Corresponding Author Information

Name

Leonard Trejo

Address

MS 269-3, NASA Ames Research Center
Moffett Field California 94035-1000
Telephone: 650-604-2187
email: ltrejo@mail.arc.nasa.gov

Abstract Title

Measures and Models for Estimating and Predicting Cognitive Fatigue

Abstract Body

We analyzed EEG and ERPs in a fatiguing mental task and created statistical models for single subjects. Seventeen subjects (4 F, 18-38 y) viewed 4-digit problems (e.g., 3+5-2+7=15) on a computer, solved the problems, and pressed keys to respond (intertrial interval = 1 s). Subjects performed until either they felt exhausted or three hours had elapsed. Pre- and post-task measures of mood (Activation Deactivation Adjective Checklist, Visual Analogue Mood Scale) confirmed that fatigue increased and energy decreased over time. We tested response times (RT); amplitudes of ERP components N1, P2, P300, readiness potentials; and amplitudes of frontal theta and parietal alpha rhythms for change as a function of time. For subjects who completed 3 h (n=9) we analyzed 12 15-min blocks. For subjects who completed at least 1.5 h (n=17), we analyzed the first-, middle-, and last 100 error-free trials. Mean RT rose from 6.7 s to 8.5 s over time. We found no changes in the amplitudes of ERP components. In both analyses, amplitudes of frontal theta and parietal alpha rose by 30% or more over time. We used 30-channel EEG frequency spectra to model the effects of time in single subjects using a kernel partial least squares classifier. We classified 3.5-s EEG segments as being from the first 100 or the last 100 trials, using random sub-samples of each class. Test set accuracies ranged from 63.9% to 99.6% correct. Only 2 of 17 subjects had mean accuracies lower than 80%. The results suggest that EEG accurately classifies periods of cognitive fatigue in 90% of subjects.

List of Authors and Affiliations

Trejo, Leonard J., NASA Ames Research Center
Kochavi, Rebekah, QSS Group, NASA Ames Research Center
Kubitz, Karla, UC Santa Cruz EA Program, Towson University
Montgomery, Leslie D., UC Santa Cruz EA Program, NASA Ames Research Center
Rosipal, Roman, National Research Council, NASA Ames Research Center
Matthews, Bryan, QSS Group, NASA Ames Research Center

Measures and Models for Predicting Cognitive Fatigue

Leonard J. Trejo[*a], Rebekah Kochavi[a], Karla Kubitz[b],
Leslie D. Montgomery[a], Roman Rosipal[a], Bryan Matthews[a]
[a]NASA Ames Research Center, MS 269-3, Moffett Field, CA, USA 94035-1000
[b]Towson University, 8000 York Rd., Towson, Md., 21252

ABSTRACT

We measured multichannel EEG spectra during a continuous mental arithmetic task and created statistical learning models of cognitive fatigue for single subjects. Sixteen subjects (4 F, 18-38 y) viewed 4-digit problems on a computer, solved the problems, and pressed keys to respond (inter-trial interval = 1 s). Subjects performed until either they felt exhausted or three hours had elapsed. Pre- and post-task measures of mood (Activation Deactivation Adjective Checklist, Visual Analogue Mood Scale) confirmed that fatigue increased and energy decreased over time. We examined response times (RT); amplitudes of ERP components N1, P2, and P300, readiness potentials; and power of frontal theta and parietal alpha rhythms for change as a function of time. Mean RT rose from 6.7 s to 7.9 s over time. After controlling for or rejecting sources of artifact such as EOG, EMG, motion, bad electrodes, and electrical interference, we found that frontal theta power rose by 29% and alpha power rose by 44% over the course of the task. We used 30-channel EEG frequency spectra to model the effects of time in single subjects using a kernel partial least squares (KPLS) classifier. We classified 13-s long EEG segments as being from the first or last 15 minutes of the task, using random sub-samples of each class. Test set accuracies ranged from 91% to 100% correct. We conclude that a KPLS classifier of multichannel spectral measures provides a highly accurate model of EEG-fatigue relationships and is suitable for on-line applications to neurological monitoring.

Keywords: EEG, cognitive, fatigue, estimation, monitoring

1. INTRODUCTION

Aerospace jobs require sustained mental work for periods of up to several hours, often without intermittent rest periods. For example, NASA astronauts perform mentally demanding intra- and extravehicular activities, which may last for several hours and airline and military pilots fly continuously for hours at a time. Laboratory experiments and operational reports from situations like these show that performance and associated cognitive functions decline with time on task. The affected cognitive functions include alertness, attention, working memory, long-term memory recall, situation awareness, judgment, and executive control.

