



EEG-derived estimators of present and future cognitive performance

Maja Stikic^{1*}, Robin R. Johnson¹, Daniel J. Levendowski¹, Djordje P. Popovic¹, Richard E. Olmstead² and Chris Berka¹

¹ Advanced Brain Monitoring Inc., Carlsbad, CA, USA

² Semel Institute, University of California Los Angeles, Los Angeles, CA, USA

Edited by:

Alvaro Pascual-Leone, Harvard Medical School, USA

Reviewed by:

Marine Vernet, Harvard Medical School, USA

Carlo Miniussi, University of Brescia, Italy

*Correspondence:

Maja Stikic, Advanced Brain Monitoring, Inc., 2237 Faraday Avenue, Suite 100, Carlsbad, CA 92008, USA.
e-mail: maja@b-alert.com

Previous electroencephalography (EEG)-based fatigue-related research primarily focused on the association between concurrent cognitive performance and time-locked physiology. The goal of this study was to investigate the capability of EEG to assess the impact of fatigue on both present and future cognitive performance during a 20-min sustained attention task, the 3-choice active vigilance task (3CVT), that requires subjects to discriminate one primary target from two secondary non-target geometric shapes. The current study demonstrated the ability of EEG to estimate not only *present*, but also *future* cognitive performance, utilizing a single, combined reaction time (RT), and accuracy performance metric. The correlations between observed and estimated performance, for both present and future performance, were strong (up to 0.89 and 0.79, respectively). The models were able to consistently estimate “unacceptable” performance throughout the entire 3CVT, i.e., excessively missed responses and/or slow RTs, while acceptable performance was recognized less accurately later in the task. The developed models were trained on a relatively large dataset ($n = 50$ subjects) to increase stability. Cross-validation results suggested the models were not over-fitted. This study indicates that EEG can be used to predict gross-performance degradations 5–15 min in advance.

Keywords: cognitive performance estimators, electroencephalography, combined performance metric, fatigue, sustained attention tasks

INTRODUCTION

The management of fatigue is a serious public health and safety concern, as impaired vigilance is a primary contributor in many transportation and industrial accidents every year. Safety, efficiency, productivity, and liability are all impacted by employee alertness. Fatigue-related accidents and decreased productivity associated with drowsiness are estimated to cost the U.S. over \$77 billion each year, with annual costs amounting to over \$377 billion worldwide (Corburn, 1997). The American Academy of Sleep Medicine reports that one in every five serious motor vehicle injuries is related to driver fatigue, with 80000 drivers falling asleep behind the wheel every day, and 250000 sleep-related accidents every year (Drowsy Driving Fact Sheet, 2009). Recent research suggests that sleep deprivation also poses a serious risk in the medical community, with one report showing that medical residents with less than 4 h of sleep a night made more than twice as many errors as residents who slept for more than 7 h a night (Baldwin and Daugherty, 2004). The effects of even small amounts of sleep loss each night accumulate over time, resulting in a “sleep debt” – as sleep debt increases, alertness, memory, and decision-making are increasingly impaired (Kribbs and Dinges, 1994; Dinges et al., 1997; Weaver et al., 1997; Van Dongen et al., 2003). Importantly, individuals become habituated to this chronic accumulation of fatigue, and are often unaware of its negative impact on their performance (Dinges, 2004; Balkin et al., 2008). Consequently, subjective estimates of fatigue are unreliable, and individuals are typically unable to self-assess the level of risk for driving or other activities when fatigued. Thus, objective

metrics of fatigue are required to evaluate risk for performance deficits that may lead to accidents, either while driving or in the workplace.

One approach that has been used to provide an objective assessment of error/accident risk is assessing neuropsychological performance prior to beginning a driving or other work shift. Neuropsychological tests of continuous performance and reaction time (RT) have proven useful as a convenient and inexpensive method for quantifying speed and accuracy of responses, as well as ability to sustain attention over time (Rosvold et al., 1956; Dinges and Powell, 1985; Wilkinson, 1990; Weinstein et al., 1999). Performance measures derived from these tasks have proven sensitive to changes in alertness as a result of acute sleep deprivation, cumulative sleep loss, or daytime drowsiness in sleep deprived (SD) patients (Dinges et al., 1997; Dinges and Weaver, 2003; Baulk et al., 2008). These measures, however, must be taken prior to the work/driving shift, and thus fail to address the likely increase in risk that develops over the course of the following 8–12 h. To address this timing issue, many researchers (Serman et al., 1992; Makeig and Jung, 1995; Berka et al., 2004) have begun to integrate neuropsychological testing with integrated electroencephalography (EEG) that can continue to monitor the worker throughout the shift and identify risk elevation over time.

Significant correlations between EEG indices of cognitive state changes and objective performance measures have been seen in studies conducted in laboratory, simulation, and operational environments (Brookhuis and de Waard, 1993; Makeig, 1993; Serman and

Mann, 1995; Brookings et al., 1996; Makeig and Jung, 1996; Gevins et al., 1997, 1998; Pleydell-Pearce et al., 2003; Wilson, 2004). Thus, the EEG-based models of drowsiness detection show great promise in providing a means to predict and, ultimately, prevent workplace accidents due to sleep loss. Error prediction has been studied recently, but most of the resulting algorithms (Balkin et al., 2002; Rajaraman et al., 2008) are insufficiently accurate on time scales useful for detection in field applications (i.e., time scales of days rather than hours/minutes). In order to provide a useful tool that can predict significant performance decrements, any predictive solution must identify the states that precede the onset of increased errors or slowed RT far enough in advance to allow for an appropriate intervention or mitigation, but close enough to the likely error to ensure that the risk profile does not change significantly in the interim, leading to increased risk over time or excessive false negatives.

In addition to setting the lead time window appropriately, a successful algorithm must also address: (1) the variability in performance that occurs as individuals choose alternative strategies for coping with sleep loss; (2) the feasibility of the algorithm for field use (e.g., while driving); and (3) the computational needs of the algorithm to ensure that the system can be field deployed. Typically, a speed–accuracy tradeoff occurs during sleep deprivation (Forstmann et al., 2008; Van Veen et al., 2008; Bogacz et al., 2009; Carp et al., 2010; Eichele et al., 2010). The tradeoff may involve slowed decision-making in order to maintain accuracy, or maintaining speed while allowing accuracy to be compromised. Thus, both accuracy and speed must be addressed when evaluating performance. The pattern of performance deficits observed may also be a function of motivation, task-related instruction provided, perceived rewards for speed or accuracy, or trait differences (Doran et al., 2001; Gray and Watson, 2002; Van Dongen et al., 2007; Hsieh et al., 2010). In order to accommodate individual differences, the models must account for subtle individual changes in the EEG data, as well as the different types of performance degradation. Most of the individualized models are based on person-dependent training (Olofsen et al., 2004; Van Dongen et al., 2007; Rajaraman et al., 2008). Although such models can provide highly accurate classifications of cognitive state, they may prove impractical in field environments (e.g., for drivers) due to the additional time required for personnel to obtain the training data, as well as the computer processing time and capacity needed to train the models. Lastly, several algorithms (Smith et al., 2002; Liang et al., 2005; Lin et al., 2006) for alertness/drowsiness detection require 20+ EEG channels, requiring processing power not yet available in a portable modality for use in industry or on the road.

This study investigates core elements of an alternative approach for managing workplace and driving safety by using continuous monitoring of neurophysiology and associated estimators of future performance. This effort expands upon previous research designed to associate time-locked associations between physiology and performance (Makeig and Jung, 1996; Trujillo et al., 2009). Fatigue resulting from a lack of sleep was assessed through the subjects' cognitive performance on a sustained attention task, and performance decrements were estimated by an EEG-based algorithm. The proposed method aims to overcome the shortcomings of the existent solutions, in that: (1) it is able to predict increased risk of errors 5–15 min in advance, limiting the lead time to a period that ensures accuracy and time for mitigation; (2) individual coping strategies

are compensated through the development of a single combined cognitive performance metric; (3) the prediction algorithm is trained in a person independent setting, but it is still individualized through self-normalization of the EEG data, ensuring that individual variability in the EEG is accommodated; (4) the model is trained on a large sample size, and evaluated across all subjects to ensure generalizability; and (5) the method is not computationally expensive, making it suitable for deployment in the field.

MATERIALS AND METHODS

STUDY PROTOCOL

A total of 65 subjects were recruited for a 3-day study, including a night of partial sleep deprivation. The protocol was approved by the independent IRB, Biomedical Research Institute of America (San Diego, CA, USA). Prior to enrollment, subjects were screened by questionnaires to exclude those likely to have a neurological, sleep, or psychiatric disorder. Further exclusion criteria included self-report of excessive daytime sleepiness (Epworth >6), excessive smoking (more than 10 cigarettes/day) or caffeine intake (more than 5 cups/day), head trauma, and inconsistent sleep patterns (<7.25 h/night on average). Individuals who qualified signed a consent form after receiving instructions for participation.

