Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model

Robin R. Johnson*, Djordje P. Popovic, Richard E. Olmstead, Maja Stikic, Daniel J. Levendowski, Chris Berka

Advanced Brain Monitoring, Inc., University of California, Los Angeles, United States

ARTICLE INFO

ABSTRACT

A great deal of research over the last century has focused on drowsiness/alertness detection, as fatigue-related physical and cognitive impairments pose a serious risk to public health and safety. Available drowsiness/alertness detection solutions are unsatisfactory for a number of reasons: (1) lack of generalizability, (2) failure to address individual variability in generalized models, and/or (3) lack of a portable, un-tethered application. The current study aimed to address these issues, and determine if an individualized electroencephalography (EEG) based algorithm could be defined to track performance decrements associated with sleep loss, as this is the first step in developing a field deployable drowsiness/alertness detection system. The results indicated that an EEG-based algorithm, individualized using a series of brief "identification" tasks, was able to effectively track performance decrements associated with sleep deprivation. Future development will address the need for the algorithm to predict performance decrements due to sleep loss, and provide field applicability.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Researchers have focused on drowsiness detection primarily due to the substantial human and economic costs in public health, public safety, and productivity associated with drowsiness-related functional impairment. Early studies estimated that up to 30% of fatal vehicular accidents are caused by fatigued (i.e., drowsy) driving, while more recent data suggests that 1.6% of all crashes and 3.6% of all fatal crashes can be attributed to fatigue (NTSB, 1990, 1999). At this time, fatigued driving is a well recognized public safety concern both for commercial (Fournier et al., 2007) and private drivers (Fletcher et al., 2005). In addition to vehicular safety concerns, sleepiness is also an issue for industrial workers, pilots, air traffic controllers and medical care givers (Sigurdson and Ayas, 2007). These safety concerns are on par with those associated with alcohol intake, with errors due to sleep loss resulting in similar consequences such as accidents with multiple fatalities (Dinges and Kribbs, 1990; Powell et al., 2001). While public safety concerns are perhaps the most urgent, public health is also significantly impacted by drowsiness (Sigurdson and Ayas, 2007). Excessive fatigue (due to poor sleep hygiene or sleep disorders) reduces productivity and neurocognitive function (Dinges et al., 1997), increases the risk of developing obesity (Levy et al., 2009),

metabolic syndrome (Calvin et al., 2009), diabetes (Surani et al., 2009), depression (Edd and Flores, 2009), and reduces quality of life. Numerous studies have also demonstrated a causal relationship between level of alertness and performance on tasks ranging from simple reaction time, to complex decision-making (Doran et al., 2001; Kamdar et al., 2004), potentially leading to deficits in quality of life and productivity. Unfortunately, the measurement of alertness/drowsiness has proven elusive, as it represents a complex interaction of both physiological (e.g., level of fatigue, overall health) and psychophysiological variables (e.g., motivation, task demands and time of day).

Neurophysiological and/or behavioral measurements such as actigraphy, eye movement/blink tracking, performance tests, and electroencephalography (EEG) have all been shown to provide objective and relatively accurate quantification of drowsiness/alertness (Dinges and Powell, 1985; Makeig et al., 1994; Blood et al., 1997; Mallis, 1999). Many, however, believe that EEG-based technologies will provide the most broadly applicable, accurate and efficient drowsiness detection systems. EEG offers technological advantages that may overcome the shortcomings of these other technologies, to provide a task independent, non-disruptive method for detecting drowsiness, as well as predicting proximate future errors (Lal and Craig, 2001). Electrooculographic (EOG) and EEG recordings provide insight into brain activity directly associated with various states of arousal from sleep to waking (Santamaria et al., 1987). EEG is often applied as the "gold standard" in the identification of states ranging from vigilant and alert

* Corresponding author. Tel.: +1 760 720 0099.
 E-mail address: rjohnson@b-alert.com (R.R. Johnson).

to drowsy or asleep. Changes in the frequency and amplitude of the EEG have also been shown to correlate directly with behavioral performance measures, particularly on tasks which require sustained attention over long periods of time (Jung and Makeig, 1994; Makeig and Jung, 1995). Previous studies (Makeig and Inlow, 1993) have demonstrated specific EEG correlates with changes in alertness that result in alterations in performance. In addition, measurements of event-related EEG signals, such as event related potentials (i.e., ERPS), have proven sensitive to changes in perception, attention, and cognition (Gevins et al., 1990; Makeig, 1993; Coenen, 1995; Lim et al., 1999; Sambeth et al., 2004).