In this study, we were concerned with decrements in cognitive function arising during sustained mental work in a controlled laboratory experiment. We refer to these decrements as *cognitive fatigue* to distinguish them from effects of sleepiness, motivation, learning, and physical fatigue. We define cognitive fatigue as the unwillingness of alert, motivated subjects to continue performance of mental work[1], a definition that has been supported by behavioral studies[2]. In this study, we examine EEG and ERP correlates of cognitive fatigue during sustained mental work. We also describe useful EEG measures and statistical models for estimating and predicting cognitive fatigue in individual subjects.

1.1. EEG Hypotheses

Previous studies have reported EEG spectral changes as alertness declines. For example, the proportion of low frequency EEG waves, such as theta and alpha rhythms, may increase while the proportion of higher frequency waves, such as beta rhythms may decrease. For example, as alertness fell and error rates rose in a vigilance task, Makeig and Inlow[3], found progressive increases in EEG power at frequencies centered near 4 and 14 Hz. Thus, the relative power

[*] ltrejo@mail.arc.nasa.gov; phone 1-650-604-2187, fax 1-650-604-4036; ic.arc.nasa.gov

of theta, alpha, and other EEG rhythms may serve to indicate the level of fatigue that subjects experience. However, the EEG spectral changes that relate to cognitive fatigue, in the absence of alertness decrements, are unclear because most experiments have used either vigilance-like paradigms or short duration/ high workload experimental sessions to examine fatigue. In contrast, we used a sustained low-workload mental arithmetic task and encouraged subjects to maintain alertness, motivation, and high response accuracy so as to minimize vigilance-related effects of arousal or alertness decrements.

We designed our study to allow for high-resolution estimation of the EEG frequency spectrum over the entire course of each experimental run. In addition, we tested the null hypothesis that EEG power in specific theta and alpha bands remained constant over the course of a fatigue-inducing task.

1.1. ERP Hypotheses

Other studies have suggested links between fatigue and changes in event-related potential (ERP) components. For example, the visual N100 component is sensitive to spatial and non-spatial directions of attention, being of larger amplitude for attended than ignored stimuli. If fatigue were to reduce subjects' ability to focus and sustain attention to task-relevant stimuli, there may be corresponding decreases in N100 amplitude. Similarly, the P300 component is known to reflect the allocation of processing resources to task-relevant stimuli, being of larger amplitude in high-workload tasks than in low-workload tasks[4]. On the other hand, long periods of extended wakefulness[5] are linked to increases in errors, non-responses, response latencies, and P300 latencies, and decreases in P300 amplitudes. So if cognitive fatigue makes a given mental task seem more difficult than during non-fatigued conditions, we may find corresponding increases in P300 amplitudes for the task-relevant stimuli. However, if the effects of cognitive fatigue resemble those of extended wakefulness, we should find correlated increases in latency and decreases in amplitudes of the P300.
We designed our experiment to allow for the accurate estimation of ERPs elicited by the onset of each mental arithmetic problem. In addition we tested the null hypotheses that specific ERP components, N100, P200, and P300 remained constant in amplitude and latency over the course of a fatigue-inducing task.

1. METHODS

1.1. Participants

Data were collected from 33 individuals recruited from the NASA Ames Research Center community. However, 17 of the 33 participants were excluded from analyses. Eight were excluded because their EEG data contained high noise levels, which could not be filtered or corrected. Four were excluded because they either fell asleep (n=3) or violated experimental protocol (n=1; wore a watch). Five 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. All participants signed an informed consent approved by the NASA Ames Research Center and were paid for their participation. Also, according to their self-reports, all of the participants had normal vision and hearing and 14 of the 16 participants were right-handed.

1.1. Experimental Design

We tested several hypotheses about the dependence of subjective moods, observed behavior, performance, and physiological measures induced by 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[6] (AD-ACL) and the Visual Analogue Mood Scales[7] (VAMS, Psychological Assessment Resources, Inc., Lutz, FL) questionnaires. Observed behavior included ratings of activity and alertness from videotaped recordings of each subject's performance. The performance measures were response time and response accuracy. The physiological measures included several measures of spontaneous EEG and event-related potentials, that is: (a) theta activity at Pz (both average power and peak amplitude in the theta band); (b) alpha activity at Fz (both average power and peak amplitude in the alpha band); and (c) peak amplitudes and latencies of the N100, P200, and P300 components of event-related potentials elicited by onset of the task stimuli.

1.1. Mental Arithmetic Task

Participants sat in front of a computer with their right hand resting on a 4-button keypad (Neuroscan STIM pad). 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; and (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.