Participants were asked to complete three sessions: baseline (BL), fully rested (FR), and SD. The BL session was scheduled to start 3 h after the subject typically awoke based on self-report (i.e., between 8 a.m. and 10 a.m.). The FR and SD sessions were conducted on consecutive days, and began at the same time as the BL session. Three tasks (described below) were initially administered in a single 2 h cycle during the BL session, and repeated in four 2 h cycles during both FR and SD sessions. The four cycles began at approximately 0900, 1100, 1300, and 1500. Wrist actigraphs were used to ensure subjects obtained at least 6 h of sleep each night in the 4-days leading up to their BL and FR sessions, and to ensure their compliance with the limited sleep requirement on the night prior to the SD sessions. In the night between the FR and SD day, subjects were required to remain awake 2 h beyond their typical bedtime, and to sleep only 3.5 h total, and leave telephone voice messages every 30 min of the required wake time. As a safety precaution, subjects were provided transportation to and from the study site for the SD sessions. A total of 50 participants were selected for the model development data set. The other subjects were eliminated due to: (a) insufficient or poor sleep the night before the BL and/or FR data collection, (b) signs of sleepiness during BL and/or FR sessions, (c) excessively poor performance on tasks, or (d) failure to leave the phone messages during the night of partial sleep deprivation (indicating compliance failure).

Three tasks were repeated across each of the 3 days of the study (one cycle for BL, four cycles each for FR and SD): visual passive vigilance task (VPVT), auditory passive vigilance task (APVT), and 3-choice active vigilance task (3CVT). All tasks were administered using a PC workstation. During the 5-min of VPVT and APVT, subjects were required to press the space bar on the keyboard every 2 s. Subjects were prompted to maintain the 2-s time intervals by a 10-cm diameter red circle that appeared in the center of the monitor during the VPVT, or by an audio tone during the APVT. The 3CVT required subjects to discriminate one primary target (presented 70% of the time) from two secondary non-target geometric shapes that were randomly interspersed over a 20-min test period, for a

total of 376 trials. Each trial consisted of a single stimulus type (target or non-target) presented for 0.2 s. The task comprised four consecutive quartiles (Q1–Q4), with different inter-stimulus intervals to increase difficulty of the task, and to challenge the subject's ability to sustain attention. The inter-stimulus interval varied by up to 10 s, and increased from 1.5 to 3 s in the first 5 min (Q1), to 1.5–6 s in the next 10 min (Q2 and Q3), and, lastly, to 1.5–10 s in the final 5 min (Q4). Participants were instructed to respond as quickly as possible to each stimulus presented by selecting the left arrow to indicate target stimuli, and the right arrow to indicate non-target stimuli. Performance measurements included RT to stimulus and accuracy (i.e., percent of the correct responses). A brief training period was provided prior to the start of the testing period to minimize practice effects for the 3CVT.

DATA RECORDING AND SIGNAL PROCESSING

The B-Alert sensor headset (Advanced Brain Monitoring Inc., Carlsbad, CA, USA; Berka et al., 2007; Johnson et al., 2011) was used to acquire the EEG data from three referential channels (Fz, Cz, and POz) and two bipolar channels (Fz–POz and Cz–POz). The sampling rate was 256 samples/s for all channels. Proprietary data acquisition software stored the EEG data on the host computer. The software additionally incorporated performance outputs from the performed sustained attention tasks into the EEG record. Thus, the EEG data was synchronized with the stimuli/response events from the neuropsychological assessment.

The EEG signals were filtered with a band-pass filter (0.5–65 Hz) before the analog-to-digital conversion. In order to remove environmental artifacts from the power network, sharp notch filters at 50, 60, 100, and 120 Hz were applied. The algorithm (Berka et al., 2007) automatically detected and removed a number of artifacts in the time-domain EEG signal, including spikes caused by tapping or bumping of the sensors, amplifier saturation, and excursions that occur during the onset or recovery of saturations. Eye blinks and excessive muscle activity were identified and decontaminated by wavelet transform (Berka et al., 2007).

From the decontaminated EEG signal, features were extracted on a second-by-second basis and then averaged for each quartile of the 3CVT. Decontaminated EEG signal was segmented into 50% overlapping, 256 data-point windows called overlays. For each 1 s epoch, three consecutive overlays were used to compute the absolute and relative power spectral density (PSD) values. Each overlay was multiplied by the Kaiser window ($\alpha = 6.0$) to reduce the edge effect when applying 256-point fast Fourier transformation (FFT). The FFT was averaged on the three successive overlays to decrease epoch-by-epoch variability. For each channel, the absolute PSD values were derived for each 1 Hz bin from 1 to 40 Hz, and transformed to a logarithmic scale. Relative PSD values were derived by subtracting the logged PSD for each 1 Hz bin from the total logged PSD power in the range from 1 to 40 Hz. Overall, this sums up to 80 EEG-based features per channel.

To normalize the data for individual variability, the absolute and relative PSD quartile data obtained from the 3CVT sessions on the FR day and the SD day were *z*-scored to the respective quartile data obtained during the BL 3CVT session. This accommodated for individual differences in the EEG data inherent in drowsiness detection.

In order to explore the applicability of alertness quantification in performance estimation, we also incorporated the outputs of the B-Alert model (Berka et al., 2007; Johnson et al., 2011) into the feature vector. The B-Alert model is an individualized model that classifies a subject's cognitive state into one of the four levels of alertness (sleep onset, distraction/relaxed wakefulness, low engagement, and high engagement). It utilizes the absolute and relative PSD values during VPVT, APVT, and Q1 of the 3CVT BL data to derive coefficients for a discriminant function that generates classification probabilities for each 1 s epoch. The resulting model was used to classify the 3CVT data from the four FR and four SD sessions (Johnson et al., 2011). The output probabilities of the B-Alert model were then also averaged for each quartile of the 3CVT, and added to the feature vector for further analysis.

PERFORMANCE METRIC

To measure speed of processing and number of lapses during the 3CVT, typical performance metrics used for the neuropsychological assessment of the sustained attention have been utilized: RT to stimulus, and *accuracy* (i.e., percent of the correct responses). Individual response patterns due to neurobehavioral impairment vary in one of three primary ways: (1) both RT and accuracy deteriorate; (2) RT is preserved by sacrificing accuracy; or (3) RT is sacrificed to preserve accuracy. In order to account for this, we combined RT and accuracy in an aggregated performance score called *F-Measure*. *F-Measure* is often used in information retrieval as a unique performance score of the classifier. It is the harmonic mean of sensitivity and positive predictive value (PPV):

$$F - \text{Measure} = 2 \cdot \frac{\text{Sensitivity} \cdot \text{PPV}}{\text{Sensitivity} + \text{PPV}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

where TP is a number of true positives, FN is a number of false negatives, and FP is a number of false positives. As there is a typical tradeoff between number of false negatives and false positives, classifiers typically have high sensitivity and low PPV, or vice versa. As these two measures have the same range (i.e., they vary between 0 and 1), they are weighted equally through *F-Measure*.

Performance metrics during the 3CVT have different ranges. Typical RT range is between 0.4 and 1.5 s, and percent of the correct responses varies between 0 and 100%. Thus, utilizing the raw performance values would have different effects on *F-Measure*, i.e., small changes in RT would be equivalent to significantly larger changes in percent of the correct responses. To equalize their influence, both RT and percent of the correct responses were linearly scaled to the same range (0,1) in the experiments. Reversed scoring of RT is utilized to form speed of response, which is then combined with accuracy to create a new unique metric that decreases with performance decrements. Thus, in our work, the *F-Measure* is defined as the harmonic mean of the scaled values of reversed RT (i.e., RT_s) and percent of correct responses (i.e., accuracy; PC_s):

$$F - \text{Measure} = 2 \cdot \frac{RT_s \cdot PC_s}{RT_s + PC_s} \quad (3)$$

The scaling was performed in the following manner:

$$RT_s = \frac{1.5 - RT}{1.5 - 0.4} \quad (4)$$

$$PC_s = \frac{PC}{100} \quad (5)$$

where RT and PC are the raw values of the utilized performance metrics.

