

Transforming innovations into applications.



EEG-based Estimation and Classification of Mental Fatigue

Leonard J. Trejo ^{*a}, Karla Kubitz ^b, Roman Rosipal ^{a,c}

Rebekah L. Kochavi ^d, Bryan L. Matthews ^d, Leslie D. Montgomery ^e

^a Pacific Development and Technology, LLC, 999 Commercial St. Ste. 205, Palo Alto, CA 94303

^b Towson University, Towson, MD

^c Medical University of Vienna, Vienna, Austria

^d Stinger Ghaffarian Technologies, Inc.

^e LDM Associates, San Jose, CA

PDT Technical Report No. QGPBF0636B1403

Published online October 31, 2006

Revised January 4, 2009

Copyright © 2006, 2009 Pacific Development and Technology, LLC

* ltrejo@pacdel.com; phone 1-650-320-8841, fax 1-650-230-1356; www.pacdel.com

Abstract

Mental fatigue was associated with increased power in frontal theta (θ) and parietal alpha (α) EEG rhythms. A statistical classifier can use these effects to model EEG-fatigue relationships accurately. Participants ($n=22$) solved math problems on a computer until either they felt exhausted or 3 h had elapsed. Pre- and post-task mood scales showed that fatigue increased and energy decreased. Mean response times rose from 6.7 s to 7.9 s but accuracy did not change significantly. Mean power spectral densities or PSDs of θ and α bands rose by 29% and 44%, respectively. A kernel partial least squares classifier trained to classify PSD coefficients (1-18 Hz) of single 13-s EEG segments from alert or fatigued task periods was 91% to 100% accurate. For EEG segments from other task periods, the classifier outputs tracked the time course of the development of mental fatigue. By this measure, most subjects became fatigued after 45-60 min of task performance.

EEG, mental fatigue, alertness, situation awareness, performance monitoring, KPLS, machine learning, classification

EEG-based Estimation and Classification of Mental Fatigue

The purpose of this study was two-fold: a) to identify the features of the electroencephalogram or EEG that change in an orderly and reliable manner with the development of mental fatigue, and b) to construct a practical system that can use EEG features to accurately estimate the instantaneous degree of mental fatigue over a performance period of up to three hours. Our approach consisted of five key steps. First, we adopted a clinical-theoretical framework that allowed us to distinguish *mental* fatigue from general fatigue or other factors that influence task performance. Second, we adopted a task and a set of procedures for inducing mental fatigue in the participants of the study. Third, we adopted psychometric tests that can independently validate the ability of the task to induce mental fatigue. Fourth, we selected an EEG measurement system that is practical enough to use in operational settings and sensitive enough to detect changes in EEG features with the development of fatigue. Finally, we adapted an algorithm for estimation of mental fatigue from EEG features that is robust in the face of signal corruption from noise or sensor loss and accurate enough to classify states of mental fatigue in individual human subjects with a few seconds of EEG recordings.

There are countless occupations today in which fatigued individuals routinely operate complex, high-risk systems. Such systems may be as common as driving a car or operating a forklift, or as rare as controlling traffic at a busy airport or docking two spacecraft in orbit. This undesirable state of affairs has contributed to many well publicized—and some not so well publicized—disasters. For example, operator fatigue was involved in the Three-Mile Island and Chernobyl nuclear power plant disasters, the EXXON Valdez grounding, and the Guantanamo Bay airline crash (U.S. Congress, Office of Technology Assessment, 1991; Fatigue Resource Directory, 2000; Dingess, 1995). “There is [also] much evidence that [operator] fatigue has

contributed to serious incidents and accidents in industrial operations ... and to accidents in *all* [emphasis added] modes of transportation” (Dinges, 1995, p. 8). Analyses of crash data confirm that fatigue, drowsiness and inattention pose the greatest known risks to automobile driver and passenger safety, surpassing all other known risks including alcohol and secondary tasks such as cell phone usage (Dingus, et al., 2006). Accordingly, there continues to be much scientific interest in assessing, monitoring, and predicting operator fatigue (Gevins et al., 1995; Kennedy, 1953; Wilson & Fisher, 1991; 1995; Russo, Stetz, & Thomas, 2005).

Unfortunately, there is no standard definition of fatigue. Clinically, fatigue is best defined as difficulty in performing voluntary activities (Chaudhuri & Behan, 2004). The main psychological symptoms are subjective feelings of low energy and motivation, which may be accompanied by physical weakness. We prefer the point of view that fatigue is distinct from drowsiness, or the feeling of a need to sleep, although fatigue and drowsiness are often felt at the same time (Hurd, 2007). Drowsiness aside, fatigue is still a broad concept, which includes mental, motor, metabolic, endocrine, sensory, and other physiological or environmental components (Chaudhuri & Behan, 2004, Winningham, 1996).

In the present study, we will attempt to experimentally isolate *mental fatigue* from general fatigue because complex, high-risk tasks such as piloting aircraft and vehicles impose minimal physical demands and heavy cognitive demands on human operators. In such tasks, mental fatigue jeopardizes performance and safety through a variety of cognitive impairments, including: a) decreased vigilance (Mast & Heimstra, 1964; Šipoš, 1971; Smit, Eling, & Coenen, 2004), b) decreased secondary task performance (van der Hulst, Meijman, & Rothengatter, 2001), c) impaired problem solving (Horne, 1988; Van Der Linden, Frese, & Meijman, 2003; Van Der Linden, Frese, & Sonnentag, 2003), d) reduced situation awareness (Vidulich, Stratton,

Crabtree, & Wilson, 1994), and e) decreased tolerance of risk and effort (Barth, Holding & Stamford, 1976; Brown, Tickner, & Simmonds, 1970; Neri, Shappell, & DeJohn, 1992; Shingledecker & Holding, 1974).

As with all mental states, clinicians and neuroscientists cannot sharply define mental fatigue in terms of specific physiological processes. So mental fatigue is usually defined as a subjective assessment that is associated with various behavioral and physiological symptoms. For example, we can define mental fatigue as the unwillingness of alert, motivated subjects to continue performing a specific mental task (Montgomery, Montgomery, & Guisado, 1995). Then at times before, during, or after performance of the task, we can validate this definition with psychometric tests.

Although mental fatigue is currently assessed subjectively, there is a pressing practical need for objective measures to estimate, monitor, and predict it. Changes in the EEG clearly signal the onset and stages of sleep, but more recent studies have shown that EEG patterns can also signal the onset of mental fatigue in awake persons (Shen, Li, Ong et al., 2008). However, despite many studies to date, the EEG signs of fatigue are still not sharply defined and using EEG for operational fatigue estimation is still a formidable challenge.

This state of affairs motivated us to rigorously test the accuracy and robustness of EEG-based estimation of mental fatigue instead of further analyzing its underlying physiology. We used a two-pronged approach: first we assessed the common across-subjects pattern of changes in power spectral density of multichannel EEG over extended periods of mental arithmetic performance. Second, we tested the feasibility of using EEG power spectral densities or *PSDs* to classify mental fatigue states at a sub-minute time scale in individual subjects. More specifically,

we trained, cross-validated, and tested statistical learning algorithms for classifying brief segments of EEG recordings as pertaining to self-reported states of alertness or mental fatigue.

To assess EEG changes common to all subjects, we tested the hypothesis that the PSDs of theta, alpha, and other EEG bands would co-vary regularly with the degree of *mental* fatigue that subjects experienced. While many studies have considered EEG correlates of general fatigue, drowsiness and sleepiness, our study focused on inducing mental fatigue through sustained performance of a cognitive task for up to three hours, while minimizing other effects. Prior studies led us to expect significant fatigue-related changes in PSDs of the frontal midline theta band and the parietal alpha band (e.g., Gevins et al., 1998). So we specified null hypotheses that EEG power in specific theta and alpha bands would remain constant over the course of a fatigue-inducing task.

To test the feasibility of using EEG to monitor mental fatigue in individuals, we trained a supervised machine learning algorithm to extract information-rich features from brief segments of EEG recordings and classify them as pertaining to states of alertness or mental fatigue. The algorithm is known as *kernel partial least squares* or *KPLS*, which we have applied successfully to classification of a wide range of data types including linear and nonlinear classification of EEG and event-related potentials (Rosipal & Trejo, 2001; Rosipal, Girolami, Trejo, et al., 2001; Wallerius, Trejo, Matthews, et al., 2005; Trejo, Matthews & Rosipal, 2006). The features we extract using KPLS are a set of orthogonal basis vectors, which exhibit high covariance with states of alertness or fatigue. Except for this covariance constraint the basis vectors are computed using methods similar to those used in principal components analysis or factor analysis. We find the necessary and sufficient set of basis vectors using algorithms (de Jong, 1993; Rosipal, Trejo & Matthews, 2003) that extract each vector sequentially, in descending rank of its covariance

with an output measure, which in this case is a distinction between alertness and fatigue. Each extracted basis vector is a unique linear combination of multichannel EEG spectral densities. We can then iterate this feature extraction procedure with cross validation methods to optimize the number of extracted components by striking a robust balance of classification accuracy and generality. We apply the KPLS algorithm in this study to accurately and robustly estimate mental fatigue in the framework of developing a useful system for continuous monitoring of mental fatigue under working conditions.