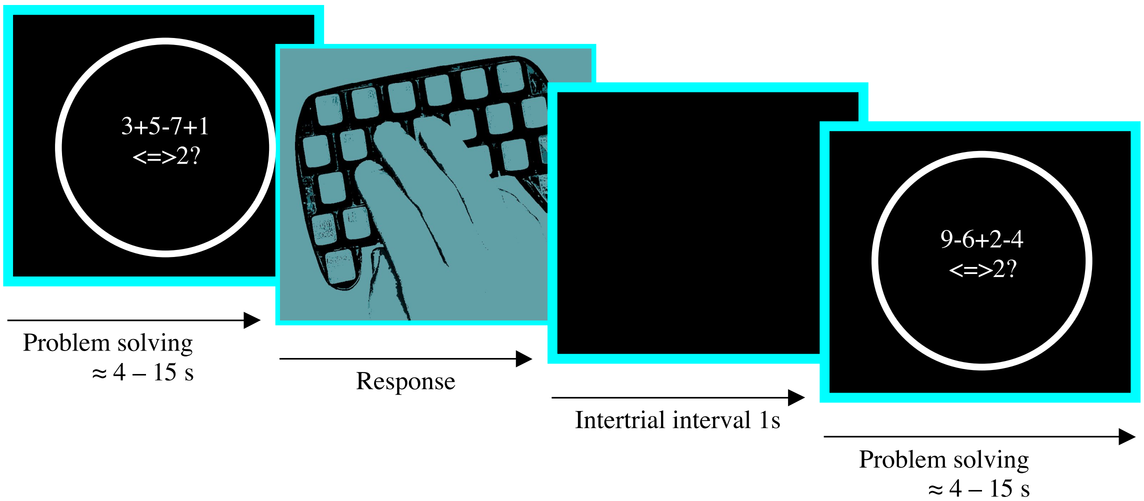

Figure 1. Schematic diagram of events in the mental arithmetic task.

1.1. Activation Deactivation Adjective Checklist

Thayer's AD ACL[6] 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), calmness (reflects low preparatory arousal). The AD ACL is a reliable and valid subjective method[6].

1.1. Visual Analogue Mood Scales

The VAMS[7] 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[7].

1.1. 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 rater judged levels of alertness and activity (unnecessary motion) on a five-point scale. Alertness was rated as a 1 if the participant was asleep; a 2 if the participant was dozing off but still responding; a 3 if the participant was awake but distracted, yawning, and only somewhat alert; a 4 if the participant was completely awake and mostly alert; and a 5 if the participant was completely awake and fully alert. Activity was rated as a 1 if the participant was sitting still with little or no movement; a 2 if the participant was mostly sitting still with little or occasional movement; a 3 if the participant was spontaneously moving and fidgeting; a 4 if the participant was almost constantly fidgeting, tapping, or shaking; and a 5 if the participant was constantly moving. Ratings were tested for correlations with response times and accuracy, and with EEG spectral measures.

1.1. EEG Activity

EEG activity was recorded continuously using 32 Ag/AgCl electrodes embedded in an elastic fabric cap (i.e., a Quik-Cap™). The electrode cap was placed on the participant according to the manufacturer's instructions (Compumedics USA, El Paso, TX). The reference electrodes were electronically linked 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 5kΩ for EEG electrodes and 10 kΩ for EOG electrodes. The electroencephalogram was amplified and digitized with a 64-channel Neuroscan Synamps™ 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 stored on hard disk drives by a Pentium II computer equipped with Neuroscan Scan 4.2 software (Compumedics USA, El Paso, TX) and archived on optical media (CD-R).

1.1. Procedures

1.1.1. Participants

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; (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 either until three hours had elapsed or until volitional exhaustion had occurred. Task termination was followed by the completion of post-test self-report measures and participant debriefing.

1.1.1. Data Processing

The EEGs, initially processed using Neuroscan Scan 4.2 Edit™ (Compumedics USA, El Paso, TX) software, the EEGs were: (a) submitted to an algorithm for the detection and elimination of eye-movement artifact; (b) visually examined and blocks of data containing artifact greater than 100 μV 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 > 100μV). 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 residuals of these predictions served to estimate the artifact-free EEG. EEG power spectra were estimated with the Welch's periodogram method at 833 frequencies from 0-64 Hz[8]. Peak and average power in the theta and alpha bands were measured at electrodes Fz and Pz, respectively.

The ERP data were initially processed using the same methods as for the EEGs, then epoched around the stimulus (–1.5s pre- to +2s post-stimulus). The overall rejection rate was 19% for the ERP data. The 'cleaned and filtered' epochs were: (a) decimated to a sampling rate of 128 Hz; (b) corrected for 'residual' EOG artifact; and (c) averaged across the 1st 100, middle 100, and last 100 trials. The average ERPs were used to measure the latencies and amplitudes of the N100, P200, and P300 components. Latencies were peak latencies and were determined based on visual examination of the spatial distribution for the component (i.e., N100, P200, and P300). Amplitudes were mean amplitudes and were calculated as mean amplitudes (at O2, Fz, and CPz for the N100, P200, and P300 components, respectively) in a window +/- 50 ms around the peak latency.