Based on the *F*-Measure, two classes of performance are defined: *unacceptable performance* and *acceptable performance*. **Figure 1A** presents the mean and STD of *F*-Measure for each quartile of the BL sessions, averaged over all subjects ($n = 50$). The mean BL *F*-Measure,

averaged over all four quartiles, is 0.88, and its STD is 0.04. The *F*-Measure threshold that defines unacceptable (F -Measure < 0.80) and acceptable (F -Measure ≥ 0.80) performance was selected as 2 STD below the mean BL *F*-Measure value (i.e., $0.88 - 2 \times 0.04 = 0.8$). The chosen threshold between acceptable and unacceptable performance represents the 35th percentile, i.e., overall, 35% of the FR and SD *F*-Measure values (**Figure 1B**) were in the defined range of unacceptable performance. Average RT and average accuracy, with their STD for unacceptable performance, were 0.87 ± 0.14 s and $81.1 \pm 17.9\%$, respectively. In the case of acceptable performance, average RT and accuracy were 0.62 ± 0.06 s and $96.6 \pm 3\%$.

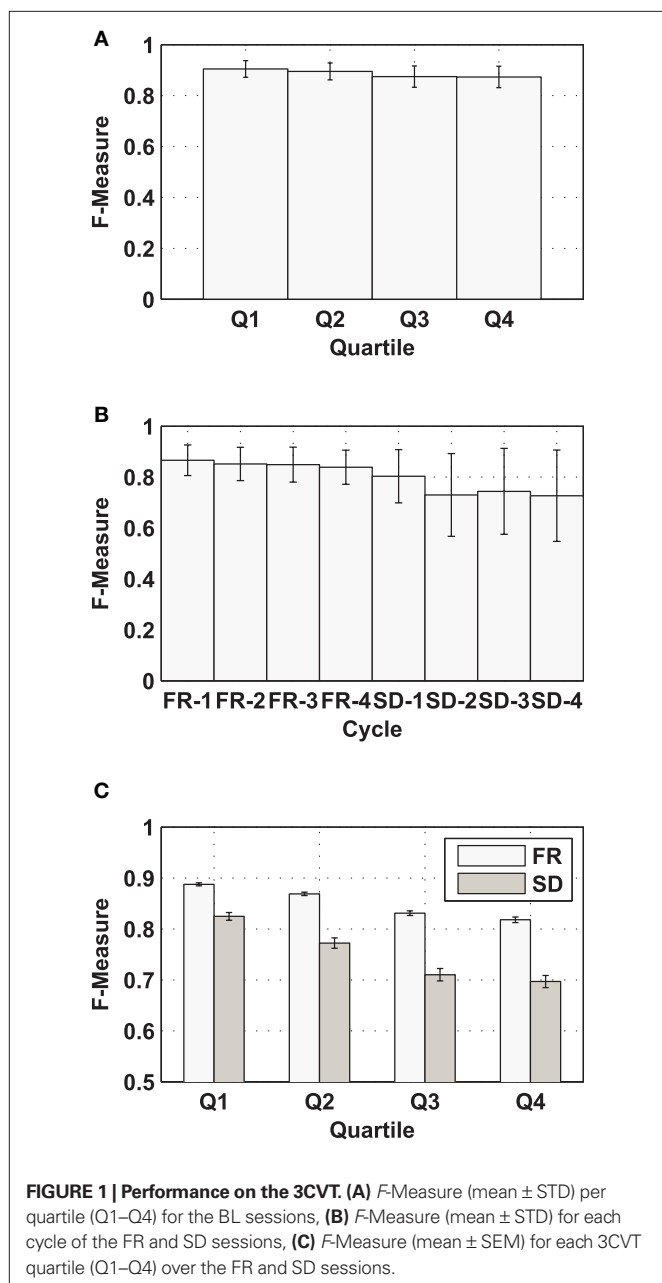
ALGORITHM DEVELOPMENT AND OUTCOME ANALYSES

Two analyses were conducted to assess the capabilities of EEG metrics to estimate cognitive performance by developing and comparing two different EEG-based models. The first approach aimed at estimating *present performance*, and the second approach estimated *future performance*. Both methods estimate the *F*-Measure during the quartiles of the 3CVT. The other tasks, VPVT and APVT, are not designed to assess RT of subjects, and data from these tasks were used only for the B-Alert model. The algorithm utilized the EEG data recorded during the FR and SD sessions. The VPVT, APVT, and 3CVT EEG data recorded during the BL session were used for training of the B-Alert model and, additionally, the 3CVT EEG data from the BL session were utilized for normalization purposes (i.e., z-scoring) of the 3CVT EEG data during FR and SD sessions.

The following steps were performed during the development of our EEG-based algorithm to recognize unacceptable or acceptable cognitive performance: feature selection, linear regression, and grouping into classes.

The EEG PSDs and B-Alert probabilities from the 3CVT quartiles were used to estimate present (i.e., Q_i to Q_i , $i = 1..4$) and future (Q_1 to Q_i , $i = 2..4$) performance for each 3CVT session. As a large number of features were extracted from the EEG data, statistical model selection procedure was applied to select the most predictive variables. Step-wise regression analysis was utilized to down-select from 400 EEG data values per quartile (i.e., 40 absolute and 40 relative PSD values from each of five EEG channels) and mean B-Alert probabilities, those variables most predictive in estimating the performance on the 3CVT defined by *F*-Measure (using data from both FR and SD data collections). In each step of step-wise regression analysis, a set of *F*-tests was performed as the selection criteria to determine the explanatory power of variables and select which variables to include, and which to exclude from the model. The maximum *p*-value for a variable to be added to the model was 0.05, and the minimum *p*-value for a variable to be removed from the model was 0.1. Furthermore, the singularity criterion was utilized during each step of feature selection to preclude the entry of the variables whose squared correlation with the variables already in the model exceeds a certain value ($p = 10^{-8}$).

The predictive variables were then used to fit a linear regression model to estimate the *F*-Measure. Thus, in all following experiments, *F*-Measure is the dependent variable, and the variables selected by step-wise analysis are used as independent variables. Furthermore, all observations in the dataset and selected variables were equally treated without usage of weighting or priors. The models were



trained using the least squares approach. This minimized the sum of squared distances between the observed and the estimated performance (i.e., *F*-Measure) by the linear approximation.

As the aim of the algorithm is to detect cognitive performance decrements, the estimated *F*-Measure was then stratified into the target class of unacceptable performance and the non-target class of acceptable performance using a threshold of 0.80. In this way, the accuracy of the model with respect to sensitivity, specificity, and positive and NPV was determined. Furthermore, a sensitivity analysis was conducted to assess the impact of borderline cases when the classes of estimated and observed performance disagreed, but the observed *F*-Measure was within 0.05 or 0.075 of the threshold.

Figure 2 shows the input and output for both the present performance and future performance models. To assess present performance, four regression equations, one for each quartile (Q1–Q4), were developed using variables from each respective quartile that were most predictive in estimating performance on that same quartile across the FR and SD sessions:

$$f_{PP-Q_i} : EEG_{Q_i} \rightarrow F\text{-Measure}_{Q_i}, i = 1..4 \quad (6)$$

The second model estimates performance on Q2–Q4 based on the EEG data from Q1 (i.e., estimating future performance). Thus, three equations were developed using EEG variables from Q1 to estimate performance during the remaining three quartiles Q2–Q4:

$$f_{FP-Q_i} : EEG_{Q_1} \rightarrow F\text{-Measure}_{Q_i}, i = 2..4 \quad (7)$$

The trained linear regression models were evaluated in two different ways, by auto-validation during the model development, and by cross-validation. Auto-validation entails testing the model on the data on which it was trained. Cross-validation assesses the generalization capabilities of the model by testing it on the data that was not used for training. In order to examine the feasibility of using the subject independent training, we performed leave-one-subject-out cross-validation of the linear regression models. First, the most explanatory variables were selected through step-wise analysis including all subjects' FR and SD data. Afterward, in each cross-validation round, new regression coefficients for the previously selected variables were derived with one subject removed from the data set, and then the resulting regression model was applied to the removed subject's data. The procedure was repeated for each subject in the study, and the results averaged across all cross-validation rounds.

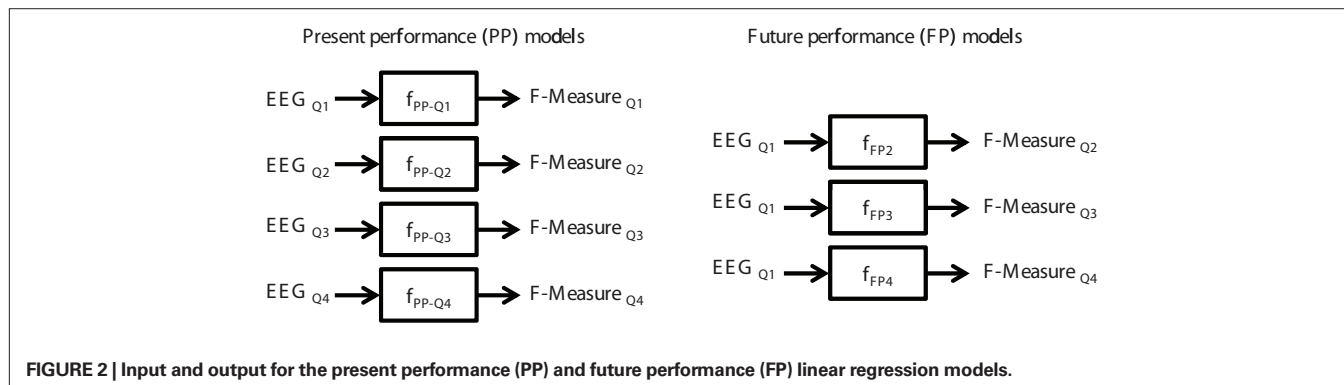
RESULTS

In this section the following results are presented: (1) performance on each of the 3CVT quartiles over the FR and SD sessions and its sensitivity to sleep loss, (2) the most predictive variables selected by step-wise analysis, (3) evaluation of the trained linear regression models that estimate *F*-Measure, (4) prediction accuracy of grouping the estimated *F*-Measure values into the previously defined classes of unacceptable and acceptable performance, and (5) analysis of the borderline cases.