Many researchers have attempted to leverage these characteristics to develop EEG based drowsiness algorithms. The statistical and/or mathematical modeling used to develop the algorithms include: EEG power spectral density (PSD) bandwidth comparisons (no underlying model) (Liang et al., 2005; Sing et al., 2005; Pal et al., 2008), ERP (event related potentials) latency increases (no underlying model) (Smith et al., 2002), linear regression (Chiou et al., 2006), stepwise linear discriminant functions, artificial neural networks (Vuckovic et al., 2002; Wilson and Russell, 2003; Subasi and Ercelebi, 2005), and principle component analysis (Fu et al., 2008). The empirical support for each of these algorithms is of mixed quality. Common weaknesses are: (1) small sample size, (2) lack of cross validation analysis (or other acknowledgement/accommodation of individual variance), (3) task dependence/specificity, and (4) algorithm complexity. Small sample sizes typically lead to over fitting models, and reduce the likelihood that such algorithms are generalizable across individuals in the general population. The largest sample size used for the cited algorithms appears to be $n = 30$ (Subasi, 2005), with many developed on sample sizes of less than $n = 10$. With the exception of the EEG PSD bandwidth comparison methods, all of these techniques have theoretical underpinnings that require much larger sample sizes in order to produce stable models (ideally $n = 30$ for each variable used in the final model) (Tabachnick and Fidell, 1983, 2007). Some algorithms reported are based solely on theory (with no actual data used to develop or evaluate the actual algorithm). In addition to small sample sizes, most algorithms noted do not accommodate individual variability, either in the algorithm methodology or through cross validation analysis (nor do they have sample sizes that would support them), limiting application across individuals. As individual differences are a major confound in all EEG based algorithm development (Karis et al., 1984; Makeig and Jung, 1995, 1996; Van Dongen et al., 2004; Wong et al., 2008), failure to accommodate this issue (either through cross-validation, or individualization as part of the modeling development) reduces the potential adoption of any of the algorithms developed thus far. Moreover, generalizability of the drowsiness algorithms across tasks is rarely, if ever, addressed, limiting interpretation and application outside of the laboratory, or beyond the specific task upon which the algorithm is developed. Finally, many of the previously proposed algorithms are computationally expensive due to their complexity and large number of channels required (up to 66) (Smith et al., 2002; Vuckovic et al., 2002; Wilson and Russell, 2003; Liang et al., 2005; Lin et al., 2006; Fu et al., 2008; Pal et al., 2008), limiting their implementation in real time settings.

In addition to the algorithm used, the equipment required may also limit adoption of any drowsiness detection system in many, if not most, personal and workplace environments. Until the past decade, the practical application of EEG measures outside the laboratory was limited by the technical difficulties of ambulatory physiological recording. Technological advances have resulted in equipment designed to record high quality EEG using lightweight, portable devices suitable for non-laboratory environments. Neurophysiologic data has been successfully collected from interstate truck drivers (Miller, 1995), train operators (Torsvall and Akerstedt, 1987), pilots (Gundel et al., 1995), and physicians (Richardson et al.,

1996) during their normal work hours. Other investigators have utilized ambulatory EEG equipment to monitor daytime drowsiness in narcoleptics (Broughton et al., 1988) and sleep disorder patients (White et al., 1995), or to record seizures in epileptic patients (Ives and Mainwaring, 1993). Although these studies clearly demonstrate the viability of recording EEG in normal workplace environments, a number of practical considerations remain unresolved. Primarily, these systems require trained technicians to apply recording electrodes secured to the scalp with collodion, or a placement cap (e.g., ElectroCap). Some of these studies also used a large number of electrode sites (as with a number of the drowsiness algorithms developed thus far), limiting portability and duration of data collection periods able to be recorded and/or monitored.