1.1.1. Classification Procedures

We classified single EEG epochs using KPLS-DLR, or *kernel partial least squares decomposition* of multichannel EEG spectra coupled with a *discrete-output linear regression classifier*. Through extensive side-by-side testing of EEG data, we have found that KPLS-DLR is just as accurate as KPLS-SVC, which uses a support vector classifier[9] for the classification step. KPLS selects the reduced set of orthogonal basis vectors or "components" in the space of the independent variables (EEG spectra) that maximizes covariance with the experimental conditions. DLR finds the linear hyperplane in the space of KPLS components that maximizes the margin between the classes. In a pilot study, and in our present data, we found that the first 15 minutes on task did not produce cognitive fatigue, whereas cognitive 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 model selection was the minimum classification error rate summed over all (five) cross-validation subsets.

1.1.1. Statistical Analyses

The data were analyzed using either singly or, when appropriate, doubly multivariate repeated measures analyses of variance with either time of measurement (for the self-report, behavior, and EEG analyses) or number of artifact-free trials as a within-subjects factor (for the ERP analyses). The 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), alpha activity data (peak and band-average amplitudes), N100 data (amplitudes and latencies), P200 data (amplitudes and latencies), and P300 data (amplitudes and latencies) were analyzed using doubly multivariate analyses. The response times and accuracy data were analyzed using singly multivariate analyses of variance. 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 cases, partial η^2 values were reported as effect size estimators.

1. RESULTS

1.1. 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 ($F(4,5)=10.4$, $p<.01$, h2= .89). Within-subjects contrasts showed significant linear trends for energy ($F(1,8)=6.46$, $p<.04$, h2= .45), calmness ($F(1,8)=21.3$, $p<.002$, h2= .73), and tiredness ($F(1,8)=6.38$, $p<.04$, h2= .44). Energy decreased from a pretest mean of 12.0 (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 VAS 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 moods.

1.1. Behavior Analyses

The behavioral observations (i.e., observed activity and alertness) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., 10 15-min periods) as a within-subjects factor. The main effect of time of measurement was significant ($F(18,178)=3.70$, $p<.0005$, h2= .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 contrast showed a linear decrease in alertness ($F(1,10)=10.4$, $p<.009$, h2= .51) and a linear increase in activity ($F(1,10)=5.88$, $p<.04$, h2= .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 $F(3,39)=3.78$, $p<.03$, h2= .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 ($F(1,12)=8.29$, $p<.01$, h2= .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., 1^{st}100, 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., 1^{st}100, 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.

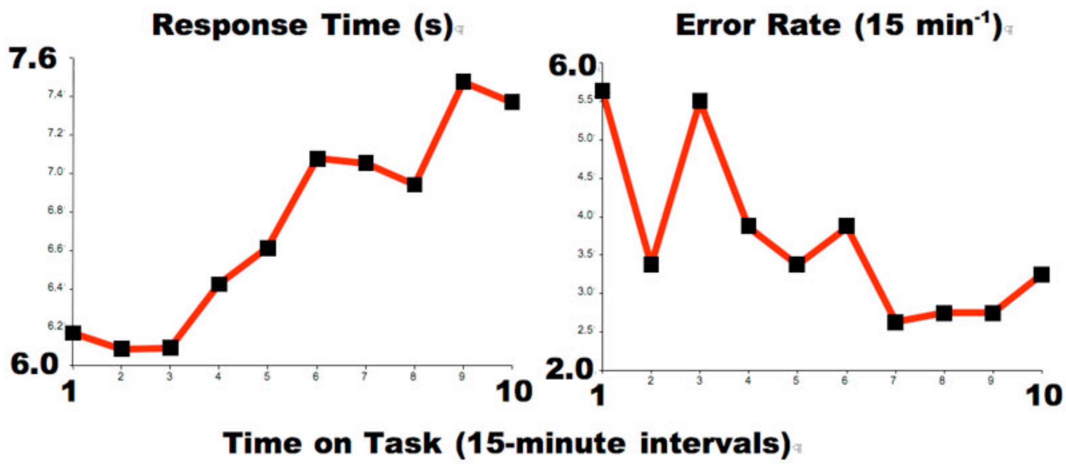

Figure 2. Effects of time on task on response time and accuracy. Response times trended linearly upwards over time beginning after block 3 or 45 minutes on task. Error rates declined over time but the trend was not significant.