COGNITIVE PERFORMANCE ON THE 3CVT

The study protocol included a night of partial sleep deprivation. In order to assess the sensitivity of the subjects' performance on the 3CVT to sleep loss, a statistical test was utilized to compare performance decrements (collapsed over all four quartiles of the 20-min task) over the four 2 h cycles for each condition. A 2 (FR, SD) \times 4 (cycles) repeated measure ANOVA showed a significant main effect of condition (i.e., FR vs. SD) on performance (i.e., observed *F*-Measure), $F(1,98) = 34.177, p < 0.0001$, but no interaction effect or main effect of cycle. *Post hoc* analysis indicated that sleep loss was associated with poorer performance at each time point (i.e., cycle).

In contrast, performance (i.e., the observed *F*-Measure) decrements within the 20-min long 3CVT were in evidence. **Figure 1C** shows *F*-Measure (mean \pm SEM) on each of the 3CVT 5 min quartiles for the FR and SD sessions (collapsed across cycles). A 2 (FR, SD) \times 4 (quartile) analysis revealed an interaction of condition \times quartile [$F(3,396) = 16.56, p < 0.0001$], as well as main effects of condition [$F(1,398) = 91.85, p < 0.0001$] and quartile [$F(3,396) = 146.96, p < 0.0001$]. Furthermore, performance decrements between FR and SD sessions were larger on later quartiles of the task (i.e., Q3 and Q4 – the last 10 min of the 20-min task). This is presumably due to the inability to sustain attention over longer periods of time when SD. Lastly, there was more variance in performance during the SD sessions compared to the FR sessions. We further investigated the tendency of the performance decrement within the same day by examining the cycle's effect on the last quartile (Q4) of the 3CVT (i.e., the quartile on which the largest drop in performance was seen). This analysis mirrored the statistical analysis over all quartiles in that a significant effect of condition was found [$F(1,198) = 2.59, p < 0.0001$], but not for cycle ($p > 0.05$). Taken together, these analyses indicate that performance decrements occur within the 20-min 3CVT, but not over repeated cycles throughout the day.



VARIABLE SELECTION RESULTS

Table 1 shows the variables selected by step-wise analysis for estimation of the present performance for each of the 3CVT quartiles. Similarly, **Table 2** shows the most predictive variables for estimation of the future performance, i.e., estimation of *F*-Measure on Q2–Q4 based on the data from Q1. The amount of variance explained with these variables ranges between 67 and 80% for the present performance, and between 53 and 63% for the future performance.

Most of the selected EEG features for estimation of the present performance came from the Fz–POz and Cz–POz channels. For the future performance estimation, the most informative features were found in the Fz–POz channel and the POz channel. Analysis of the selected EEG features per frequency bands showed that the majority of the selected features were from the Delta and Alpha frequency bands in both cases.

The B-Alert classification probabilities for Sleep Onset, Distraction, and Low Engagement were selected for estimation of the present performance. Even though the B-Alert model was originally developed for drowsiness detection (Berka et al., 2004), it also proved useful for drowsiness prediction (i.e., estimation of

the future performance decrements), as the B-Alert classification probabilities for Sleep Onset and Distraction were selected by step-wise regression for the future performance model.

REGRESSION RESULTS

Figure 3 provides regression plots (i.e., the observed and estimated *F*-Measure values) for the model that estimates present performance for all four quartiles across the FR and the SD sessions. Both the model development (left) and cross-validation (right) results are shown. The relationship between the observed and estimated performance was examined using Pearson's correlation analysis. The Pearson correlations for the model development were strong (greater than 0.80) and relatively consistent across the quartiles. Performance on Q2 and Q3 ($r = 0.89$ and $r = 0.85$, respectively) was slightly better than on Q1 and Q4 ($r = 0.83$ and $r = 0.82$, respectively). Cross-validation confirmed this pattern of results. The goodness of fit of the cross-validation regression models did not decrease significantly (10% on average), and it was still in a moderate to strong range (i.e., $r^2 [0.58–0.70]$).

Table 1 | List of variables used to estimate present performance by quartile.

Channel	Absolute/Relative PSD	Q1 Hz-Bins	Q2 Hz-Bins	Q3 Hz-Bins	Q4 Hz-Bins
Fz–POz	Absolute	3,8,10,21,28,33,39	1,2,4,10,23,33	8,10,12,16,32,36	25,31,32
	Relative	3,8,10,24,25,26	3,7,8,10,27,30	6,8,9,10,12,16,19,39	40
Cz–POz	Absolute	–	5,32,33,35,39,40	2,3,14,34	18,22,35,37
	Relative	2,16,18,19	11,14	9,11,30,31,32	7,8
Fz	Absolute	20,24,25,40	7,14	17,26	7,12,19
	Relative	8,29,39	4,9,31,37,38,39	9,12	23
Cz	Absolute	2,3,21	1,10,11,33	2,10	14,19,22
	Relative	1,36	16,27,32	28	26
POz	Absolute	1,3,36	1,6,9	1	16,17,34
	Relative	15,26	2,13	6,9,12,16,33,37	9
B-Alert probabilities		Sleep onset	Sleep onset	Sleep onset	Sleep onset
		Distraction		Low engagement	

Table 2 | List of variables from quartile 1 (Q1) used to estimate future performance by quartile.

Channel	Absolute/Relative PSD	Q2 Hz-Bins	Q3 Hz-Bins	Q4 Hz-Bins
Fz–POz	Absolute	16,23	16,26,34,39	3,8,10,11,25,26
	Relative	3,5,6,9	3,9,11	9,12
Cz–POz	Absolute	34,35	6	1,2,7
	Relative	11,37	37	–
Fz	Absolute	–	–	7,10
	Relative	7,36	9	–
Cz	Absolute	2,6	3	3,4,13,14
	Relative	14,20,23,36	20	11
POz	Absolute	11,15,17	10	15,16,34
	Relative	1,13,16	1,2,11,13	1,10,13,35,37
B-Alert probabilities		Sleep onset	Sleep onset	–
		Distraction		