The current study describes the development of a system that sought to address each of these issues: (1) ensure maximal stability and inter-individual generalizability by using a large sample size and individualizing the model, (2) be applicable across tasks (task generalizability), (3) be computationally accurate and efficient for use in a portable hardware application, and (4) provide a hardware platform to apply the algorithm in the field. In addition, the final system will enable field studies to determine the true applicability of the algorithm in future studies. EEG collected during four neuropsychological tasks conceptually associated with four cognitive states on the sleep to alertness continuum was used to build and train the algorithm. These tasks included: the Osler modified maintenance of wakefulness task (Krieger et al., 2004), standard eyes closed and eyes open vigilance tasks, and a proprietary 3-choice vigilance task, similar to the PVT-192. The cognitive states defined by each of these tasks were, respectively, sleep onset, distraction/relaxed wakefulness, low engagement, and high engagement. Drowsiness is determined by combing sleep onset and distraction probabilities. We previously reported applications of this algorithm, as well as the artifact identification, decontamination, general signal processing, and a brief description of the algorithm, in its final format (Berka et al., 2004, 2007a,b). Two studies were used to develop and validate the algorithm and system. Both studies included a fully rested "baseline" assessment, as well as a sleep deprived session used to build and train the algorithm. The subsequent protocols of each study allowed for across task validation evaluation by comparing the drowsiness classification to performance decrements over time of sleep deprivation. The first study examined daytime rested and sleep-deprived performance (and EEG) during two consecutive 8 h daytime sessions. The second study followed a smaller set of subjects for 48 consecutive hours of sleep deprivation.

2. Method

2.1. Participants

2.1.1. Study 1

A sample of $n = 200$ participants were enrolled after screening for the following exclusion criteria: self-report of excessive daytime sleepiness (Epworth > 6); excessive smoking (more than 10 cigarettes/day) or caffeine intake (more than 5 cups/day); history of sleep, neurological or psychiatric disorder; head trauma; symptoms of a sleep disorder; and inconsistent sleep patterns (<7.25 h/night on average). A total of $n = 135$ participants were selected for the model development data set, with $n = 65$ eliminated due to: (a) insufficient or poor sleep the night before data collection, (b) signs of sleepiness during rested session tasks, or (c) excessively poor performance on tasks. These participants had a mean age of 26.8 yr (range: 18–71 yr), and were ethnically diverse and gender balanced (30.3% non-white, 48.1% female). Of the $n = 135$ subjects, a subset of $n = 65$ underwent sleep deprivation and provided data with transitions between awake and sleep onset (mean age 28.0 yr, range 19–63; 31.4% non-white; 49.2% female).

2.1.2. Study 2

Participants were recruited for the second study, using the same exclusion criteria as study 1 ($n = 25$). The algorithm includes data from all $n = 25$ participants from this study. These participants had a mean age of 24.8 (range: 18–44 yr, 48% non-white; 24% female).

All participants in both studies provided signed, informed consent prior to participation in these studies.

2.2. Equipment/Materials

2.2.1. EEG/EOG acquisition

In study 1, a prototype version of the B-Alert System Sensor headset was used to locate disposable Ag/AgCl electrodes (Physiometrix Corp., New Benita, MA) at Fz, Cz, Pz, POz and Oz (based on the international 10–20 system), with left and right earlobes as the reference and ground. Referential recordings were acquired from each of the electrode sites and differential recordings for Fz-POz, Cz-Pz, and Cz-POz. Differential recordings from CzOz were calculated offline by the B-Alert software. Vertical and horizontal EOG were recorded referentially (for use in confirming the eye blink algorithm based on Fz-POz signal). The prototype AC coupled data acquisition system included Teledyne (Marina Del Rey, CA) amplifiers, 12 bit ADC, a low pass filter at 75 Hz with a dynamic range of $\pm 125\,\mu V$ and fixed gain of 10,000, and a high pass filter at 0.5 Hz with dynamic range of $\pm 625\,\mu V$ and fixed gain of 2000. The sampling rate was 256 samples for all channels.