Methods

Experimental Design

We tested several hypotheses about the association of subjective moods, observed behavior, performance, and physiological measures during continuous performance of mental arithmetic for up to three hours. We manipulated a single factor, that is, time on task, and used a repeated measures design. Subjective moods were indexed by the Activation Deactivation Adjective Checklist (AD-ACL; Thayer, 1986: 1989) and the Visual Analogue Mood Scales (VAMS; Stern, 1997) questionnaires. Observed behavior included ratings of activity and alertness from videotaped recordings of each participant's performance. The performance measures were response time and response accuracy, recorded in real time by the task computer. The physiological measures were derived from spontaneous EEGs, including: a) theta activity (4-8 Hz) at Fz (both average power and peak amplitude in the theta band) and b) alpha activity (8-13 Hz) at Pz (both average power and peak amplitude in the alpha band).

Participants

We collected data from 33 individuals from the NASA Ames Research Center community. All participants gave their informed consent using a form approved by the NASA

Ames Research Center Institutional Review Board and were paid for their participation. We report only the data from the 22 participants who met the following inclusion criteria: they had to have a) been able to wear an EEG electrode cap without significant discomfort, b) stayed awake for at least 90 minutes, c) provided stable EEG recordings within a tolerance of $\pm 100 \mu\text{V}$. We dismissed participants who failed to meet these criteria. Six of the 22 eligible participants were excluded from analyses. One was excluded because he violated experimental protocol and wore a watch during the testing, and attended to the watch repeatedly. Five more were excluded because their response times were extremely slow (and consequently they provided too few EEG epochs for analysis). The remaining 16 participants included 12 males and 4 females with a mean age of 26.9 ($SD = 7.4$) years. Also, according to self-reports, all of the participants had normal vision and hearing. Fourteen of the 16 participants reported being right-handed and two reported being left-handed.

Mental Arithmetic Task

Participants sat in front of a computer with their right hands resting on a 4-button keypad (Neuroscan STIM pad, Compumedics USA, El Paso, TX). Arithmetic summation problems, consisting of four randomly generated single digits, three operators, and a target sum (e.g., $4 + 7 - 5 + 2 = 8$), were displayed on a computer monitor (cathode ray tube) continuously until the subject responded (Fig. 1). Only addition and subtraction were used, and equations for which answers were obvious (such as those including several repeated digits) were excluded. The participants: a) solved the problems, b) decided whether their ‘calculated sums’ were less than, equal to, or greater than the target sums provided, c) indicated their decisions by pressing the appropriate key on the keypad. The keypad buttons were labeled “<,” “=,” and “>,” respectively. Subjects were instructed to answer as quickly as possible without sacrificing accuracy. After a

response, there was a 1 s inter-trial interval, during which the monitor was blank. Participants performed the task until either they quit from exhaustion or three hours had elapsed. As can be seen in Table 1, all participants completed at least six of the twelve 15-min blocks and eleven participants completed all twelve of the 15-min blocks (i.e., the total 3-hour period).

Activation Deactivation Adjective Checklist

Thayer's AD-ACL (Thayer, 1986; 1989) is a multi-dimensional checklist reflecting perceptions of activation. Individuals respond to 20 items using a 4-point rating scale (definitely feel, feel slightly, cannot decide, and definitely do not feel). The scoring procedure includes four subscales: energy (reflects general activation), tiredness (reflects general deactivation), tension (reflects high preparatory arousal), and calmness (reflects low preparatory arousal). The AD-ACL is a reliable and valid subjective method (Thayer, 1986; 1989).

Visual Analogue Mood Scales

The VAMS (Stern, 1997) measure eight specific mood states, including afraid, confused, sad, angry, energetic, tired, happy, and tense. The VAMS have a neutral schematic. That is, they have a "mood-neutral" face (and word) at the top of a 100 mm vertical line and they have a "mood-specific" face (and word) at the bottom of the line. Individuals mark the point along the line that best illustrates how they feel at present. Scores range from 0 to 100, with 100 indicating the maximum level of the mood and 0 indicating the minimum level of a mood. Like the AD-ACL, the VAMS are also reliable and valid (Nyenhuis, Stern, Yamamoto, Luchetta, & Arruda, 1997; Stern, 1997).

Observed Activity and Alertness

Activity and alertness were measured by visual inspection of videotapes of each participant's performance. The videotapes showed combined overall scene- and facial views of

the participants. For each 15-min interval, a single rater judged levels of alertness and activity (unnecessary motion) on a five-point scale (Table 2). The alertness scale considered the subjective appearance of sleep, sleepiness, dozing off, distraction, yawning, and general alertness. The activity scale considered the frequency of subject motion, including moving in the chair, fidgeting, tapping, or shaking. The rater was blind to the specific aims of the study, although not to the time interval being rated (as the videotaped segments were viewed sequentially). However, changes in activity and alertness were related to obvious changes in behavior and, as a result, were fairly easy to quantify. Moreover, a second rater randomly rated a subset of the videotapes and found good agreement with the ratings. Ratings of activity and alertness were tested for correlations with response time, accuracy, and EEG spectral measures.

EEG Activity

EEG was recorded continuously using 32 Ag/AgCl electrodes embedded in an elastic fabric cap (i.e., a Quik-CapTM, Compumedics USA, El Paso, TX). The electrode cap was placed on the participant according to the manufacturer's instructions. The reference electrodes were averaged mastoids and the ground electrode was located at AFz. Vertical and horizontal electrooculograms (VEOG and HEOG) were recorded using bipolar pairs of 10 mm Ag/AgCl electrodes (i.e., one pair superior and inferior to the left eye; another pair to the right and to the left of the orbital fossi). Impedances were maintained at less than 5 k Ω for EEG electrodes and 10 k Ω for EOG electrodes. The EEG was amplified and digitized with a 64-channel Neuroscan SynampsTM system (Compumedics USA, El Paso, TX), with a gain of 1,000, sampling rate of 500 s⁻¹ and a pass band of 0.1 to 100 Hz. Amplifiers were calibrated with a 50 μ V signal prior to each testing session. The signals were digitized and stored on hard disk drives by a computer

equipped with Neuroscan Scan 4.2 software (Compumedics USA, El Paso, TX) and archived on optical media (CD-R).

Procedures

Participants: a) were given an orientation to the study, b) read and signed an informed consent document, c) completed a brief demographic questionnaire (age, handedness, hours of sleep, etc.), d) practiced the mental arithmetic task for 10 minutes, and e) were prepared for data collection by having the electrode cap, EOG, and reference electrodes applied. They then completed the pretest self-report measures (i.e., the AD-ACL and VAMS) and performed the mental arithmetic task until either three hours had elapsed or volitional exhaustion had occurred. Task termination was followed by the completion of post-test self-report measures and participant debriefing.

Data Processing

The EEGs, initially processed using Neuroscan Scan 4.2 Edit™ (Compumedics USA) software, were: a) submitted to an algorithm for the detection and elimination of eye-movement artifact, b) visually examined and blocks of data containing artifact were manually rejected, c) epoched around the stimulus (i.e., from -5 s pre-stimulus to +8 s post-stimulus), d) low pass filtered (50 Hz; zero phase shift; 12 dB/octave roll off), and e) submitted to an automated artifact rejection procedure (i.e., absolute voltages > 100 μ V). The overall single-epoch rejection rate was 47%. The 'cleaned and filtered' epochs were decimated to a sampling rate of 128 Hz. EOG artifact was removed by using wavelet-denoised VEOG and HEOG signals as predictors of the artifact voltages at each EEG electrode in a multivariate linear regression. The ocular artifact rejection algorithm consisted of four main steps: a) wavelet smoothing of the EOG channels, b) blink detection, c) computing the linear regression of the EEG on the EOG channels, and

d) subtracting the resulting weighted EOG from the EEG channels. These steps were repeated for each single EEG epoch. The wavelet smoothing process decomposed the EOG channels by using a Daubechies (DWT-8) wavelet filter. The lowest 2% of the coefficients were retained and the signal was reconstructed from the retained coefficients (e.g., Trejo & Shensa, 1999). The voltage of each sample in the VEOG signal was then compared to an absolute voltage threshold set at 100 μ V and marked as a blink or non-blink accordingly. A multivariate linear regression algorithm was used to calculate the regression weights for the EEG on the VEOG and HEOG channels. Separate regression weights were estimated for blink and non-blink epochs. The final step was a linear subtraction of the weighted VEOG and HEOG signals from each EEG channel. EEG power spectra were estimated with Welch's periodogram method at 833 frequencies from 0-64 Hz. Peak and average power in the theta and alpha bands were measured at electrodes Fz and Pz, respectively.

