake, D. E., & Moorman, J. R. (2004). Sample entropy *Numerical Computer Methods, Pt E* (Vol. 384, pp. 172-184).

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology, 278*(6), H2039-H2049.

Scranton, R., Runger, G. C., Montgomery, D. C., & Keats, B. J. (1996). Efficient Shift Detection Using Multivariate Exponentially-Weighted Moving Average Control Charts and Principle Components. *Quality and Reliability Engineering International, 12*, 6.

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

Stone, J. V. (2002). Independent component analysis: an introduction. [Review]. *Trends in Cognitive Sciences, 6*(2), 59-64.

Tou, J., & Gonzalez, R. (1974). *Pattern Recognition Principles*. Reading, MA: Addison-Wesley.

Vigario, R., Sarela, J., Jousmaki, V., Hamalainen, M., & Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *Ieee Transactions on Biomedical Engineering, 47*(5), 589-593.

Wilson, G. F., & Fisher, F. (1991). THE USE OF CARDIAC AND EYE BLINK MEASURES TO DETERMINE FLIGHT SEGMENT IN F4-CREWS. *Aviation Space and Environmental Medicine, 62*(10), 959-962.

Wilson, G. F., & Fisher, F. (1995). COGNITIVE TASK CLASSIFICATION BASED UPON TOPOGRAPHIC EEG DATA. [Article]. *Biological Psychology, 40*(1-2), 239-250.

Wilson, G. F., & Russell, C. (1999). *Operator Functional State Classification Using Neural Networks with Combined Physiological and Performance Features*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting.

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

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

Yu, Zhaohan. (2009). *OPTIMIZATION TECHNIQUES IN DATA MINING WITH APPLICATIONS TO BIOMEDICAL AND PSYCHOPHYSIOLOGICAL DATA SETS.* University of Iowa, Iowa City.

Zhang, D. D., Ding, H. Y., Liu, Y. F., Zhou, C., Ding, H. S., & Ye, D. T. (2009). Neurodevelopment in newborns: a sample entropy analysis of electroencephalogram. *Physiological Measurement, 30*(5), 491-504.