For study 2, a wireless, portable sensor headset was utilized. The sensor headset acquired six channels of EEG and EOG, using a mixed referential/differential montage, with 3 sensors down the midline at Fz, Cz, and POz (based on initial algorithm development indicating that Fz-POz and Cz-POz were the only required channels). Data were sampled at 256 Hz with a band pass from 0.5 Hz to 65 Hz (at 3 dB attenuation) obtained digitally with Sigma-Delta A/D converters. The RF link was frequency-modulated to transmit at a rate of 57 kBaud in the 915 MHz ISM band. By utilizing the bidirectional mode, the firmware allowed the host computer to initiate impedance monitoring of the electrodes, select the transmission channel (so two or more headsets can be used in the same room), and monitor battery power of the headset. Data were acquired across the RF link on a host computer via an RS232 interface. Data acquisition software then stored the EEG data on the host computer. The proprietary acquisition software used also includes artifact decontamination algorithms for eye blink, muscle movement, and environmental/electrical interference such as spikes and saturations (Berka et al., 2004).

2.2.2. Software/Neuropsychological tasks

The software incorporates performance signals from the neuropsychological assessments into a separate event channel of the EEG record for synchronized data analysis. The neuropsychological tasks used to build the algorithm, and subsequently used to individualize the algorithm's centroids (see algorithm section) were presented using proprietary acquisition software. The algorithm was trained using EEG data collected during the Osler maintenance of wakefulness task (OSLER) (Krieger et al., 2004), eyes closed passive vigilance (EC), eyes open passive vigilance (EO), and 3-choice active vigilance (3CVT) tasks to define the classes of sleep onset (SO), distraction/relaxed wakefulness (DIS), low engagement (LE), and high engagement (HE), respectively. Subsequent performance of the algorithm was evaluated in additional neuropsychological tests including image recognition and interference tasks.

In study 1, a hand-held finger tapping device (similar to a counter device) was used during the OSLER, EC and EO tasks. The device was directly connected to the data acquisition unit to provide synchronized performance signals to incorporate into the EEG record. EO presented a 10 cm circular target image for 200 ms in the center of the computer monitor, repeated every 2 s for 5 min; the subject was asked to tap the spacebar in time with the target image. For EC, an auditory tone every 2 s prompted the participant to tap in time with the noise for 5 min. The OSLER required participants to recline (45° angle) in a comfortable chair in a dimly lit room, and attempt to remain awake while pressing the hand-held event marker every 2 s (to a visual stimuli similar to the EO stimuli). OSLER sessions were terminated after: (a) 40 min, or (b) the subject was unable to remain awake without constant technician intervention. A limitation of OSLER procedure, as a measure of sleep onset, was that subjects were unable to consistently keep their eyes open prior to sleep onset. In addition, when subjects awoke from a long sleep-interval (>30 s), they would momentarily forget to finger-tap. For the purpose of identifying SO epochs, these cases were resolved by using video recordings of the subjects to distinguish sleep onset and awakening periods. In study 2, the response tracking was integrated into the software, and, rather than use the separate hand held tracker, the spacebar was used for each of these tasks. In addition, the OSLER was limited to 40 min to maintain the repeated measures schedule.