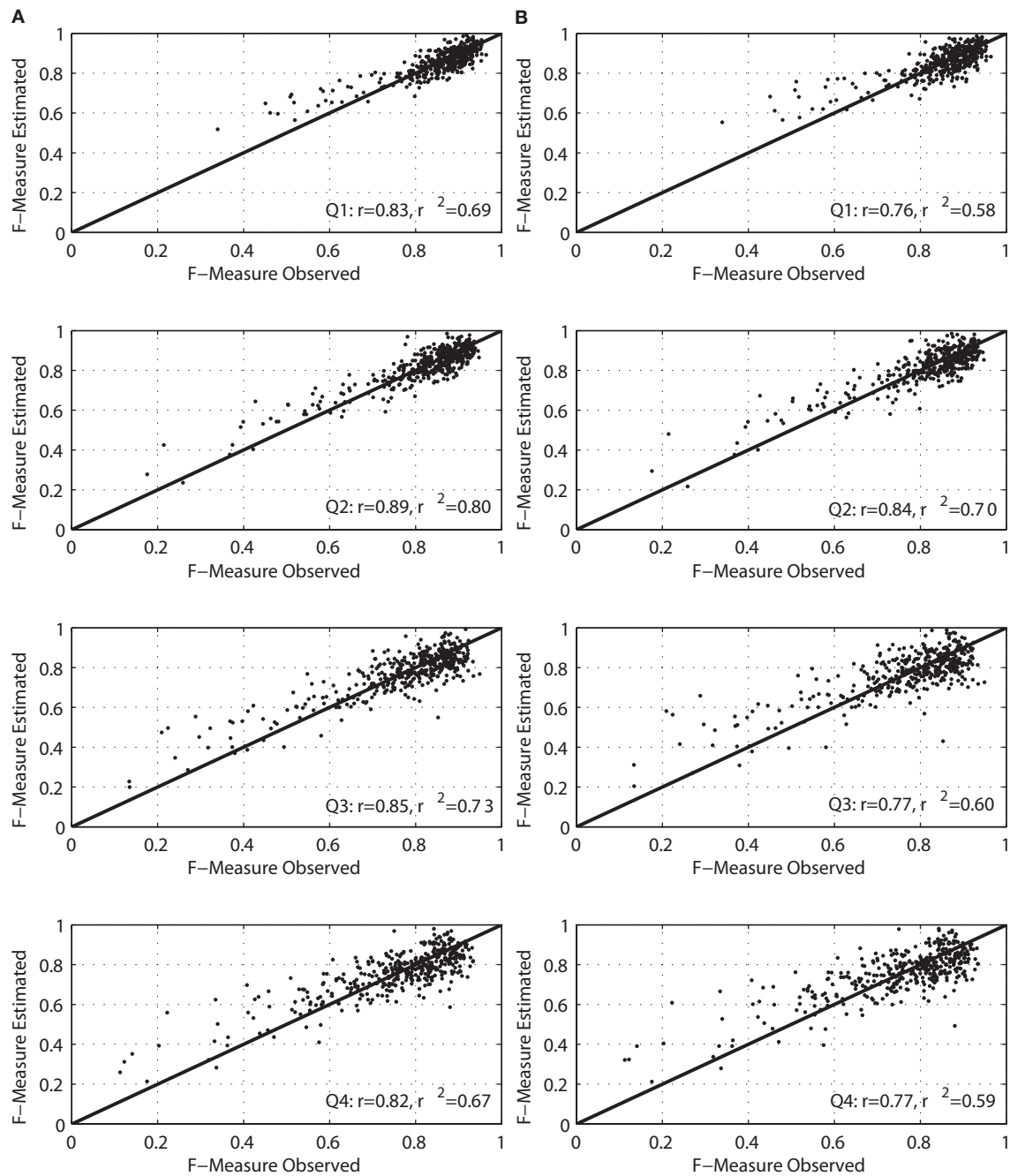


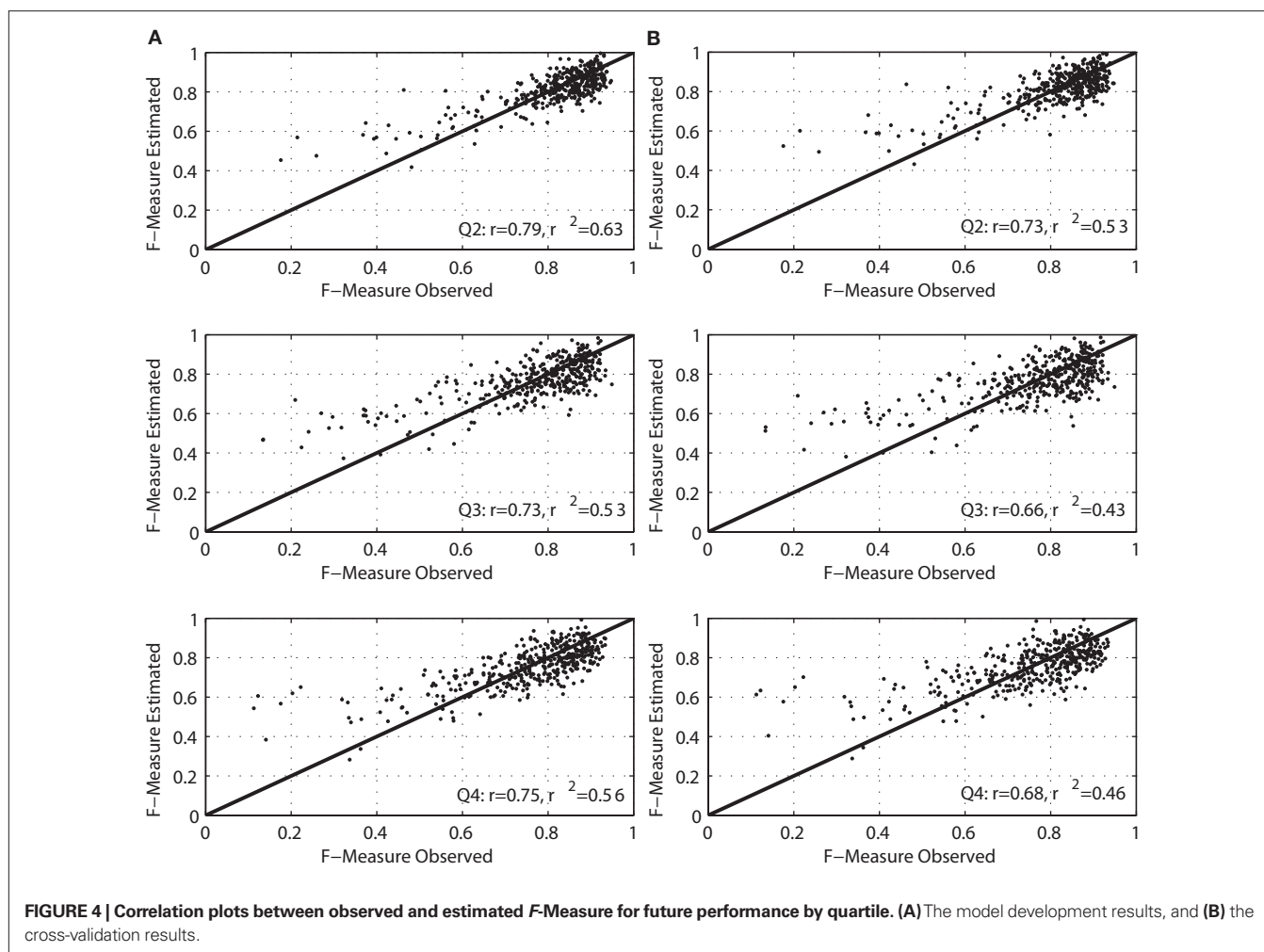
FIGURE 3 | Correlation plots between observed and estimated *F*-Measure for present performance by quartile. (A) The model development results, and (B) the cross-validation results.

Similarly, in **Figure 4**, regression plots for the model that estimates future performance (i.e., *F*-Measure on Q2–Q4 based on the EEG data from Q1) are shown. Again, both the model development (left) and cross-validation (right) results are shown. In this case, the Pearson correlations between observed and estimated performance were slightly weaker for the model development ($r = 0.79, 0.73, \text{ and } 0.75$ for Q2, Q3, and Q4, respectively) and greater than 0.65 for the cross-validation models. The capability of the EEG

variables to account for performance variability across subjects was relatively compromised in the case of the cross-validation models (approximately 10%).

PREDICTION ACCURACY

After grouping the observed and estimated *F*-Measure values into the classes of unacceptable (target class) and acceptable (non-target class) performance using the threshold of 0.8, the following



measures were calculated: sensitivity, specificity, positive and negative predictive value with the 95% confidence intervals for both the present and future performance models. The results are shown in **Tables 3 and 4**. Again, the model development and cross-validation results are displayed.

The present performance models (**Table 3**) tended to more consistently estimate unacceptable performance during all four quartiles of the 3CVT [i.e., relatively similar sensitivity ($\approx 83\%$) and PPV ($\approx 70\%$)], while acceptable performance was recognized less accurately during later quartiles of the 3CVT (i.e., specificity and NPV decreased later in the task by 28 and 20%, respectively).

The results of the EEG models that estimate future performance are presented in **Table 4**. Difficulties in recognizing acceptable performance later in the task, apparent during estimation of present performance, also occurred during estimation of future performance. Interestingly, in this case, estimation of unacceptable performance was improved later in the task (Q3 and Q4), and the model that estimates future performance occasionally obtained higher sensitivity and PPV than the present performance model.

Overall, the algorithm achieved better results when the cognitive performance (i.e., *F*-Measure) was poorer, i.e., future performance decrements were better estimated in Q3 and Q4 than in Q2. This is presumably due to the fact that the lower values of *F*-Measure

were better distinguished from acceptable performance in the EEG spectral content and, as *F*-Measure was decreasing over the course of the 3CVT (see **Figure 1C**), it was better recognized during the later quartiles of the task.

The differences between the model development and cross-validation results were notably small, suggesting that the models were not over-fitted. For the model that estimates present performance, averaged results per quartile dropped between 2.5% (specificity and NPV) and 4% (sensitivity and PPV). Similarly, the decrease in averaged performance per quartile for the model that estimates future performance ranged from 3.3% (specificity and PPV), over 4% (negative predictive value), to 4.6% (sensitivity).

ANALYSIS OF THE BORDERLINE CASES

As estimation of future performance, i.e., performance on Q4 of the 3CVT (based on the EEG data from Q1 of the same task), is the most beneficial from the application point of view, and, as the obtained results were very promising, a deeper analysis of this prediction model is presented.