Classification Procedures

We classified single EEG epochs using kernel partial least squares decomposition of multichannel EEG spectra coupled with a discrete-output linear regression classifier, or *KPLS-DLR*. Through extensive side-by-side testing of EEG data, we found that *KPLS-DLR* is just as accurate as *KPLS-SVC*, which uses a support vector classifier for the classification step (Rosipal, Trejo, & Matthews, 2003). *KPLS* selects the reduced set of orthogonal basis vectors or *components* in the space of the input variables (EEG spectra) that maximizes covariance with the experimental conditions. *DLR* finds the linear hyperplane in the space of *KPLS* components that separates the classes. In a pilot study (Montgomery, Montgomery & Guisado, 1995) and in the present study, we found that the first 15 minutes on task did not produce mental fatigue, whereas mental fatigue was substantial in the final 15 minutes. So we randomly split EEG epochs from

the first and last 15-min periods into equal-sized training and testing partitions for classifier estimation. Only the training partition was used to build the final models. The number of KPLS components in the final models was set by five-fold cross-validation. The criterion for KPLS-DLR model selection was the minimum classification error rate summed over all (five) cross-validation subsets.

Statistical Analyses

The data were analyzed using either singly or, when appropriate, doubly multivariate repeated measures analyses of variance with time of measurement as a within-subjects factor. Doubly multivariate analyses were carried out for AD-ACL subscale scores (energy, tension, calmness, and tiredness), VAMS subscale scores (afraid, confused, sad, angry, energetic, tired, happy, and tense), behavioral observation data (observed activity and alertness), theta activity data (peak and band-average amplitudes), and alpha activity data (peak and band-average amplitudes). Singly multivariate analyses of variance were carried out for response time and accuracy. For the doubly multivariate analyses, significant multivariate F-ratios were decomposed using single degree of freedom, within-subjects contrasts. For the singly multivariate analyses, Huynh-Feldt-corrected degrees of freedom and p-values were reported (i.e., because of sphericity). In both analyses, partial η^2 values were reported as effect size estimators.

RESULTS

Self-report Analyses

The AD-ACL subscale scores were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., pretest vs. posttest) as a within-subjects factor. The main effect of time of measurement was significant ($F(4, 5) = 10.4, p < .01, \eta^2 = .89$). Within-subjects contrasts

showed significant linear trends for energy ($F(1, 8) = 6.46, p < .04, \eta^2 = .45$), calmness ($F(1, 8) = 21.3, p < .002, \eta^2 = .73$), and tiredness ($F(1, 8) = 6.38, p < .04, \eta^2 = .44$). Energy decreased from a pretest mean of 12.0 (standard deviation (SD) = 4.1) to a posttest mean of 8.6 (SD = 3.7). Calmness decreased from a pretest mean of 16.8 (SD = 1.6) to a posttest mean of 14.1 (SD = 1.8). Tiredness increased from a pretest mean of 10.1 (SD = 4.3) to a posttest mean of 15.3 (SD = 5.7). There was also a non-significant linear trend for tension ($F(1, 8) = .92, p < .37$). Thus, the AD-ACL data indicate that our manipulation decreased general activation (i.e., self-reported energy) and preparatory arousal (i.e., self-reported calmness) and increased general deactivation (i.e., self-reported tiredness).

The VAMS subscale scores (i.e., for afraid, confused, sad, angry, energetic, tired, happy, and tense) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., pretest vs. posttest) as a within-subjects factor. The main effect of time of measurement was non-significant (multivariate $F(8, 1) = 1.31, p = .59$). This analysis suggests that our manipulation, despite its effects on activation and arousal, did not influence other, more durable, moods.

Behavior Analyses

The behavioral observations (i.e., observed activity and alertness) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., ten 15-min periods) as a within-subjects factor. The main effect of time of measurement was significant ($F(18, 178) = 3.70, p < .0005, \eta^2 = .27$). This analysis suggests that time on task influenced behavior (i.e., observed activity and alertness levels). Moreover, time on task had a progressive effect on behavior. Within-subjects contrasts showed a linear decrease in alertness ($F(1, 10) = 10.4, p < .009, \eta^2 = .51$) and a linear increase in activity ($F(1, 10) = 5.88, p < .04, \eta^2 = .51$). Alertness decreased from a mean of 5.00 (SD = 0.00) in the first 15-min period to a mean of 2.43 (SD = 0.98) in the

last 15-min period. Activity increased from a mean of 1.36 (SD = 0.51) to a mean of 2.45 (SD = 1.30), respectively.

The response times (RT) were analyzed in an ANOVA with time of measurement (i.e., 15-min periods) as a within-subjects factor. The main effect of time of measurement was significant (Huynh-Feldt corrected $F(3, 39) = 3.78, p < .03, \eta^2 = .24$). This analysis suggests that time on task influenced performance. Moreover, time on task had a progressive effect on performance (Fig. 2). Within-subjects contrasts showed a significant linear increase in RT ($F(1, 12) = 8.29, p < .01, \eta^2 = .41$) rising from a mean of 6.70 s (SD = 2.18) in the first 15-min period to a mean of 7.87 s (SD = 2.64) in the last 15-min period. We found the same pattern of significant effects for RT analyzed in an ANOVA with fraction of artifact-free trials (i.e., 1st100, middle 100, and last 100) as a within-subjects factor.

Response accuracy was analyzed in an ANOVA with time of measurement (i.e., ten 15-min periods) as a within-subjects factor. The main effect of time of measurement was not significant (Huynh-Feldt corrected $F(5, 43) = 1.74, p = .14$). Response accuracy was also analyzed in an ANOVA with fraction of artifact-free trials (i.e., 1st100, middle 100, and last 100) as a within-subjects factor. The main effect of number of trials was not significant (Huynh-Feldt corrected $F(2, 19) = 2.84, p = .09$). This analysis suggests that, despite its effects on other aspects of behavior, time on task did not have a substantial influence on response accuracy.

EEG Analyses

Because the small number of participants precluded analysis of the full montage of electrodes, we examined the average spectra for pronounced effects (Fig. 3) and focused our analyses on the electrodes that had maximal average voltage for the alpha and theta band (i.e., Fz for theta and Pz for alpha).

The changes in frontal midline theta (i.e., average PSDs and peak amplitudes at Fz) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., 15 min-periods) as a within-subjects factor. The main effect of time of measurement was significant (multivariate $F(18, 178) = 2.05, p < .01, \eta^2 = .17$). Average power in the theta band increased from a mean of $199.36 \mu\text{V}^2/\text{Hz}$ (SD = 97.50) in the first 15-min period to a mean of $256.58 \mu\text{V}^2/\text{Hz}$ (SD = 135.57) in the last 15-min period. Peak amplitude in the theta band increased from a mean of $272.4 \mu\text{V}^2/\text{Hz}$ (SD = 146.0) in the first 15 min period to a mean of $390.8 \mu\text{V}^2/\text{Hz}$ (SD = 227.1) in the last 15-min period. This analysis suggests that theta increased with time on task. Moreover, this analysis suggests that time on task had a progressive effect on frontal midline theta activity. Within-subjects contrasts showed significant linear increases in average theta PSDs ($F(1, 10) = 7.42, p < .01, \eta^2 = .48$) and in peak theta amplitudes ($F(1, 10) = 9.31, p < .01, \eta^2 = .48$).

The changes in midline parietal alpha activity (i.e., average PSDs and peak amplitudes at Pz) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., 15-min periods) as a within-subjects factor. The main effect of time of measurement was significant (multivariate $F(18, 178) = 2.20, p < .005, \eta^2 = .18$). Average alpha PSDs increased from a mean of $307.4 \mu\text{V}^2/\text{Hz}$ (SD = 434.3) in the first 15-min period to a mean of $459.0 \mu\text{V}^2/\text{Hz}$ (SD = 593.9) in the last 15-min period. This analysis suggests that alpha increased progressively with time on task. Within-subjects contrasts showed significant linear increases in average alpha PSDs ($F(1, 10) = 6.07, p < .03, \eta^2 = .38$). Peak alpha amplitudes increased and trended similarly, but not significantly so ($F(1, 10) = 4.11, p = .07$).

Classification

We applied our classification procedure to EEG recordings from 14 of the 16 subjects (two subjects had too few EEG epochs for model estimation). The EEG epochs were synchronized with the onset of each math problem, extending from -5 s to $+8$ s relative to each stimulus onset. As such there was some overlap among the EEG segments. However a second analysis of 3.5 s segments with no overlap produced highly similar results, so we will focus only on the long-epoch results. We also reduced the likelihood of electromyogram artifacts by low-pass filtering the EEG with an 18-Hz cutoff.

For each subject we constructed a KPLS-DLR model using either linear or Gaussian (nonlinear) kernels and selected the best model as described above. Results for Gaussian kernels were not superior, and on average linear kernels had slightly better results, so we focus on linear kernels here. Classification accuracies (Fig. 4) across both classes for 18-Hz filtered EEG ranged from 91.12 to 100% (mean $\bar{=}$ 98.30, Table 3). The corresponding range for 11-Hz filtered EEG was 89.53 to 98.89% (mean $\bar{=}$ 98.30%). The number of KPLS components ranged from 1 to 4 (mean 2.77) for 18-Hz EEG and from 1 to 5 (mean 3.76) for 11-Hz EEG (Table 3). With as few as two components, the separation of classes was usually evident from the distribution of KPLS scores for single EEG epochs. The test-set data for the first and last 15-min blocks occupied distinct regions in the space of the KPLS scores (Fig. 5).