The 3CVT developed by ABM incorporates features of the most common measures of sustained attention, such as the Continuous Performance Test (Weinstein et al., 1999; Randerath et al., 2000), Wilkinson Reaction Time (Wilkinson and Houghton, 1982; Wilkinson, 1990; Valencia-Flores et al., 1996), and the PVT-192 (Dinges and Kribbs, 1991; Dinges et al., 1997), and it is designed to allow simultaneous monitoring and quantification of the EEG. These tasks are sensitive to changes in alertness as a result of acute sleep deprivation, cumulative sleep loss (Dinges et al., 1997), daytime drowsiness in untreated sleep apnea in clinical populations (Dinges and Weaver, 2003), and commercial truck drivers (Baulk et al., 2008). The 3CVT was directly compared to the PVT-192 (Ambulatory Monitoring Inc., Ardsley, NY) to determine the validity of using the 3CVT in a manner similar to the PVT-192 (Dinges and Powell, 1985; Dinges et al., 1997). The 3CVT and PVT-192 do not

result in significantly different performance metrics of reaction time and lapses, Fs $(1, 174) \leq 1.205$, ps > .62, and they are highly correlated ($r = .74$, .84 respectively).

The 3CVT requires subjects to discriminate one primary (70% occurrence) from two secondary (30% occurrence) geometric shapes with stimulus presentation intervals of either 0.2 or 1 s for two separate versions of the task, over a 20-min test period. Participants were instructed to respond as quickly as possible to each stimulus presentation. A training period was provided prior to the beginning of the task to minimize practice effects. During the first 5 min of the session, the inter-stimulus interval ranged from 1.5 to 3 s, while the middle 10-min period had an inter-stimulus interval range of 1.5 to 6 s. During the final 5-min, the inter-stimulus interval range was 1.5 to 10 s. Participants were instructed to select the left arrow to indicate target stimuli, and the right arrow to indicate non-target stimuli. The stimulus presentation rate did not significantly impact performance accuracy or reaction time ($p > .05$), and the two versions (.2 s and 1 s stimulus presentation rate) were applied to the protocol in a counterbalanced order across participants.

Image Recognition/Interference Learning (IR/IIR) tasks were used to evaluate attention and short-term memory, with each task taking 6 min. The initial IR task included both training and testing periods. During the training period, participants were asked to memorize a series of 20 target images that were presented twice per image. To ensure the participant was attending to the target images, they were required to respond to each image by pressing the space bar. In the testing period, the participants were then asked to identify the target images (select left arrow for targets, right arrow for non-target/interference as with the 3CVT) in a field of 100 total images (20 target/80 non-target). The IIR task also included both training and testing periods. In the image case, a new set of 20 target images were trained, with the testing consisting of 20 target/20 interference (the targets from the original task)/60 non-target images. In the number based IIR task, the new targets were the same images as the original image recognition tasks, with a corresponding number added to each (paired). The interference images were then the target images with incorrect numbers, in the same ratio of 20 target/20 interference/60 non-target images. The two interference tasks did not differ in performance accuracy or reaction time ($p > .05$), and were used in a counterbalanced manner across participants. In addition, there were four image categories available (food, animal, sports, and travel), and these were used in a counterbalanced order across participants as needed to ensure that there were not carryover interference effects over time for each participant. Performance accuracy and reaction time were not altered due to category ($p > .05$).

The driving simulator system (STI System Inc., Hawthorne, CA) was programmed to present a 45-min scenario that included a 24-min monotonous section and an 18-min challenging section. Two unique but equitable scenarios were developed to avoid learning effects. Each scenario was designed with a four-lane road that could be driven at a constant speed of 55 mph. For the monotonous scenarios, driving challenges (i.e., traffic lights, cross traffic at an intersection, and vehicles entering traffic) were presented periodically and randomly, interspersed with 10–45 s of straight-aways and occasional 12° curves. A small number of vehicles traveling in the same direction were presented periodically in the slow lane. Response time to a divided attention task was measured randomly approximately once per minute with an up or down arrow that was presented in the upper left or right corner of the driving simulator, and disappeared if the subject did not respond by pushing (up arrow) or pulling (down arrow) the gear shifter within 5 s. The challenging scenarios required the driver to change lanes often to avoid and/or pass slower cars, and monitor the rear view for vehicles traveling at a faster speed. Traffic lights, intersections with cross traffic and/or pedestrians, and curves in the highway were presented routinely. Divided attention tasks were presented randomly, approximately every 20 s. For both scenarios, speeding tickets were tallied when the driver exceeded 62.5 mph. The vehicle was repositioned in the center of the fast lane immediately after any accidents, and the driving scenario continued. Subjects were allowed a 15 min orientation session on the rested session to become familiar with both types of scenarios (monotonous and challenging). Variables recorded by the driving simulator system for subsequent off-line analysis included the number of: (a) accidents, (b) collisions, (c) hit pedestrians, (d) speeding tickets, (e) traffic tickets, (f) correct divided attention task (DAT) responses, (g) incorrect DAT responses, and (h) non-responses to the DATs.