1.3. EEG Analyses

Average spectra revealed changes in frontal theta and parietal alpha bands over time (Fig. 3). The changes in frontal midline theta (i.e., average power densities 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 $F(18,178)=2.05$, $p<.01$, h2= .17). Average power in the theta band increased from a mean of 199.36 (SD=97.50) in the first 15-min period to a mean of 256.58 (SD=135.57) in the last 15-min period. Peak amplitude in the theta band increased from a mean of 272.4 (SD=146.0) in the first 15-min period to a mean of 390.8 (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 power densities ($F(1,10)=7.42$, $p<.01$, h2= .48) and in peak theta amplitudes ($F(1,10)=9.31$, $p<.01$, h2= .48).

The changes in midline parietal alpha activity (i.e., average power densities 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 $F(18,178)=2.20$, $p<.005$, h2= .18). Average alpha power densities increased from a mean of 307.4 (SD=434.3) in the first 15-min period to a mean of 459.0 (SD=593.9) in the last 15-min period. This analysis suggests that alpha increased with time on task. Moreover, this analysis suggests that our manipulation had a progressive effect on parietal alpha activity. Within-subjects contrasts showed significant linear increases in average alpha power densities ($F(1,10)6.07$, $p<.03$, h2= .38). Peak alpha amplitudes increased and trended similarly, but not significantly so ($F(1,10)=4.11$, $p=.07$).

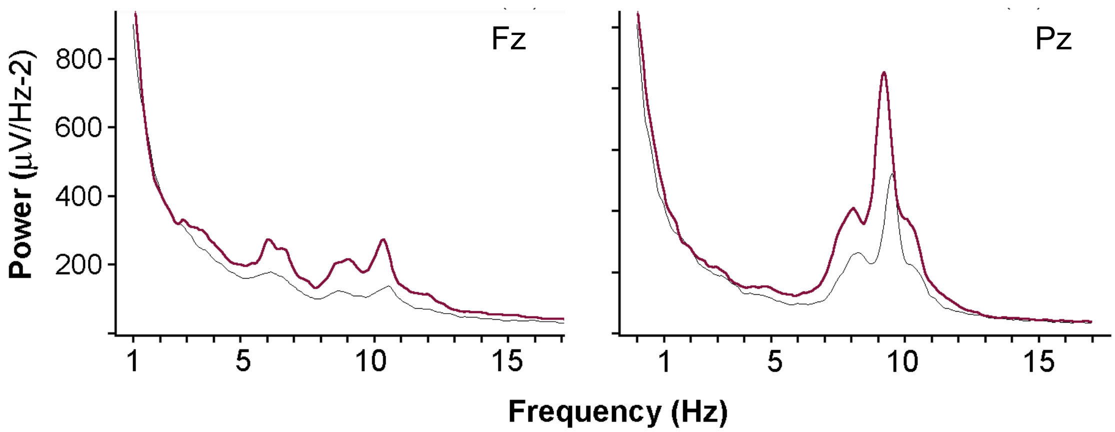

Figure 3. Average EEG spectra across all subjects for the first (black/fine line) and final (red/heavy line) 15-min blocks of the math task. Left: Electrode Fz shows an increase in theta power near 6-7 Hz. Right: electrode Pz shows an increase in alpha power near 8-11 Hz.

1.4. ERP Analyses

The N100, P200, and P300 latencies and amplitudes were analyzed in separate doubly multivariate ANOVAs with number of artifact-free trials (i.e., 1^{st}100, middle 100, and last 100) as a within-subjects factor. For N100 and P300 alike, the main effect of number of trials was not significant (N100 multivariate $F(4,54)=1.59$, p=.19; P300 multivariate $F(4,34)=1.02$, p=.41). This analysis suggests that time on task did not influence the N100 or P300. In the P300 range above 500 ms, P300 amplitudes in the last 100 trials were slightly larger than the 1^{st} 100, but not significantly.

For P200 the main effect of number of trials was significant (multivariate $F(4,54)=7.77$, p<.0005, h2= .37). This analysis suggests that time on task influenced the P200s. Moreover, this analysis suggests that time on task had a curvilinear effect on the P200s. Within-subjects contrasts showed a significant quadratic trend for the P200 amplitudes ($F(1,14)=16.1$, p<.001, h2= .54). Within-subjects contrasts did not show a significant quadratic (or linear) trend for the P200 latencies, $F(1,14)=2.55$, p=.13. P200 amplitudes averaged 4.82 (SD=1.64) in the 1^{st} 100 trials, 6.21 (SD=2.38) in the middle 100 trials, and 4.29 (SD=1.72) in the last 100 trials.

1.5. Classification

We applied our classification procedure to EEG recordings from 14 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 EMG artifact by low-pass filtering the EEG with 11- or 18-Hz cutoffs.