The scalp topography of the KPLS weights can serve as an indicator of which regions or electrodes strongly influence classification. For example, by plotting the product of the KPLS model weights and the average EEG spectra in limited frequency bands in one subject (Fig. 6), we found that a broad set of frontocentral midline sites was important for classification in the theta band. In the alpha band, the discriminating electrodes were tightly concentrated over midline parietal site Pz.

KPLS Model Prediction

We also examined the predictive validity of the KPLS-DLR models by testing them with data from the first nine intervening 15-min periods (between first and last). The behavior of the classifiers for these periods was consistent with an orderly, progressive migration of single-trial KPLS predictions from the non-fatigued to the fatigued class. This observation agrees with the trends we observed in response times, EEG measures, and behavioral observations. We examined these patterns of migration by inspecting graphs of the predicted scores for the first two components of the KPLS models for single subjects (Fig. 7). Initially, the predicted points overlapped with the region occupied by the non-fatigued training set. Over time, the predicted points shifted towards the fatigue region. For each of 11 subjects, we computed the center of mass (mean) of the KPLS classification scores for each of 10 fifteen-minute blocks (Fig. 8). These means correspond to the centers of the clouds of points shown moving from the alert region to the fatigue region in Fig. 7. Negative scores represent alert states and positive scores represent fatigue states. For most subjects there is a progressive shift of the means toward the fatigue state, with some intermittent reversals. Three subjects entered fatigue as early as block 1 (15-30 min) whereas one subject did not enter fatigue until block 7 (105-120 min). The majority of subjects (n=6) entered fatigue by block 3 (45-60 min).

DISCUSSION

Our results showed that the power spectral densities of theta and alpha EEG bands co-varied with the degree of fatigue that subjects experienced. In contrast to our expectations based on prior studies (see Introduction), power in both theta and alpha bands increased between the beginning and end of the task. These increases were confirmed with behavioral and subjective reports as pertaining to alert and fatigued conditions. Moreover, we found that we could

accurately model and estimate fatigue as a function of EEG spectral features using KPLS-DLR, a statistical learning theory based classifier.

Behavioral Measures

Time on task produced decreased general activation (i.e., self-reported energy) and preparatory arousal (i.e., self-reported calmness) and increased general deactivation (i.e., self-reported tiredness) but did not significantly influence other moods. These effects support the assertion that our task produced a state of substantial mental fatigue. Observations of the participants showed that behavioral activity progressively increased while alertness progressively decreased over time. Moreover, there was a progressive slowing of response times. However, time on task did not significantly influence response accuracy. Together, these results suggest that our subjects experienced mental fatigue, but did not sacrifice accuracy as might be expected if motivation had waned. The moderate, general increases in RT over time also indicate increasing mental fatigue, but not a severe increase in RT, as might be expected if lapses or microsleep episodes had occurred frequently.

Fatigue due to prolonged vigilance or drowsiness appears to have different effects on error rates than cognitive fatigue. For example, in a simulated sonar task, an EEG-based alertness index followed closely the time course of detection error rates (Makeig & Inlow, 1993). However, we found that cognitively fatigued subjects continued to perform with low and stable error rates for up to three hours. Our task differed from the sonar task by virtue of having strong cognitive demands, whereas the demands of the sonar task were mainly sensory or perceptual (detecting sonar sounds). The main behavioral impact of *cognitive* fatigue in our arithmetic task appeared to be a slowing of mental processes, because response times trended significantly and progressively higher over time. Because the response demands of our task were simple and

highly practiced, we do not consider that cognitive fatigue substantially slowed response selection processes. Instead, our results are more consistent with the view that cognitive fatigue slows down central executive functions such as working memory and decision making as compared to perception and response-selection.

EEG Measures

Our EEG analyses indicate that time on task had a progressive influence on PSDs of frontal midline theta and parietal alpha activity. PSDs in both of these EEG bands increased as a function of time on task. Our inspection of the EEG spectra did not indicate effects outside the theta and alpha bands. In particular, there were no indications of effects at 14 Hz or in the beta band. Our results do not support an overall slowing of the EEG in mental fatigue, as much as they indicate specific increases in frontal midline theta and midline parietal alpha power. The lack of effects in bands above 14 Hz also discounts electromyographic artifacts or other broad-band noise sources as determinants of the theta and alpha effects.

The hypothesized fatigue-related increase in frontal midline theta band PSDs was supported by our analyses. One current hypothesis states that increases in theta band PSDs are associated with increases in the demands of cognitive processes (Klimesch, 1999, Pfurtshceller & Lopes de Silva, 2005). In our mental arithmetic task we consider that the critical cognitive processes are working memory, long-term memory, and selective attention. A large and diverse body of experimental reports has substantiated that working memory (WM) capacity and speed of access to long-term memory (LTM) are critical limiting processes in cognitive performance (Anderson, 1995). In particular, a theoretical framework has emerged, which relates alpha, theta, and beta activity in the anterior cingulate cortex and the hippocampus to WM and LTM (Gevins et al., 1997, Klimesch et al., 1997). Distinctions are now recognized among the spatial and

frequency parameters that characterize: diffuse theta rhythms associated with sleepiness, focused, rhythmic theta rhythms associated with cognitive effort and WM load, low alpha rhythms (8-10 Hz) associated with relaxed wakefulness, and focused high alpha rhythm (10-12 Hz) suppression associated with spatial and non-spatial shifts of attention (reviewed in Klimesch, 1999). These experimental and theoretical results are generally consistent with the theta band effects we observed here.

The unexpected increases in power of parietal midline alpha band power are not likely to be an artifact of our measurements and analyses, which were done as carefully and precisely as possible, including expert visual inspection of every EEG epoch analyzed. One possible explanation for this result concerns the timing of our EEG spectral measures. Our EEG epochs were synchronized with the onset of each stimulus presented in the mental arithmetic task. This means that the EEG features associated with cognitive processing for the task will tend to dominate the analysis, as compared to EEG features that are randomly selected over time. In this way our study differs from most others, in which EEG epochs were regularly sampled at a rate that did not match the appearance of task-relevant stimuli or events.

One possible explanation of the observed alpha band PSD effect derives from a neurophysiological model about the significance of varying alpha band power levels. The model states that “when patches of neurons display coherent activity in the alpha band, a depressed state of active processing of information in the underlying cortical neuronal populations can be assumed to exist” (Pfurtscheller & Lopes da Silva, 2005, pp 1008-1009). This neurophysiological model is consistent with a phenomenological model of mental fatigue in which the likelihood of depressed or reduced processing surrounding the onset of a task-relevant

stimulus increases as a function of time on task or in relation to subjectively perceived mental fatigue.

Yet another possibility is that mental fatigue may impair or slow the shifting of non-spatial selective attention, thereby producing less alpha rhythm suppression than that observed in states of normal or heightened alertness. Clearly, more work needs to be done to sort out the reasons for this unexpected, but significant effect of mental fatigue on alpha band power.

Classification and Prediction

Overall, our KPLS-DLR classification of single-trial EEG epochs ranged from 90% to 100% with a mean of 97% to 98%. The performance of these classifiers is highly accurate for single trials, and may serve as the basis for predictive models of mental fatigue in operational settings. Inspections of the predictive behavior of the KPLS models showed an orderly relationship between the scores and time on task as well as between the scores and correlated behavioral, subjective, and performance measures. In the 11 subjects that we examined for model predictions, our classifier generalized in an orderly manner to data from the 15-minute blocks in between the initial and final 15-minute blocks used to train the classifier. Specifically, the ensembles of KPLS scores used to estimate mental states from single EEG epochs in these intervening blocks exhibited an orderly progression from alertness to fatigue. According to the estimates of our classifier, subjects fatigued at widely different rates, but most subjects entered the region of KPLS scores corresponding to fatigue after about 45-60 min on task. We have also assessed the clustering of KPLS-derived features with these data and have found evidence that mental fatigue comprises three distinct states of alertness: normal alertness, heightened alertness, and mental fatigue (Trejo, Knuth, Prado et al., 2007).

The present study also shows that the individual EEG spectral signatures and patterns of change are useful for estimation of cognitive fatigue. The general trend across subjects in EEG spectral change is explained by an alpha + theta model, with both alpha and theta PSDs being greater in fatigued states than in normal states. However, the estimation of momentary cognitive fatigue states from single EEG trials of 13 s duration was most effective with multi-electrode frequency spectra of individual subjects. The KPLS algorithm correctly identified features of EEG spectra for each individual, and accurately classified over 90% of single EEG trials as having come from a period of alertness or mental fatigue in most subjects.