Wrist actigraphs were used to track sleep/wake patterns throughout the study. Precision Control Design Inc. model # OBMA-0.2 were used in conjunction with the PCD proprietary software to ensure compliance with maintenance of regular sleep/wake cycles and to confirm compliance with sleep deprivation requirements in both studies, prior to the participants arriving at the laboratory.

3. Protocol/Procedures

3.1. Study protocol

Both studies assessed subjects under fully rested conditions and sleep deprived conditions to acquire algorithm training data. Fully rested data from EO, EC, and 3CVT were used to train DIS, LE and HE; while the OSLER data collected in the sleep deprived condition were used to train the SO class. In order to validate the algorithm, addi-

tional repeated assessments of the 3CVT and IR/IIR tasks, as well as the driving simulation (from study 1), were used. These data were examined to determine if the drowsiness metric (SO + DIS) from the algorithm tracked performance decrements. Both studies were reviewed and approved by an independent Institutional Review Board (Children's Hospital of San Diego Independent IRB and Biomedical Research Institute of America, San Diego, CA for studies 1 and 2, respectively).

In study 1, participants ($n = 135$) were asked to complete up to two separate experimental sessions over 3 days: Screening (fully rested, morning only), and a 2-day study. The 2-day study ($n = 65$) consisted of back-to-back sessions: rested (fully rested morning and afternoon) and sleep deprived (morning and afternoon). Each participant wore a wrist actigraph for 4 days prior to each session appointment and completed sleep logs to determine sleep consistency, sufficiency, and patterns. Due to changes in the EEG signal associated with diurnal variations (Higuchi et al., 2001), all subjects were scheduled to begin the screening protocol prior to 1000, and the 2-day protocol at 0900 on both days. All subjects completed identical tasks, however, the categorical order of the IR/PAL and 3CVT were varied, in 4 sets of tasks beginning at 0900, 1100, 1300, and 1500. Participants were required to wear an actigraph and call into a voicemail every 30 min the night between the rested and sleep deprived parts of the 2-day study.

In study 2, participants ($n = 25$) were asked to complete 2 sessions: 1 fully rested baseline session (data from this time point were used for the non-Sleep Onset epochs to train the algorithm) and 1 sleep deprived session. The sleep deprivation session required subjects to be awake at 0700, and arrive for the experimental session at 1900. Repeated cycles of the same set of tasks (EO, EC, OSLER, 3CVT, IR/IIR) occurred in 3 h blocks.

3.2. Signal processing

Signal processing consisted of the following three steps: filtering and digitization, artifact identification/decontamination, and feature extraction. The two EEG signals were passed through a 1st-order analog band-pass filter (0.5–65 Hz) before the analog-to-digital conversion at a rate of 256 samples per second. The signals were subsequently subject to a cascade of sharp digital notch filters with stop-bands centered at 50, 60, 100 and 120 Hz, which corresponded to the main frequency and the first harmonics of the power network in the US and Eurasia. Such a combination (analog band-pass and digital, rather than analog, notch filters) was selected because it significantly reduced the size of the analog board of the EEG Headset device and, at the same time, allowed for a straightforward addition of filters to the cascade if needed (e.g., in environments with strong higher harmonics). Identification and removal of artifacts was incorporated into the system, including: eye blinks, muscle activity contamination (EMG), spikes (due to motion or QRS complexes, for example), and saturations (Berka et al., 2004).