Table 5 displays a summary of the number of misclassified cases in which the observed and estimated *F*-Measure disagreed based on the *F*-Measure cut-off of 0.80. The second and the third rows contain the number of misclassifications that

Table 3 | Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) with 95% confidence intervals of the EEG-based present performance model after grouping the observed and estimated *F*-Measure into classes of unacceptable (*F*-Measure < 0.80) and acceptable (*F*-Measure ≥ 0.80) performance.

Quartile	Sensitivity	Specificity	PPV	NPV
MODEL DEVELOPMENT				
1	0.85 (0.76–0.95)	0.92 (0.89–0.95)	0.64 (0.53–0.74)	0.98 (0.96–0.99)
2	0.82 (0.75–0.89)	0.87 (0.83–0.91)	0.70 (0.62–0.78)	0.93 (0.90–0.96)
3	0.83 (0.77–0.88)	0.74 (0.69–0.80)	0.73 (0.67–0.79)	0.83 (0.78–0.89)
4	0.83 (0.78–0.88)	0.64 (0.57–0.71)	0.70 (0.64–0.76)	0.78 (0.72–0.85)
CROSS-VALIDATION				
1	0.78 (0.67–0.89)	0.90 (0.87–0.94)	0.57 (0.45–0.68)	0.96 (0.94–0.98)
2	0.80 (0.73–0.88)	0.86 (0.82–0.90)	0.68 (0.60–0.76)	0.92 (0.89–0.95)
3	0.78 (0.72–0.84)	0.70 (0.63–0.76)	0.68 (0.62–0.75)	0.79 (0.73–0.85)
4	0.81 (0.75–0.86)	0.61 (0.54–0.68)	0.68 (0.62–0.74)	0.75 (0.69–0.82)

Table 4 | Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) with 95% confidence intervals of the EEG-based future performance models after grouping the observed and estimated *F*-Measure into classes of unacceptable (*F*-Measure < 0.80) and acceptable (*F*-Measure ≥ 0.80) performance.

Quartile	Sensitivity	Specificity	PPV	NPV
MODEL DEVELOPMENT				
2	0.76 (0.68–0.84)	0.82 (0.77–0.86)	0.60 (0.52–0.68)	0.90 (0.87–0.94)
3	0.83 (0.78–0.88)	0.62 (0.56–0.69)	0.65 (0.59–0.71)	0.81 (0.75–0.87)
4	0.85 (0.80–0.90)	0.63 (0.56–0.70)	0.70 (0.65–0.76)	0.81 (0.74–0.87)
CROSS-VALIDATION				
2	0.69 (0.60–0.78)	0.81 (0.77–0.86)	0.57 (0.49–0.66)	0.88 (0.84–0.92)
3	0.81 (0.75–0.87)	0.58 (0.51–0.65)	0.62 (0.56–0.68)	0.78 (0.72–0.85)
4	0.80 (0.75–0.86)	0.58 (0.51–0.65)	0.66 (0.60–0.72)	0.74 (0.67–0.81)

Table 5 | Number of misclassifications before and after edge effect removal.

Session	Model development								Overall percentage of misclassified
	Fully rested				Sleep deprived				
	1	2	3	4	1	2	3	4	
With edge effect	15	18	21	14	9	9	8	9	25.75
Edge effect 0.05 removed	6	6	9	6	1	1	6	5	10.00
Edge effect 0.075 removed	3	4	5	1	0	0	3	1	4.25
Session	Cross-validation								Overall percentage of misclassified
	Fully rested				Sleep deprived				
	1	2	3	4	1	2	3	4	
With edge effect	17	22	23	16	12	10	9	14	30.75
Edge effect 0.05 removed	8	9	10	8	4	2	6	7	13.50
Edge effect 0.075 removed	4	5	6	3	1	1	3	3	6.50

remained after the observed *F*-Measure that fell within 0.05 or 0.075 from the *F*-Measure cut-off were discarded, respectively. The last column shows the overall percentage of the misclassified cases. Results show that a greater number of misclas-

sifications occurred during the FR sessions than during the SD sessions, with approximately 80% more quartiles misclassified in the former case, both for the model development and cross-validation models. Conversely, there was approximately

the same number of misclassified quartiles in the morning (first two sessions) and afternoon sessions (last two sessions) on both days of the study.

During the model development and cross-validation, 25.75 and 30.75% of the quartiles were originally misclassified, respectively. Still, the majority of the misclassified quartiles were within 0.05–0.075 of the *F*-Measure cut-off between acceptable and unacceptable performance. After discarding borderline cases, the number of misclassifications was decreased by 15.75 and 21.5% in the case of auto-validation and $\pm 0.05/\pm 0.075$ edge effect removal, respectively, and by 17.25 and 24.25% in the case of leave-one-subject-out cross-validation and $\pm 0.05/\pm 0.075$ edge effect removal, respectively.

As there were many misclassified borderline cases in which the difference between the observed *F*-Measure and a cut-off of 0.80 was within a $\pm 0.05/\pm 0.075$ range, we assessed the necessary changes in RT and accuracy for these borderline cases. As *F*-Measure combines two performance measures (i.e., RT and accuracy), one of the measures was kept fixed in order to determine the necessary changes in another measure. Thus, all $\pm 0.05/\pm 0.075$ borderline cases in the dataset were analyzed in this manner and the average necessary changes in accuracy and RT were small. When keeping RT fixed, accuracy had to be changed on average by 7% for ± 0.05 edge effect, and by 10% for ± 0.075 edge effect, so that *F*-Measure became 0.80. Likewise, when keeping accuracy fixed, in order to make the *F*-Measure equal 0.80, RT had to be changed on average by 0.04 s for ± 0.05 edge effect, and by 0.06 s for ± 0.075 edge effect.

Thus, tuning of the *F*-Measure threshold might bring further improvement. The *F*-Measure threshold should respect two criteria. On the one hand, it should allow for good recognition results after grouping the observed and estimated *F*-Measure values into the classes of unacceptable and acceptable performance. At the same time, it should provide a useful detection of fatigue, i.e., it should not allow too slow RT or too low accuracy to be included in the acceptable performance. Decreasing the threshold would generally improve the recognition results, but at some point definition of unacceptable performance would not be strict enough. The lower boundary we considered for *F*-Measure threshold is 0.78 that is based on the average values of RT and accuracy in the dataset (FR and SD sessions) increased, i.e., decreased by 2 STD, respectively. In that case, the range of the unacceptable performance recognition results was slightly higher than when the threshold of 0.8 was used: sensitivity 71–83%, specificity 65–88%, PPV 58–69%, and NPV 80–93%.

DISCUSSION

The current study sought to develop a method for predicting cognitive performance decrements. In order to achieve this goal, a partial sleep deprivation was performed to deteriorate the subjects' performance and the EEG data were utilized to provide insight into brain activity. The objective was to meet the following criteria: (1) The algorithm had to predict risk/errors within a time window that ensures accuracy (i.e., reduces false negatives and accommodates changing risk over time), while allowing time for intervention or mitigation; (2) The approach had to address the variability in coping strategies that occur across individuals as they become fatigued; (3) The algorithm had to accommodate individual variability in the EEG data; (4) The model had to be trained on a large dataset to avoid overfitting; and (5) The resulting algorithm had

to be computationally simple enough to allow it to be applied in driving and industrial applications without excessive processing power. In the following, we summarize how each of these criteria were met, contrast the proposed algorithm to the related work, and discuss the limitations of the proposed algorithm and future work directions.

Predicting future performance. Previous studies have demonstrated a causal relationship between the concurrent level of alertness and performance on tasks ranging from simple RT to complex decision-making (Doran et al., 2001; Kamdar et al., 2004). The purpose of this study was to extend this one step further by examining whether the EEG data can be used to estimate future cognitive performance. An emphasis was placed on predicting significant performance decrements, in terms of a large number of errors and/or slower RT, due to drowsiness caused by partial sleep deprivation. Overall, strong association between observed and estimated performance was seen. For the present performance model, after grouping the observed and estimated *F*-Measure into classes of acceptable and unacceptable performance, the achieved results were rather high (sensitivity 78–85%, PPV 57–73%, specificity 61–92%, and NPV 75–98%). Thus, the developed algorithm has proven to be robust against the false positives and false negatives that typically limit the adoption of the drowsiness prediction algorithms into real-world applications. The model that estimates future performance was able to group performance into unacceptable and acceptable performance almost as accurately as the present performance model. The obtained results were in a similar range (sensitivity 69–85%, PPV 57–70%, specificity 58–82%, and NPV 74–90%). These results support the hypothesis that the EEG data can be successfully used to predict future performance. The achieved prediction period of up to 10–15 min should provide sufficient time to intervene, and thereby prevent potential accidents in a number of real-world scenarios.