Our results suggest several important implications for future EEG studies of mental fatigue. First, as others have found (e.g., Galbraith & Wong, 1993; Smith et al., 2001), EEG classification algorithms benefit greatly by being both individualized and sensitive to multiple sensors and frequency bands. The development of simpler and more general models, which apply to a broad set of subjects, will require considerable additional research. Another well-known problem in the applied EEG community is that the performance of classification algorithms from day to day, or at different times of day is unstable (Trejo, Matthews, & Allison, 2007; Wilson & Russell, 2003a). Additional research is needed to develop methods for stabilizing the link between EEG features and mental states such as fatigue or alertness over long periods of time. Finally, some type of saliency analysis, such as our maps of spectral features (Fig. 6), is important in order to determine which fundamental aspects of the EEG yield the most accurate predictions (Wilson & Russell, 2003b). Additional research is also needed to develop quantitative methods for saliency analyses that are meaningful across studies and feature extraction or classification algorithms.

References

- Anderson, J. R. (1995). *Cognitive Psychology and Its Implications: Fourth Edition*. New York: Freeman.
- Barth, J.L., Holding, D.H., & Stamford, B.A. (1976). Risk versus effort in the assessment of motor fatigue. *Journal of Motor Behavior*, 8, 189-194.
- Brown, I.D., Tickner, A.H., & Simmonds, D.C.V. (1970). Effect of prolonged driving on overtaking criteria. *Ergonomics*, 13, 239-242.
- Chadhuri, A., & Behan, P.O. (2004). Fatigue in neurological disorders. *Lancet*, 363, 978-988.
- De Jong (1993). SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems* 18, 251-263.
- Dinges, D.F. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research*, 4 (Suppl. 2), 4-14.
- Dingus, T.A., Klauer, S.G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., and Knippling, R.R. (2006). *The 100-Car Naturalistic Driving Study, Phase II - Results of the 100-Car Field Experiment Performed by Virginia Tech Transportation Institute*, Blacksburg, VA, Sponsored by National Highway Traffic Safety Administration, Washington, D.C., April 2006, DOT HS 810 593- April 2006.
- Fatigue Resource Directory (2000). Remarks by Jim Danaher. In National Transportation Safety Board Proceedings: Opening Remarks. Retrieved October 4, 2005, from http://human-factors.arc.nasa.gov/zteam/PDF_pubs/NTSB_Proceedings/NTSB_proceedings_opremarks.pdf.

- Galbraith, G.C., & Wong, E.H. (1993). Moment analysis of EEG amplitude histograms and spectral analysis: Relative classification of several behavioral tasks. *Perceptual and Motor Skills*, 76, 859-866.
- Gevins, A., Leong, H., Du, R., Smith, M.E., Le, J., DuRousseau, D., et al. (1995). Towards measurement of brain function in operational environments. *Biological Psychology*, 40, 169-186.
- 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, 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, 374-385.
- Horne, J.A. (1988). Sleep loss and “divergent” thinking ability. *Sleep*, 11(6), 528-536.
- Hurd, R. (2007). Fatigue. Medline Plus Medical Encyclopedia. U.S. National Library of Medicine and National Institutes of Health.
<http://www.nlm.nih.gov/medlineplus/ency/article/003088.htm>. Accessed 09/11/2008.
- Kennedy, J.L. (1953). Some practical problems of the alertness indicator. In W. F. Floyd & A. T. Welford (Eds.), *Symposium on Fatigue* (pp. 149-153). Oxford, England: H. K. Lewis & Co.
- Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev* 29, 169–195.
- Klimesch, W, Doppelmayr, M., Schimke, H. & Ripper, B (1997). Theta synchronization and alpha desynchronization in a memory task. *Psychophysiology* 34, 69–176.

- Makeig, S., & Inlow, M. (1993). Lapses in alertness: coherence of fluctuations in performance and EEG spectrum. *Electroencephalography and clinical Neurophysiology*, 86, 23-35.
- Mast, T.M., & Heimstra, N.W. (1964). Effects of fatigue on vigilance performance. *Journal of Engineering Psychology*, 3(3), 73-79.
- Mecklinger, A., Kramer, A.F., & Strayer, D.L. (1992). Event-related potentials and EEG components in a semantic memory search task. *Psychophysiology*, 29, 104-119.
- Montgomery, L.D., Montgomery, R.W., & Guisado, R. (1995). Rheoencephalographic and electroencephalographic measures of cognitive workload: analytical procedures. *Biological Psychology*, 40, 143-159.
- Neri, D.F., Shappell, S.A., & DeJohn C.A. (1992). Simulated sustained flight operations and performance, Part 1: Effects of fatigue. *Military Psychology*, 4(3), 137-155.
- Nikolaev, A.R., Ivanitskii, G.A., & Ivanitskii, A.M. (1998). Reproducible EEG alpha-patterns in psychological task solving. *Human Physiology*, 24, 261-268.
- Nyenhuis, D.L., Stern, R.A., Yamamoto, C., Luchetta, T., & Arruda, J.E. (1997). Standardization and validation of the Visual Analogue Mood Scales, *Clinical Neuropsychologist*, 11, 407-415.
- Pfurtscheller G, Lopes da Silva F H. (2005). Event-related desynchronization (ERD) and event-related synchronization (ERS). In: Niedermeyer E, Lopes da Silva F H. *Electroencephalography, basic principles, clinical applications and related fields*, 1003-1016.
- Rosipal, R. & Trejo, L. J. (2001). Kernel partial least squares in reproducing kernel Hilbert space. *Journal of Machine Learning Research*, 2, 97-123.

- Rosipal, R., Girolami, M. and **Trejo**, L., Cichocki, A. (2001). Kernel PCA for feature extraction and de-noising in non-linear regression. *Neural Computing & Applications*, 10, 231-243.
- Rosipal, R., Trejo, L.J., Matthews, B. (2003). Kernel PLS-SVC for linear and nonlinear classification. In T. Fawcett and Nina Mishra (Eds.), *Proceedings of the Twentieth International Conference on Machine Learning*, (pp. 640-647), Washington DC, Aug. 21-24.
- Russo, M.B., Stetz, M.C., & Thomas, M.L. (2005). Monitoring and predicting cognitive state and performance via physiological correlates of neuronal signals. *Aviation, Space, and Environmental Medicine*, 76, C59-C63.
- Shen, A., Li, X., Ong, C., Shao, S. & Wilder-Smith, P. V. (2008). EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clinical Neurophysiology*, 119, 1524–1533.
- Shingledecker, C.A., & Holding, D.H. (1974). Risk and effort measures of fatigue. *Journal of Motor Behavior*, 6, 17-25.
- Šípoš, I. (1971). Verbal vigilance performance and visual fatigue. *Studia Psychologica*, 13, 141-142.
- Smit, A.S., Eling, P.A.T.M., & Coenen, A.M.L. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta Psychologica*, 115, 35-42.
- 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, 366-380.
- Stern, R.A. (1997). Visual Analogue Mood Scales. Odessa, FL: Psychological Assessment Resources.

- Thayer, R. E. (1986). Activation-Deactivation Adjective Check List: Current overview and structural analysis. *Psychological Reports, 58*, 607-614.
- Thayer, R. E. (1989). *The Biopsychology of Mood and Arousal*. New York: Oxford University Press.
- Trejo, L.J., & Shensa, M.J. (1999). Feature extraction of event-related potentials using wavelets: An application to human performance monitoring. *Brain and Language, 66*, 89-107.
- Trejo, L. J., Knuth, K., Prado, R., Rosipal, R., Kubitz, K. Kochavi, R., Matthews, B., & Zhang, Y. (2007). EEG-based estimation of mental fatigue: Convergent evidence for a three-state model. HCI International 2007 and Augmented Cognition International Conference, Beijing, China, July 22-27. In D.D. Schmorow, L.M. Reeves (Eds.): *Augmented Cognition, HCII 2007*, LNAI 4565, pp. 201-211, New York: Springer LNCS.
- Trejo, L. J., Matthews, B., and Rosipal, R. (2006). Brain-computer interfaces for 1-D and 2-D cursor control: designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14*, 225-229.
- Trejo, L. J., Matthews, R., & Allison, B. Z. (2007). Experimental Design and Testing of a Multimodal Cognitive Overload Classifier. In D. D. Schmorow, D. M. Nicholson, J. M. Drexler, & L. M. Reeves (Eds.), *Foundations of Augmented Cognition*, (4th ed., pp. 13-22). Arlington, VA: Strategic Analysis, Inc.
- U.S. Congress, Office of Technology Assessment (1991). *Biological Rhythms: Implications for the Worker*, OTA-BA-463. Washington, DC: U.S. Government Printing Office.