In the feature extraction step, features were computed from the decontaminated EEG signal on a second-by-second basis. For each 1 s epoch, three consecutive 50% overlapping windows were used to compute the absolute and relative power spectral density (PSD) values in the following manner. First, the Kaiser window ($\alpha = 6.0$) was applied to each overlay to reduce the edge effect in the next step when the 256-point FFT (fast Fourier transformation) was performed. The FFT was averaged on the three overlays to decrease epoch-by-epoch variability. Finally, the corresponding absolute and relative PSD values for each 1 Hz bin from 1 Hz to 40 Hz were derived, summing up to 80 features per channel. Data from both Fz-POz and Cz-POz were incorporated, resulting in 160 original features.

Table 1
Algorithm features.

Channel	Absolute/Relative	Hz bins
FzPOz	Absolute	14 25 27 30 36 38 40
CzPOz	Absolute	4 5 6 10 28 32 33 35
	Relative	8 9 11 12 14 16 26 35 40

3.3. Algorithm development

The algorithm was comprised of three steps: feature selection, training, and classification. Feature selection was performed by submitting the extracted EEG features to stepwise analysis in two different ways. First, stepwise regression was performed for all subjects together, and then for each subject separately (to begin to individualize the model). The final set of the most predictive features consisted of the 19 features selected through stepwise analysis of the total group of subjects, together with 5 additional features selected through individual stepwise analysis. The final algorithm includes the 24 features noted in Table 1.

Training of the classifier utilized discriminant function analysis (DFA) to calculate the centroids and covariance matrices for each of the 4 classes based on the population data. Once the underlying model was determined using the population data, individual differences in the EEG data were addressed through model individualization. Individual data from the "identification" (ID) tasks (3CVT, EO, and EC) collected at baseline were used to calculate the centroids and covariance matrices for the three classes distraction/relaxed wakefulness (DIS), low engagement (LE), and high engagement (HE), respectively, and the final discriminant function was refined using the individual's class centroids and covariance matrices (see Fig. 1 for example of how this process clarifies and delineates the classes). Sleep onset was trained in two steps, in order to develop an algorithm that does not require sleep onset data *a priori*. Step one used DFA based on OSLER data from sleep deprived sessions and the rested EO, EC, and 3CVT data. In the second step, SO was derived using a secondary stepwise linear regression (based on the DFA defined features from the first step) using data from the HE, LE and DIS, to predict the sleep onset centroid, with the individual's pooled covariance matrix supplying the covariance matrix for the sleep onset class. This method allowed for SO to be classified based solely on the 3 "ID" tasks.

The classification step was performed by calculating the Mahalanobis distances of the input feature vector from each class centroid. Then, the posterior probabilities of the input vector belonging to each of the four output classes (HE, LE, DIS, and SO) were computed by taking into account the prior probabilities of each class occurrence. The vector was classified based on the highest posterior probability. Therefore, when epoch by epoch classification is utilized, an epoch will be classified as HE if the HE probability is greater than LE, DIS and SO.

Drowsiness is defined as the combined probability of SO and DIS classes.