Combined performance metric. To the best of our knowledge, RT and accuracy are treated separately in all previous studies. We developed an appropriate performance metric, *F*-Measure, to ensure that all coping strategies were accommodated and equalized in the algorithm. It takes into account performance decrements due to both slower responses to stimuli and less accurate responses and it is, therefore, sensitive to a typical speed–accuracy tradeoff that has been reported and studied in related literature (Forstmann et al., 2008; Van Veen et al., 2008; Bogacz et al., 2009; Carp et al., 2010; Eichele et al., 2010). Thus, the combined performance metric utilized in our work accounted for individual differences in dealing with drowsiness.

Algorithm individualization. Most of the previous work proposed individualized models by using person specific EEG data for training a model (i.e., person-dependent training). The models developed in our study utilized person-independent training (i.e., data from all subjects were used for training), but they were still individualized through normalization of the EEG data with respect to the subjects' BL session. This way, individual EEG variability inherent in performance decrements associated with sleep deprivation was addressed.

Size of the training dataset. The proposed model was developed on a large sample size (50 subjects) and cross-validated across all subjects to assess the generalization ability of the model. Cross-validation results did not show any signs of overfitting.

Computational costs. Many researchers have attempted to develop EEG-based drowsiness detection algorithms ranging from EEG PSD bandwidth comparisons (Liang et al., 2005; Sing et al., 2005), over multivariate linear regression (Chiou et al., 2006), artificial neural networks (Vuckovic et al., 2002; Wilson and Russell, 2003; Subasi and Ercelesi, 2005), and unsupervised approaches (Pal et al., 2008) to principle component analysis (Fu et al., 2008). Most of these algorithms, however, are computationally expensive due to their complexity and/or number of EEG channels required. The algorithm proposed in this study, based on five EEG channels and a linear regression approach using PSDs, will need minimal processing power to allow for different applications in real-world settings.

A body of fatigue-related research has been conducted using a 10-min psychomotor vigilance task in which performance is based on reaction to a single type of stimuli, e.g., PVT-192 (Dinges et al., 1997). This approach requires that workers regularly step out of their regular role to complete the test, or complete it prior to an 8- to 12-h shift, leaving risk changes over time within the shift unaddressed. The current study proposed an alternative approach. Using the 3CVT, we demonstrated that EEG alone (once individualized to a baseline, a 20-min procedure) can be used to predict risk/errors. Future studies will be required in order to validate the models developed on the 3CVT in other real-world tasks, such as driving.

Furthermore, the 20-min 3CVT improved the capability to detect subtle compromises in sustained attention which only become apparent as time on task is extended. Performance on the 20-min 3CVT substantially decreased within the 20-min of the task, with Q4 having greater performance decrements than Q1, due to the lack of motivation or inability to concentrate for a relatively long period of time. Our results support this, as the developed algorithm can better estimate future performance decrements, i.e., the class of unacceptable performance, during the third and fourth quartile of the 3CVT than during the second quartile of the 3CVT, most likely because of the larger performance decrements between the FR and SD sessions on later quartiles of the task. Interestingly, the future performance model is trained on the first quartile of the 3CVT with relatively short inter-stimulus interval, yet it can still accurately estimate performance on the other quartiles of the 3CVT with larger inter-stimulus intervals.

Previous models often take into account circadian rhythm effect by, for example, incorporating a two-process model (sleep homeostasis and circadian process) of sleep regulation (Van Dongen et al., 2007; Rajaraman et al., 2008). In our dataset, however, there is no clear circadian rhythm effect on the performance measures. The third and the fourth session on both the FR and the SD day were typically recorded during the mid-afternoon circadian rhythm dip. Even so, previous analysis (Johnson et al., 2011) has shown there was no significant overall difference in performance measures (i.e., RT and percent of correct responses) between morning and afternoon sessions in the dataset. The statistical analysis in Section

“Cognitive Performance on the 3CVT” has also shown there was no significant time of day effect on *F*-Measure, as the lack of a main effect for the cycles indicated that there is no significant drop in performance across the 3CVT cycles throughout the day of testing. These findings are consistent with others that have found that, despite a clear mid-afternoon dip evident in sleep propensity (Dijk et al., 1992; Van Dongen et al., 2007), there is no clear evidence of a consistent mid-afternoon dip in performance measures due to numerous masking factors. Thus, we do not incorporate circadian rhythm process into the model.

The current algorithm is less effective in recognizing performance under the FR than the SD conditions. One potential explanation for this is the individual variability inherent in performance, with some persons learning a task more quickly and accurately than others. Furthermore, there was also potential lack of motivation on the FR day, as they already performed the same task during the BL session. Thus, poor performance in such cases was not affected by drowsiness, and the proposed EEG-based algorithm could not recognize such changes. Additionally, recent studies indicate that individual vulnerability to sleep deprivation with respect to performance varies greatly, and a few different phenotypes can be identified (Doran et al., 2001; Van Dongen and Dinges, 2001; Rajaraman et al., 2008; King et al., 2009). While some individuals are impervious to sleep deprivation and perform within normal performance measures even after 40 h without sleep, some are highly vulnerable and cannot retain high performance. In order to ensure broad adoption of the developed algorithm in vehicular and industrial environments, future development will aim to reduce the number of misclassifications associated with rested conditions by taking into account personal differences in objective performance measures. This issue could potentially be overcome by normalization of an individual's performance (i.e., the corresponding *F*-Measure) with respect to the BL session.

The ability to predict significant cognitive performance decrements well in advance could have immense value in the military, transportation industry (e.g., airline pilots, railroad operators, and surface or underwater ship operators), and any other shift-work jobs. The EEG-derived estimators of cognitive performance presented in this paper were sensitive to performance decrements due to either slower RT or less accurate responses. In the future, we plan to validate the models on a new independent dataset and to further improve the models by taking into account different levels of vulnerability to sleep deprivation, and individual speed–accuracy tradeoff strategies.

ACKNOWLEDGMENTS

This work was supported by NIH contracts N44-NS92367, N43-NS62344, N43-NS72367, and grant R43-NS35387. The authors would also like to thank Ms. Stephanie Korszen for her excellent editing advice.

REFERENCES

- Baldwin, D. C. Jr., and Daugherty, S. R. (2004). Sleep deprivation and fatigue in residency training: results of a national survey of first- and second-year residents. *Sleep* 27, 217–222.
- Balkin, T. J., Belenky, G. L., Hall, S. W., Kamimori, G. H., Redmond, D. P., Sing, H. C., Thomas, M. L., Thorne, D. R., and Wesensten, N. J. (2002). System and method for predicting human cognitive performance using data from an actigraph. 340/573.1 USPO. US 2002/0005784 A1: 21.
- Balkin, T. J., Rupp, T., Picchioni, D., and Wesensten, N. J. (2008). Sleep loss and sleepiness: current issues. *Chest* 134, 653–660.
- Baulk, S. D., Biggs, S. N., Reid, K. J., van den Heuvel, C. J., and Dawson, D. (2008). Chasing the silver bullet: measuring driver fatigue using simple and complex tasks. *Accid. Anal. Prev.* 40, 396–402.

- Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., and Olmstead, R. (2004). Real-time analysis of EEG indices of alertness, cognition, and memory with a wireless EEG headset. *Int. J. Hum. Comput. Interact.* 17, 151–170.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. T., and Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* 78, B231–B244.
- Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., and Nieuwenhuis, S. (2009). The neural basis of the speed-accuracy tradeoff. *Trends Neurosci.* 33, 10–16.
- Brookhuis, K. A., and de Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics* 36, 1099–1110.
- Brookings, J. B., Wilson, G. F., and Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biol. Psychol.* 42, 361–377.
- Carp, J., Kim, K., Taylor, S. F., Fitzgerald, K. D., and Weissman, D. H. (2010). Conditional differences in mean reaction time explain effects of response congruency, but not accuracy, on posterior medial frontal cortex activity. *Front. Hum. Neurosci.* 4:231. doi: 10.3389/fnhum.2010.00231.
- Chiou, J. C., Ko, L. W., Lin, C. T., Hong, C. T., Jung, T. P., Liang, S. F., and Jeng, J. L. (2006). “Using novel MEMS EEG sensors in detecting drowsiness application,” in *IEEE Biomedical Circuits and Systems Conference*, London, 33–36.
- Corburn, E. (1997). Shiftworker fatigue: the \$77 billion problem. *Cost Eng.* 39, 26–28.
- Dijk, D. J., Duffy, J. F., and Czeisler, C. A. (1992). Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *J. Sleep Res.* 1, 112–117.
- Dinges, D. F. (2004). Sleep debt and scientific evidence. *Sleep* 27, 1050–1052.
- Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., and Pack, A. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep* 20, 267–277.
- Dinges, D. F., and Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav. Res. Methods Instrum. Comput.* 17, 652–655.
- Dinges, D. F., and Weaver, T. E. (2003). Effects of modafinil on sustained attention performance and quality of life in OSA patients with residual sleepiness while being treated with nCPAP. *Sleep Med.* 4, 393–402.
- Doran, S. M., Van Dongen, H. P., and Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: evidence of state instability. *Arch. Ital. Biol.* 139, 253–267.
- Drowsy Driving Fact Sheet. (2009). Available at: <http://www.aasmnet.org/Resources/FactSheets/DrowsyDriving.pdf>
- Eichele, H., Juvodden, H. T., Ullsperger, M., and Eichele, T. (2010). Maladaptation of event-related EEG responses preceding performance errors. *Front. Neurosci.* 4:65. doi: 10.3389/fnhum.2010.00065
- Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Yves von Cramon, D., Ridderinkhof, K. R., and Wagenmakers, E. J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17538–17542.
- Fu, J., Li, M., and Lu, B. L. (2008). “Detecting drowsiness in driving simulation based on EEG,” in *Proceedings of the 8th International Workshop on Autonomous Systems – Self-Organization, Management, and Control*, Shanghai, 21–28. doi: 10.1007/978-1-4020-8889-6_3
- Gevens, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., and Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Hum. Factors* 40, 79–91.
- Gevens, A., Smith, M. E., McEvoy, L., and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb. Cortex* 7, 374–385.
- Gray, E. K., and Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *J. Pers.* 70, 177–206.
- Hsieh, S., Li, T. H., and Tsai, L. L. (2010). Impact of monetary incentives on cognitive performance and error monitoring following sleep deprivation. *Sleep* 33, 499–507.
- Johnson, R. J., Popovic, D. P., Olmstead, R. E., Stikic, M., Levendowski, D. J., and Berka, C. (2011). Drowsiness determination through EEG: development and validation. *Biol. Psychol.* 87, 241–250.
- Kamdar, B. B., Kaplan, K. A., Kezirian, E. J., and Dement, W. (2004). The impact of extended sleep on daytime alertness, vigilance, and mood. *Sleep Med.* 5, 441–448.
- King, A. C., Belenky, G., and Van Dongen, H. P. (2009). Performance impairment consequent to sleep loss: determinants of resistance and susceptibility. *Curr. Opin. Pulm. Med.* 15, 559–564.
- Kribbs, N. B., and Dinges, D. F. (1994). *Vigilance Decrement and Sleepiness. Sleep Onset Mechanisms*. Washington, DC: American Psychological Association, 113–125.
- Liang, S. F., Lin, C. T., Wu, R. C., Chen, Y. C., Huang, T. Y., and Jung, T. P. (2005). “Monitoring driver’s alertness based on the driving performance estimation and the EEG power spectrum analysis,” in *Proceedings of the IEEE Engineering in Medicine and Biology Society Conference*, Shanghai, 5738–5741.
- Lin, C. T., Ko, L. W., Chung, I. F., Huang, T. Y., Chen, Y. C., Jung, T. P., and Liang, S. F. (2006). Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks. *IEEE Trans. Circuits Syst.* 53, 2469–2476.
- Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* 86, 283–293.
- Makeig, S., and Jung, T. P. (1995). Changes in alertness are a principal component of variance in the EEG spectrum. *Neuroreport* 7, 213–216.
- Makeig, S., and Jung, T. P. (1996). Tonic, phasic, and transient EEG correlates of auditory awareness in drowsiness. *Brain Res. Cogn. Brain Res.* 4, 15–25.
- Olofsen, E., Dinges, D. F., and Van Dongen, H. P. (2004). Nonlinear mixed-effects modeling: individualization and prediction. *Aviat. Space Environ. Med.* 75(Suppl. 3), A134–A140.
- Pal, N. R., Chuang, C. Y., Ko, L. W., Chao, C. F., Jung, T. P., Liang, S. F., and Lin, C. T. (2008). EEG-based subject- and session- independent drowsiness detection: an unsupervised approach. *EURASIP J. Adv. Signal Process.* 2008, doi: 10.1155/2008/519480
- Pleydell-Pearce, C. W., Whitecross, S. E., and Dickson, B. T. (2003). “Multivariate analysis of EEG: predicting cognition on the basis of frequency decomposition, inter-electrode correlation, coherence, cross phase and cross power,” in *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, Waikoloa, Hawaii. doi: 10.1109/HICSS.2003.1174299
- Rajaraman, S., Gribok, A. V., Wesensten, N. J., Balkin, T. J., and Reifman, J. (2008). Individualized performance prediction of sleep-deprived individuals with the two-process model. *J. Appl. Physiol.* 104, 459–468.
- Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D. Jr., and Beck, L. H. (1956). A continuous performance test of brain damage. *J. Consult. Psychol.* 20, 343–350.
- Sing, H. C., Kautz, M. A., Thorne, D. R., Hall, S. W., Redmond, D. P., Johnson, D. E., Warren, K., Bailey, J., and Russo, M. B. (2005). High-frequency EEG as measure of cognitive function capacity: a preliminary report. *Aviat. Space Environ. Med.* 76(Suppl. 7), C114–C135.
- Smith, M. E., McEvoy, L. K., and Gevens, A. (2002). The impact of moderate sleep loss on neurophysiologic signals during working-memory task performance. *Sleep* 25, 784–794.
- Sterman, M. B., and Mann, C. A. (1995). Concepts and applications of EEG analysis in aviation performance evaluation. *Biol. Psychol.* 40, 115–130.
- Sterman, M. B., Mann, C. A., and Kaiser, D. A. (1992). Quantitative EEG patterns of differential in-flight workload. *Space Operat. Appl. Res.* 4:66–473.
- Subasi, A., and Ercelebi, E. (2005). Classification of EEG signals using neural network and logistic regression. *Comput. Methods Programs Biomed.* 78, 87–99.
- Trujillo, L. T., Kornguth, S., and Schnyer, D. M. (2009). An ERP examination of the different effects of sleep deprivation on exogenously cued and endogenously cued attention. *Sleep* 32, 1285–1297.
- Van Dongen, H. P., and Dinges, D. F. (2001). Modeling the effects of sleep debt: on the relevance of inter-individual differences. *Sleep Res. Soc. Bull.* 7, 69–72.
- Van Dongen, H. P., Maislin, G., Mullington, J. M., and Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 26, 117–126.
- Van Dongen, H. P., Mott, C. G., Huang, J. K., Mollicone, D. J., McKenzie, F. D., and Dinges, D. F. (2007). Optimization of biomathematical model predictions for cognitive performance impairment in individuals: accounting for unknown traits and uncertain states in homeostatic and circadian processes. *Sleep* 30, 1129–1143.
- Van Veen, V., Krug, M. K., and Carter, C. S. (2008). The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *J. Cogn. Neurosci.* 20, 1952–1965.
- Vuckovic, A., Radivojevic, V., Chen, A. C., and Popovic, D. (2002). Automatic recognition of alertness and drowsi-

- ness from EEG by an artificial neural network. *Med. Eng. Phys.* 24, 349–360.
- Weaver, T. E., Laizner, A. M., Evans, L. K., Maislin, G., Chugh, D. K., Lyon, K., Smith, P. L., Schwartz, A. R., Redline, S., Pack, A. I., and Dinges, D. F. (1997). An instrument to measure functional status outcomes for disorders of excessive sleepiness. *Sleep* 20, 835–843.
- Weinstein, M., Silverstein, M. L., Nader, T., and Finelli, L. (1999). Sustained attention on a continuous performance test. *Percept. Mot. Skills* 89(Pt 1), 1025–1027.
- Wilkinson, R. T. (1990). Response-stimulus interval in choice serial reaction time: interaction with sleep deprivation, choice, and practice. *Q. J. Exp. Psychol. A* 42, 401–423.
- Wilson, G. F. (2004). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *Int. J. Aviat. Psychol.* 12, 3–18.
- Wilson, G. F., and Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Hum. Factors* 45, 381–389.
- Conflict of Interest Statement:** Authors Stikic, Johnson, Levendowski, Popovic, Olmstead, and Berka are paid salaries or consulting fees by Advanced Brain Monitoring, Inc., and Johnson, Levendowski, Popovic, Olmstead, and Berka are shareholders of Advanced Brain Monitoring, Inc.
- Received: 25 January 2011; accepted: 15 July 2011; published online: 05 August 2011.
- Citation: Stikic M, Johnson RR, Levendowski DJ, Popovic DP, Olmstead RE and Berka C (2011) EEG-derived estimators of present and future cognitive performance. *Front. Hum. Neurosci.* 5:70. doi: 10.3389/fnhum.2011.00070
- Copyright © 2011 Stikic, Johnson, Levendowski, Popovic, Olmstead and Berka. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.