- van der Hulst, M., Meijman, T., & Rothengatter, T. (2001). Maintaining task set under fatigue: a study of time-on-task effects in simulated driving. *Transportation Research Part F: Traffic Psychology & Behavior*, 4, 103-118.
- van der Linden, D., Frese, M., & Sonnentag, S. (2003). The impact of mental fatigue on exploration in a complex computer task: Rigidity and loss of systematic strategies. *Human Factors*, 45, 483-494.
- van der Linden, D., Frese, M., & Meijman, T.F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, 113, 45-65.
- Vidulich, M.A., Stratton, M., Crabtree, M., & Wilson, G. (1994). Performance-based and physiological measures of situational awareness. *Aviation, Space, and Environmental Medicine*, 65, A7-A12.
- Wallerius, J., Trejo, L. J, Matthews, R., Rosipal, R., and Caldwell, J. A. (2005). Robust feature extraction and classification of EEG spectra for real-time classification of cognitive state. *11th International Conference on Human Computer Interaction jointly with 1st International Conference on Augmented Cognition*. Las Vegas, NV, 22-27 July, 2005.
- 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, 959-962.
- Wilson, G.F., & Fisher, F. (1995). Cognitive task classification based upon topographic EEG data. *Biological Psychology*, 40, 239-250.
- Wilson, G.F., & Russell, C.A. (2003a). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors*, 45 (3), 381-389.

- Wilson, G.F., & Russell, C.A. (2003b). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors*, 45 (4), 635-643.
- Winningham, M.L. (1996). Fatigue. In S.L. Groenwald, M. Hansen Frogge, M. Goodman, & C.H. Yarbro (Eds.), *Cancer symptom management* (pp. 42–54). Sudbury, MA: Jones and Bartlett.

Author Note

This research was supported in part by the *NASA Aerospace Operations Systems Program* directed by Dr. Malcolm Cohen and the *Advanced Estimation of Cognitive Status Project (ARO STIR)* sponsored by Dr. Elmar T. Schmeisser of the Life Sciences Division of the U.S. Army Research Laboratory's Army Research Office. We would also like to acknowledge the support of the NASA/ASEE Faculty Fellowship and the NASA Ames Educational Associates programs for supporting Dr. Karla Kubitz. We also acknowledge the contributions of Richard W. Montgomery and Bernadette Luna for theoretical discussions and use of the mental arithmetic task, and Melina L. Trejo for extensive support of data analysis tasks.

Table 1. Number of participants per 15-min block.

Block	1	2	3	4	5	6	7	8	9	10	11	12
Time (min)	15-30	30-45	45-60	60-75	75-90	90-105	105-120	120-135	135-150	150-165	165-180	180-205
Participants	16	16	16	16	16	16	15	14	14	13	12	11

Table 2. Scale used to quantify observed activity and alertness.

Value	Activity	Alertness
1	Little or no movement (0-10% time)	Asleep
2	Occasional movement (10-25% time)	Dozing off but still responding
3	Intermittent movement (25-50% time)	Distracted, yawning or only somewhat alert
4	Frequent movement (50-75% time)	Completely awake and mostly alert
5	Constant movement (75-100% time)	Completely awake and fully alert

Table 3. KPLS-DLR classification accuracies (test proportion correct or TPC) by filter cutoff and class membership.

Low pass cutoff	TPC Train	TPC Test	TPC Class 1	TPC Class 2	Mean number of components
11 Hz	99.96	97.01	97.37	95.57	3.76
18 Hz	99.86	98.30	98.78	96.97	2.77

Figure Captions

Figure 1. Schematic diagram of events in the mental arithmetic task.

Figure 2. Effects of time on task on response time and accuracy. Response times trended linearly upwards over time beginning after Block 3, or 45 minutes on task. Error rates declined over time but the trend was not significant.

Figure 3. Average EEG spectra across all subjects for the first (fine line) and final (heavy line) 15 min blocks of the math task. Left: Electrode Fz shows an increase in theta power spectral density (PSD) near 6-7 Hz. Right: electrode Pz shows an increase in alpha PSD near 8-11 Hz.

Figure 4. Classification accuracies for 14 subjects and the averages across subjects with 1-18 Hz band pass. Hatched bars show test set accuracies for overall classification; black bars are for Class 1 (alert) epochs, and white bars are for Class 2 (fatigued) epochs.

Figure 5. Example of KPLS scores predicted for single-trial EEG spectra for early (dark circles) and late (light triangles) blocks of the mental arithmetic task in one subject.

Figure 6. Topographical maps fit to the first KPLS component weights in theta and alpha bands for one subject (819). The colored areas are smoothed normalized absolute values, with the largest values in white and the smallest in black.

Figure 7. Estimating the development of fatigue over time in one subject (819). KPLS scores were predicted for EEG epochs from nine 15 minute blocks between the training set blocks (B1 & B12). Block 2 = 15-30 min, block 3 = 30-45 min, block 4 = 45-60 min, ... , block 10 = 135-150 min. Black circles and purple crosses are the KPLS C1 and C2 scores of single EEG epochs from fatigued (block 1) and non-fatigued (block 12) training sets, respectively. Colored diamonds are the KPLS C1 and C2 scores ($x = C1$, $y = C2$) of single EEG epochs for intervening 15-minute

blocks 2-10. In this subject, the drift of the orange diamonds in block 3 away from the black circles and towards the purple crosses marks the onset of fatigue after 30-45 minutes on task. By the tenth block most predicted scores fell in the fatigue region, as defined by the training data set.

Figure 8. Means of the KPLS classification scores for each of ten 15-minute blocks in 11 subjects. These means correspond to the centers (arithmetic means) of the clouds of points shown moving from the alert region to the fatigue region in Fig. 7. Negative scores represent alert states and positive scores represent fatigue states.

Figure 1.

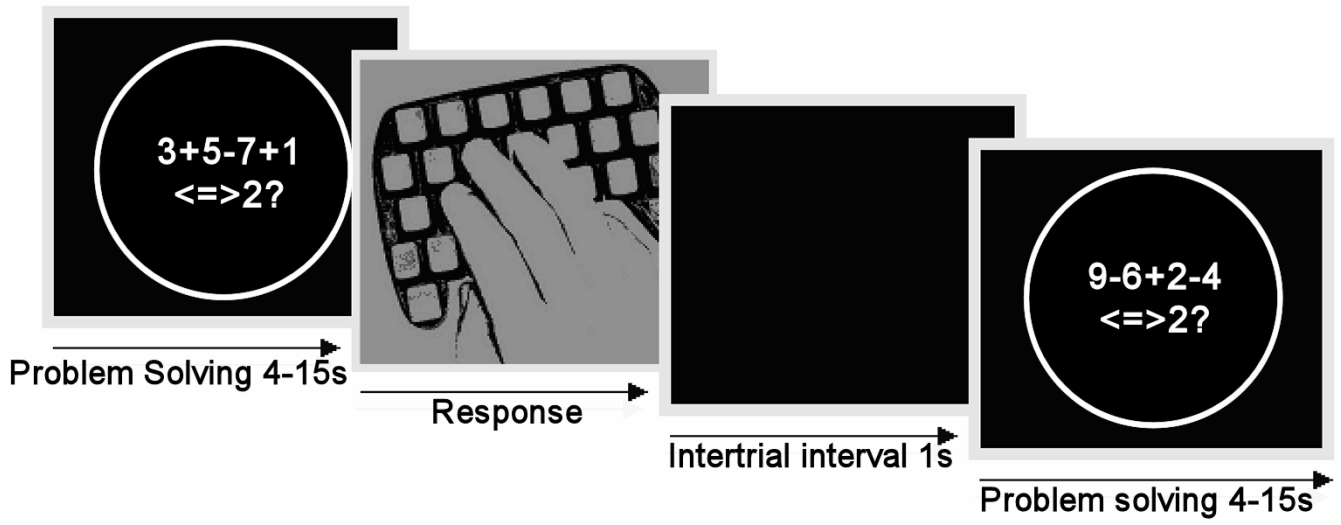


Figure 2.

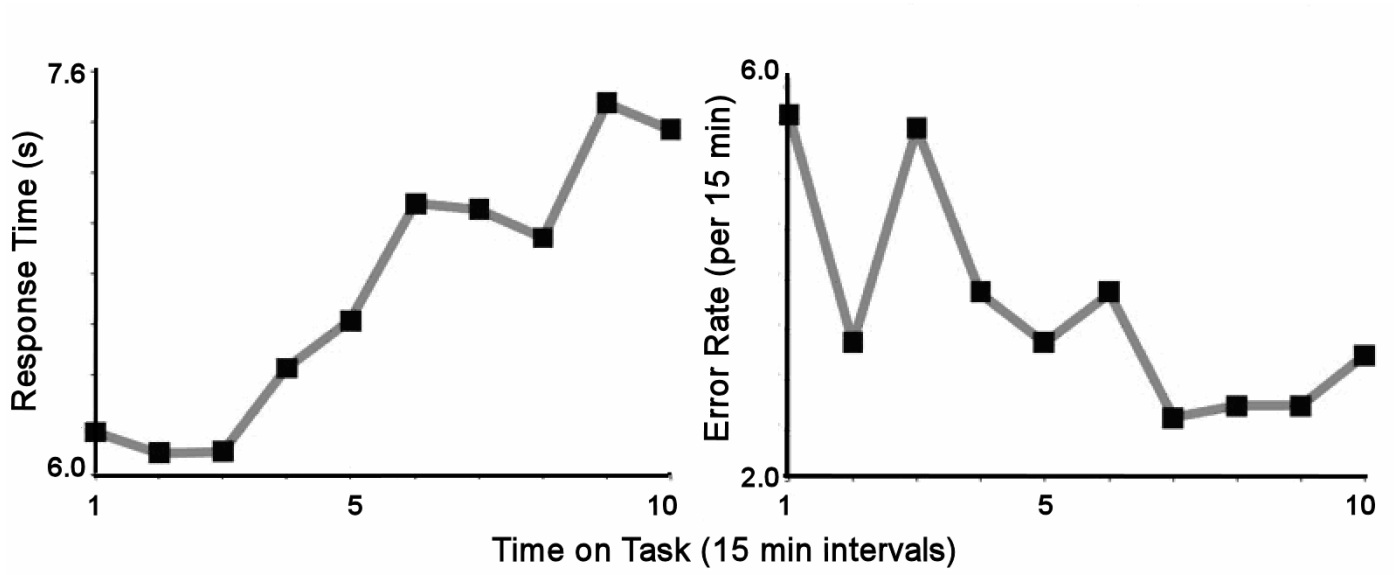


Figure 3.