For each subject we constructed a KPLS model using either linear or Gaussian (nonlinear) kernels and selected the best model as described above. We then constructed a support vector classifier for each model, which served to classify the KPLS component scores for each EEG epoch. Results for linear and 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 = 98.30, Table 1). The corresponding range for 11-Hz filtered EEG was 89.53 to 98.89% (mean = 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 1). 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).

Table 1. KPLS-DLR classification accuracies 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

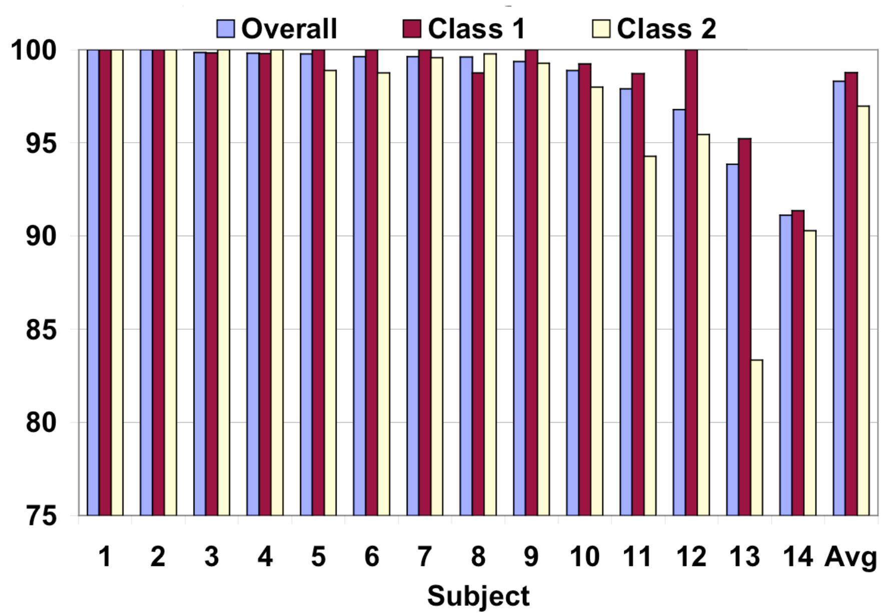

Figure 4. Classification accuracies for 14 subjects and the averages across subjects with 18 Hz band pass. Blue (light grey) bars show test set accuracies for overall classification; maroon (dark grey) bars are for Class 1 (alert) epochs, and yellow (white) bars are for Class 2 (fatigued) epochs.

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 weights in limited frequency bands in one subject (Fig. 6), we found that a broad set of fronto-central 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.

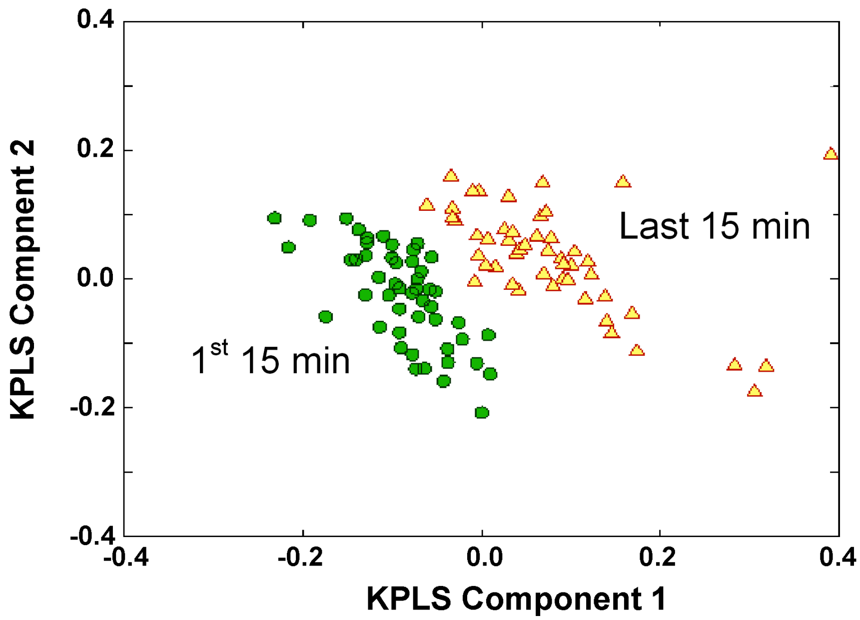

Figure 5. Example of KPLS scores predicted for single-trial EEG spectra for early (green/dark circles) and late (yellow/light triangles) blocks of the mental arithmetic task in one subject.

1.6. 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-minute 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.