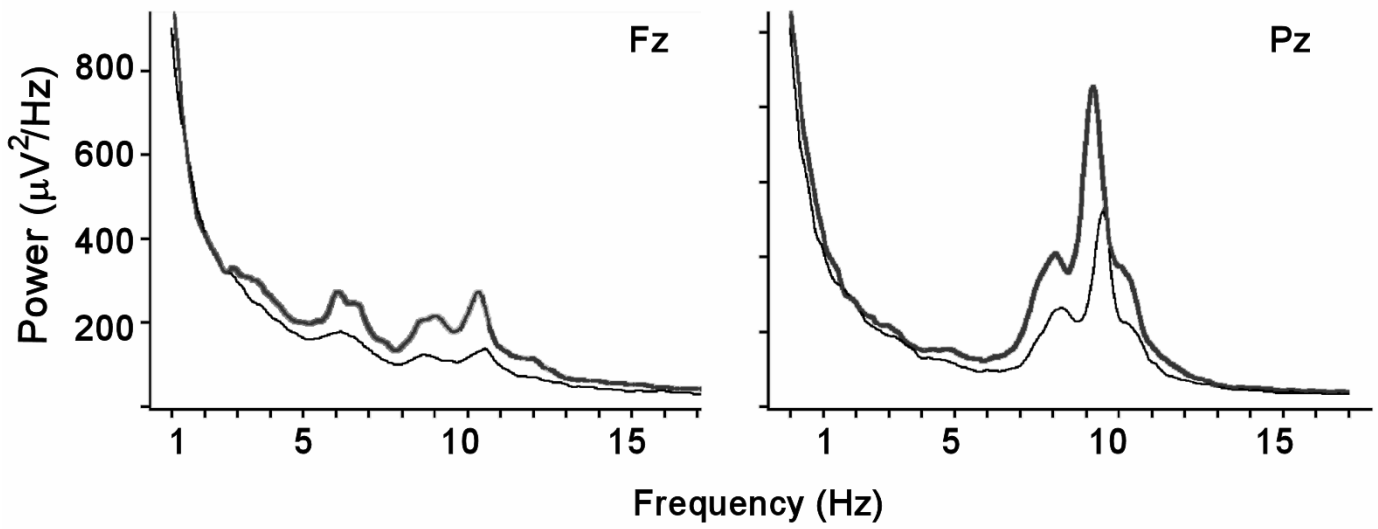


Figure 4.

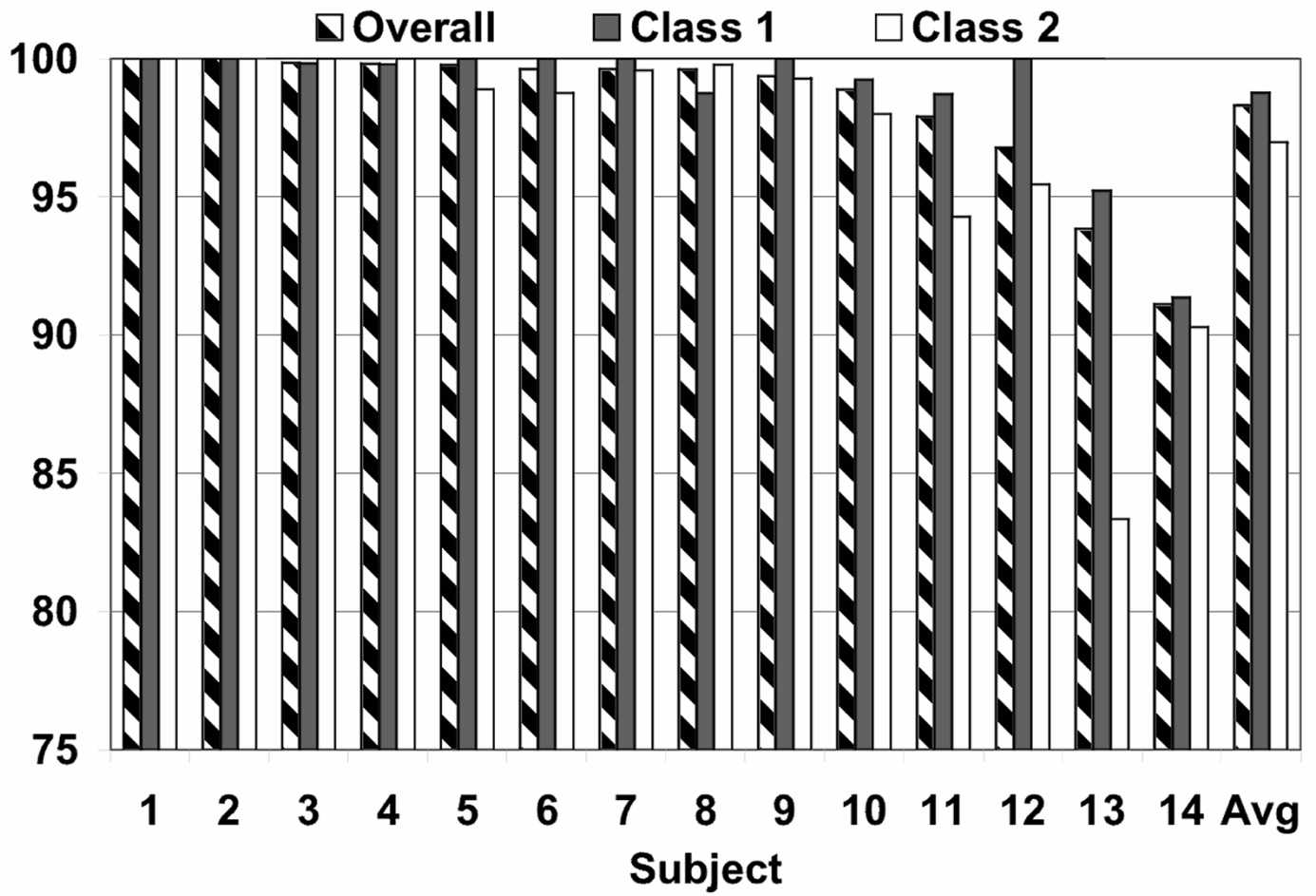


Figure 5.