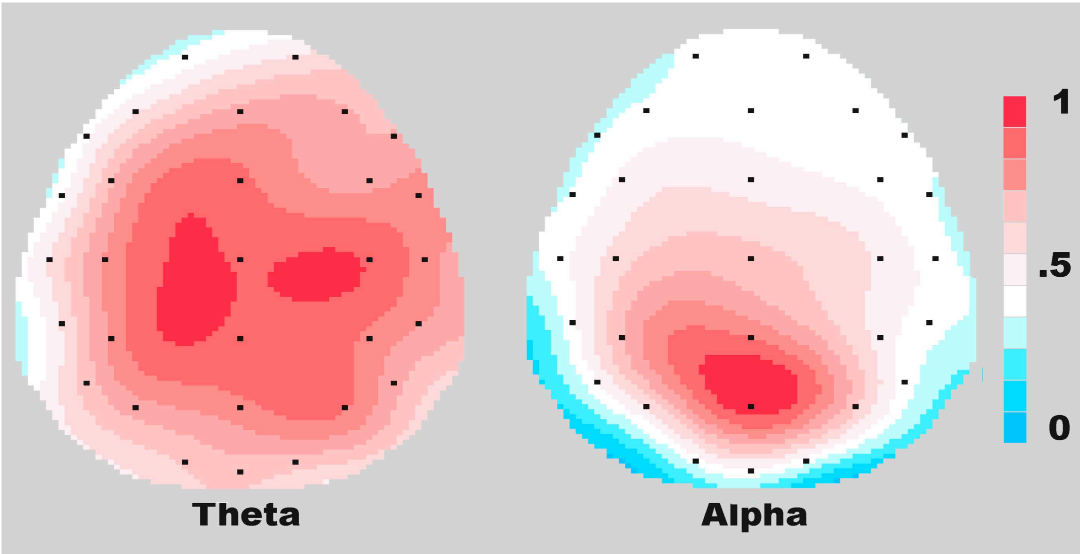

Figure 6. Topographical maps fit to the first KPLS component weights in theta and alpha bands for one subject (819). The colored areas are smoothed normalized absolute values, with the largest values in red and the smallest in blue.

2. DISCUSSION

2.1. Behavioral Measures

Time on task produce 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 influence moods. These effects support the assertion that our task produced a state of cognitive fatigue. Observed activity progressively increased while observed alertness progressively decreased over time. Moreover, there was a progressive, but moderate, slowing effect on response times. However, time on task did not influence response accuracy. Together, these results suggest that our subjects experienced cognitive fatigue, but did not sacrifice accuracy as may be expected if motivation had waned. The moderate, general increases in RT over time also indicate increasing cognitive fatigue, but not a severe increase as may be expected if lapses or sleep episodes had occurred frequently.

2.2. EEG and ERP Measures

The EEG analyses suggest that time on task had a progressive influence on frontal midline theta and parietal alpha activity. Both rhythms 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 cognitive fatigue, as much as they indicate specific increases in frontal midline theta and midline parietal alpha power. A detailed analysis of our classification results can provide more specific details for individual subjects, which we will report in the future.

2.3. ERP Measures

The ERP analyses suggest that time on task did not have substantial effects on N100, P200, and P300 amplitudes or latencies. The one exception was P200 amplitude, for which time on task had a curvilinear influence, with larger amplitudes in the middle of the task. There was no specific hypothesis about P200 and its sensitivity to fatigue in other contexts is poorly documented. The non-significant effect on P300 amplitudes was in the direction predicted by the "increasing workload" hypothesis. P300 amplitudes were larger during the periods of relatively high cognitive fatigue as compared to fresh performance.