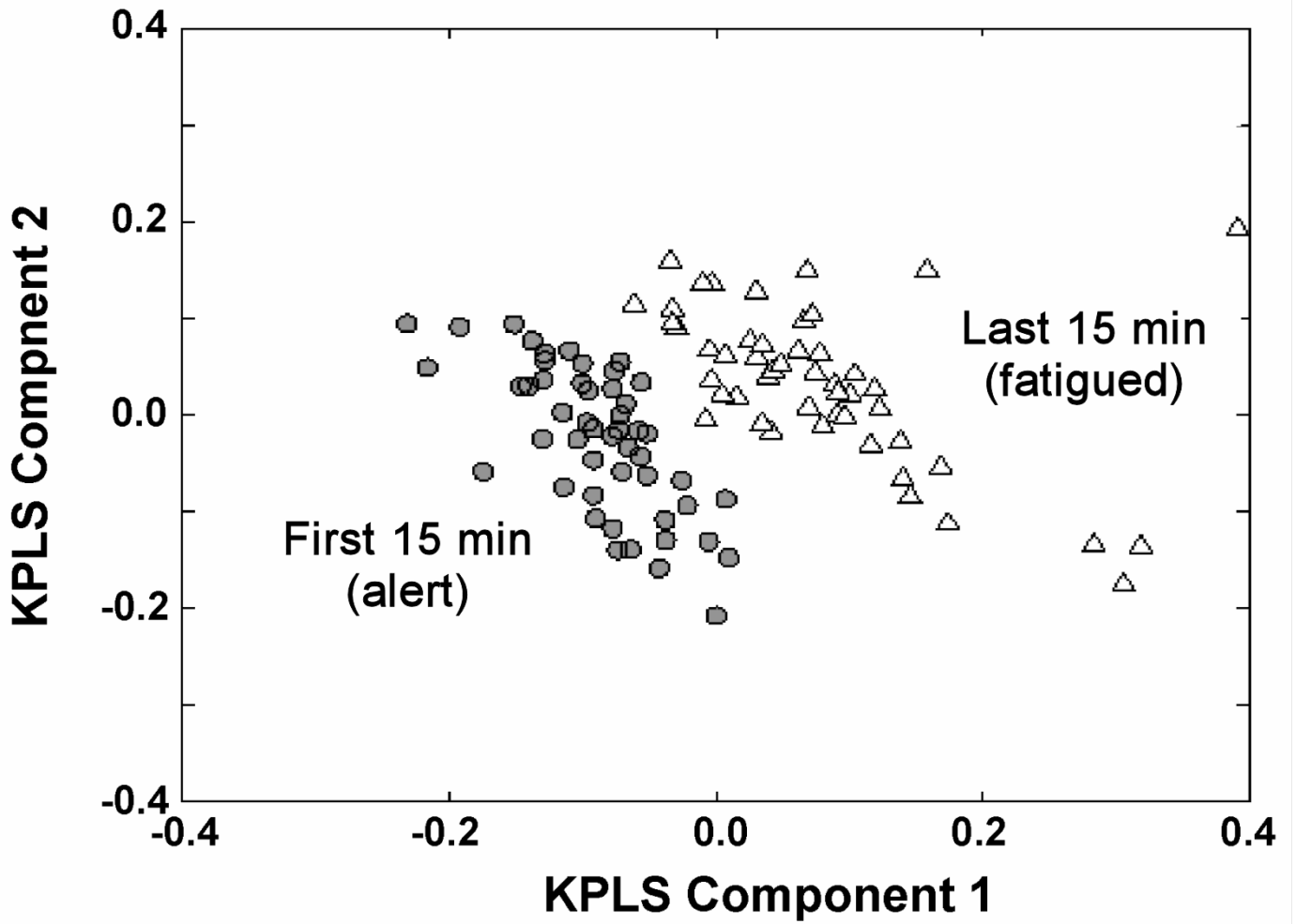


Figure 6.

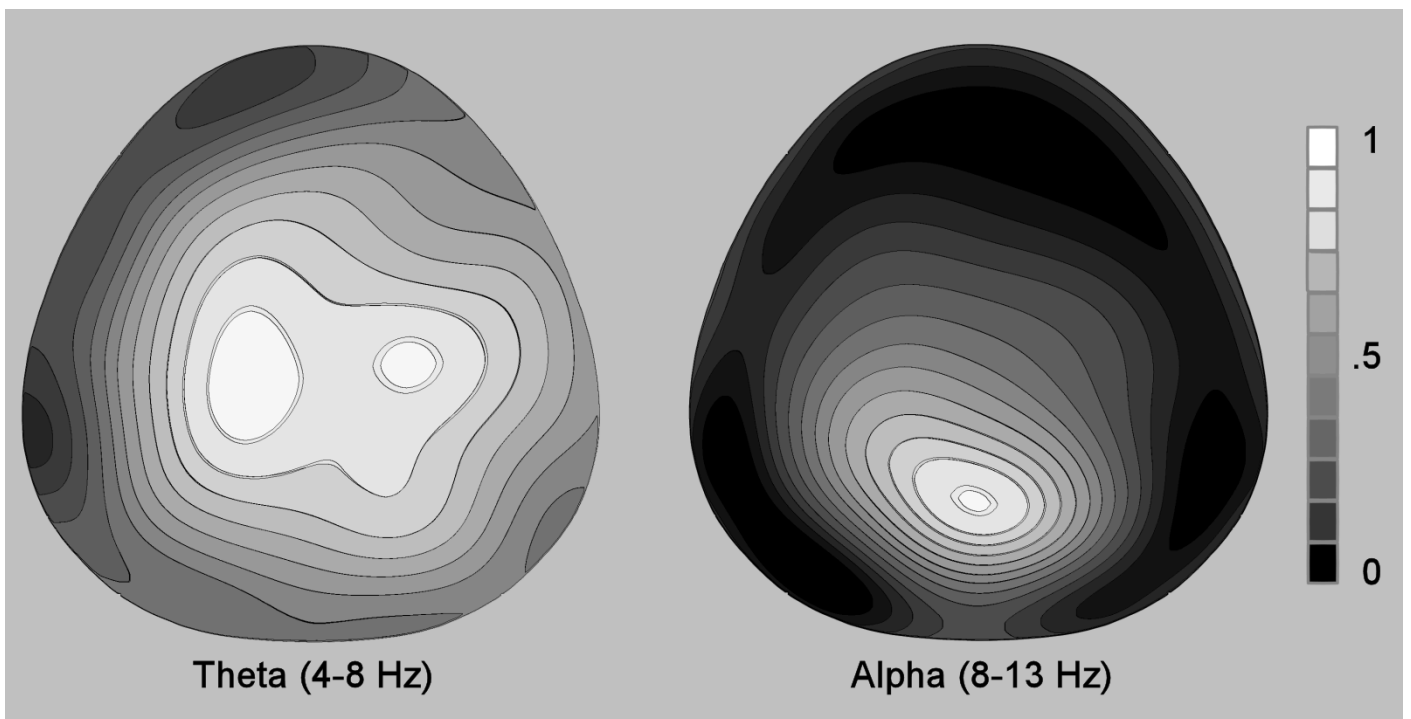


Figure 7.

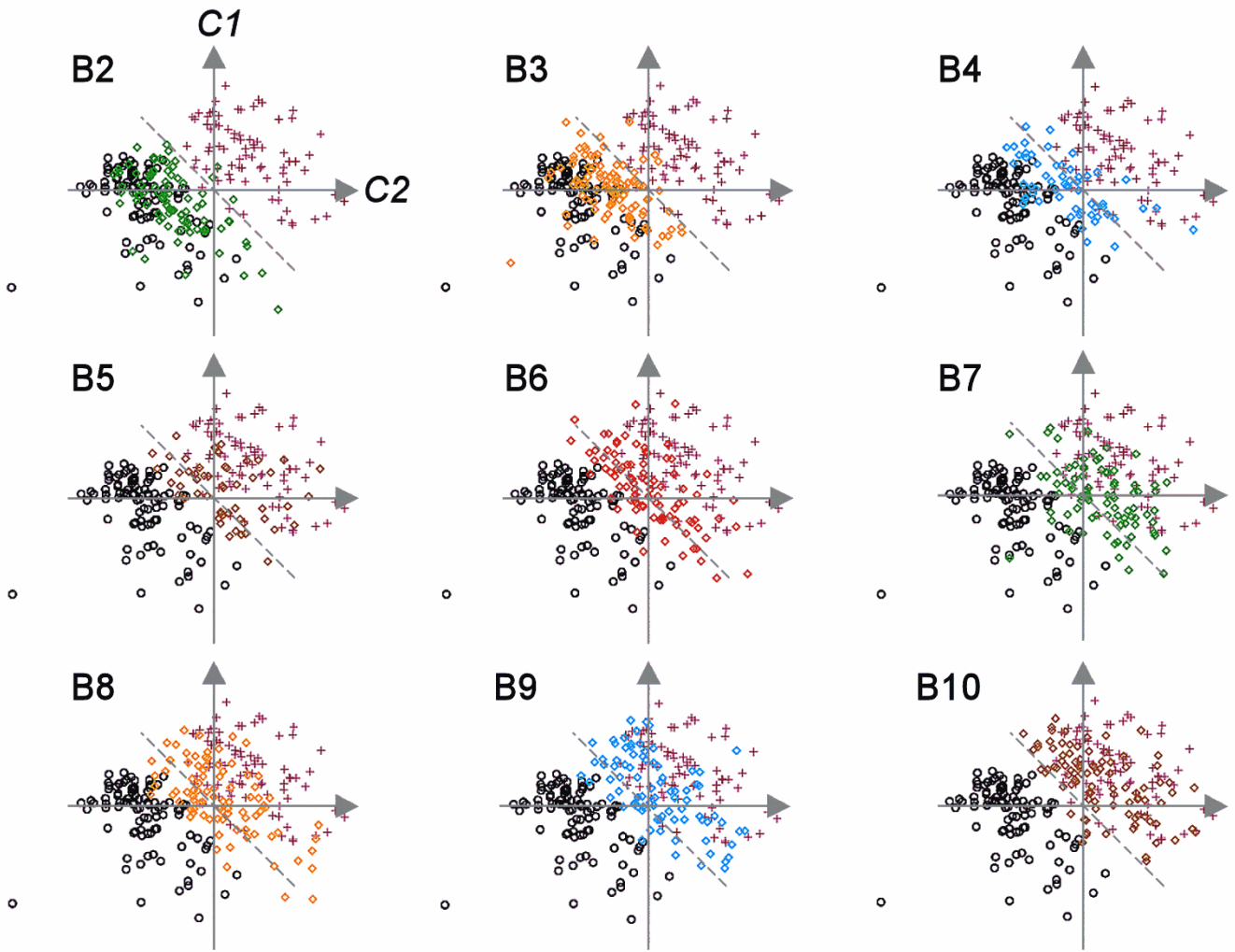


Figure 8.