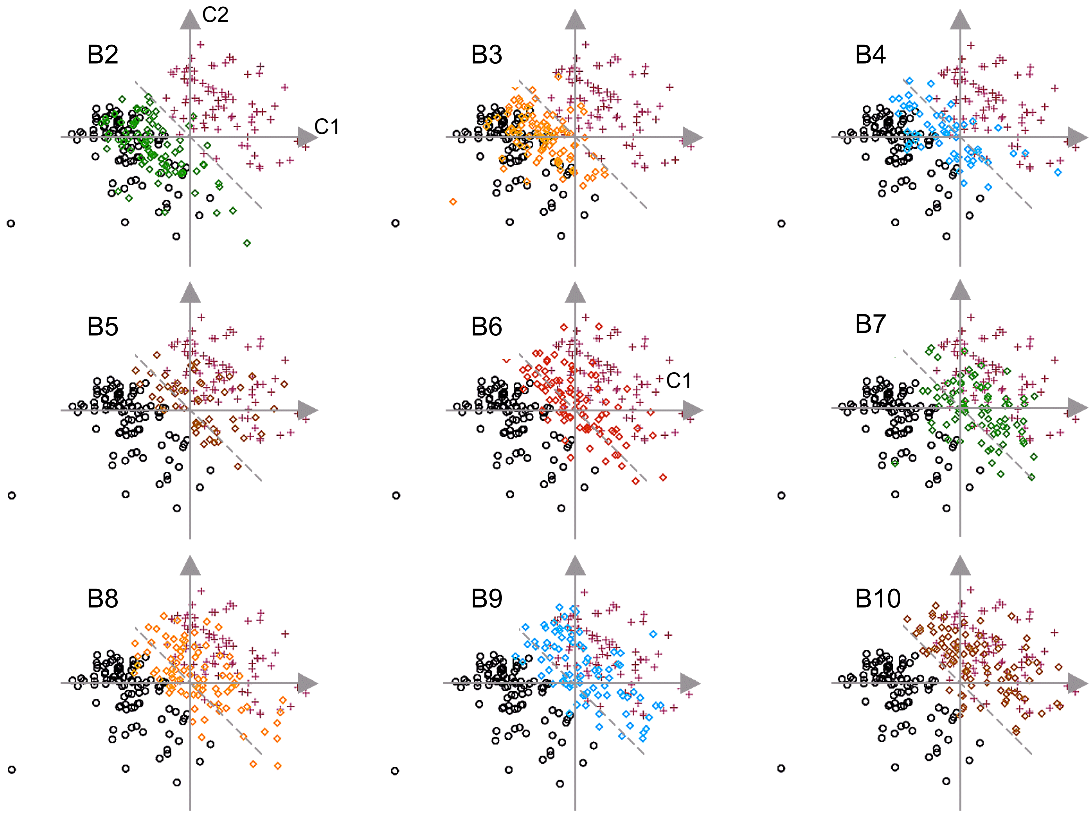

Figure 7. Example of estimating the development of fatigue over time in one subject (819). KPLS scores were predicted for EEG epochs from nine 15-minute blocks between the training set blocks (B1 & B12). Block 2 = 15-30 min. block 3 = 30-45 min, block 4 = 45-60 min, … , block 10 = 135-150 min. Black 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.

2.4. Classification and Prediction

KPLS-DLR classification of single trial EEG epochs was about 90% to 100% for with a mean of 97 to 98% depending on the low-pass cutoff. A small increase in classification accuracy appears to derive from including EEG in the 11-18 Hz range. The performance of these classifiers is highly accurate for single trials, and may serve as the basis for predictive models of cognitive fatigue in operational settings. Inspections of the predictive behavior of the KPLS models showed an orderly relationship of the scores to time on task and to correlated behavioral, subjective, and performance measures.

Future work will examine details of the KPLS models to describe individual frequency/electrode effects. For operational applications, we will also develop methods for minimizing the number of electrodes in the models, testing predictions of the models with new experiments, and developing adaptive statistical classifiers for on-line use.

3. REFERENCES

1. L. D. Montgomery, R. W. Montgomery and R. Guisado, "Rheoencephalographic and electroencephalographic measures of cognitive workload: analytical procedures," *Biol. Psychol.* **40**, pp. 143-59, 1995.
2. G. R. J. Hockey, "Compensatory control in the regulation of human performance under stress and high workload: A cognitive energetical framework." *Biol. Psychol.*, **45**, pp. 73-93, 1997.
3. S. Makeig, and M. Inlow, M, "Lapses in alertness: coherence of fluctuations in performance and the EEG spectrum," *Electroencephalogr. Clin. Neurophysiol.*, **86**, pp. 23-35, 1993.
4. A. F. Kramer, L. J. Trejo and D. Humphrey, "Psychophysiological measures of workload: Potential applications to adaptively automated systems." In R. Parasuraman & J. Mouloua, Eds., *Automation and Human Performance: Theory and Applications*," pp. 137-162, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.
5. D. G. Humphrey, A. F. Kramer and R. R. Stanny, "Influence of extended wakefulness on automatic and nonautomatic processing." *Human Factors*, **36**, pp. 652-669, 1994.
6. R.E. Thayer, "Activation-Deactivation Adjective Check List: Current overview and structural analysis." *Psychological Reports*, **58**, pp. 607-614, 1986.
7. B. S. Plake, J . C. Impara, and R. A. Spies, Eds., *The fifteenth mental measurements yearbook.* Buros Institute of Mental Measurements, Lincoln, NE, 2003.
8. P.D. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms." *IEEE Trans. Audio Electroacoust.* **AU-15**, pp. 70-73, 1967.
9. R. Rosipal, L. J. Trejo, B. Matthews B. Kernel PLS-SVC for Linear and Nonlinear Classification. In T. Fawcett and Nina Mishra, Eds., *Proc. of the Twentieth International Conference on Machine Learning*, pp. 640-647, Washington DC, Aug. 21-24, 2003.