3.4. Statistics

Data from each study were analyzed separately, outside of the algorithm development. In order to assess the sensitivity of task performance to sleep loss performance decrements, repeated measures-ANOVA was used to compare changes over time in each task. Reaction time and inaccuracy (incorrect + missed responses) were compared for each task, with Duncan's New Multiple Range Test ($p < .05$) used to compare group means of main effects; data is presented as mean ± SEM. For the DRIVE task, inaccuracy was substituted with overall errors (summation of each error type over the entire task). In order to assess the efficacy of the algorithm in track-
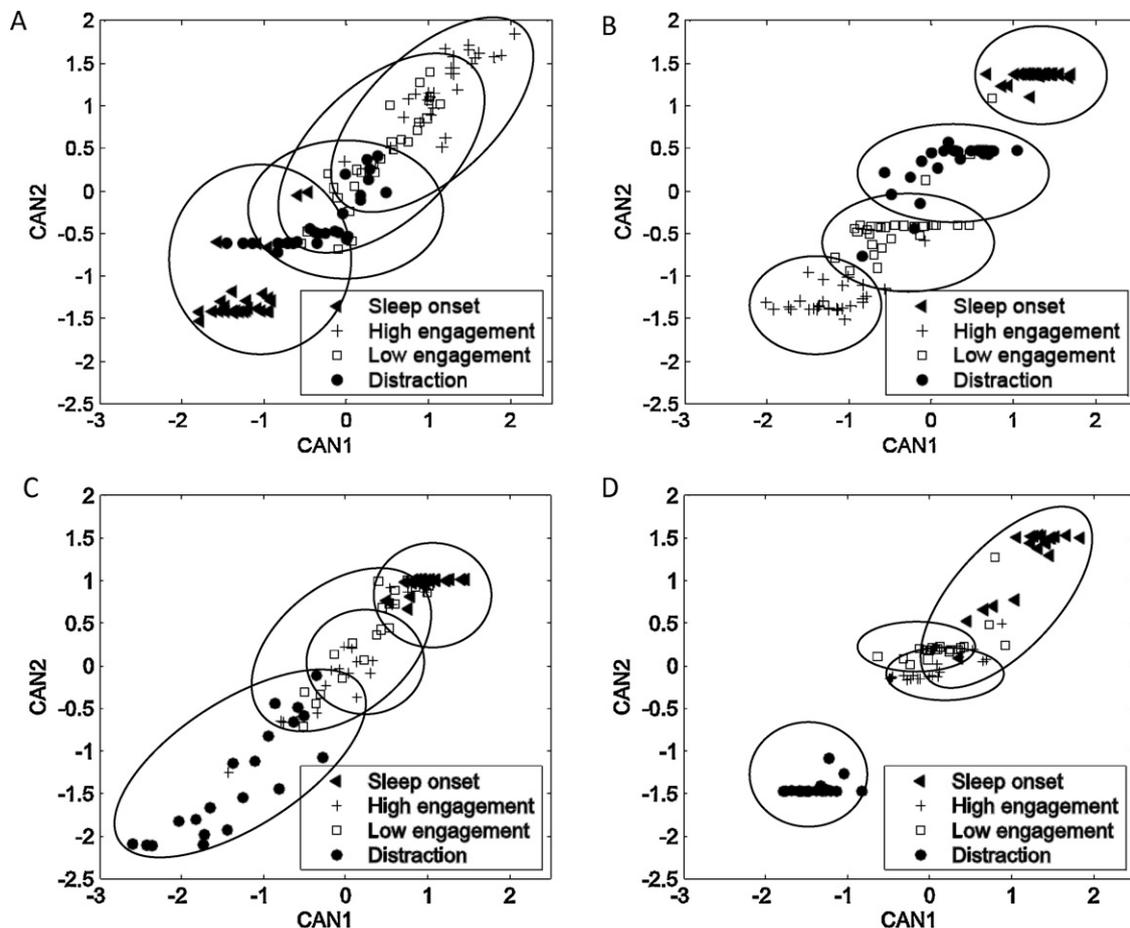
Fig. 1. Cannonical plots of population based centroids (A and C) and the individualized models (B and D) for two different subjects (subject 1: A, C; subject 2: B and D), demonstrating how the individualization shifts and delineates the four classes.

ing performance related errors, mean Pearson's correlations were conducted comparing total inaccuracy and mean task drowsiness probability at each time point available for each task.

4. Results

The results presented herein address: (a) the validity of the neuropsychological tests used, and (b) across task validation of the algorithm's drowsiness classification, using repeated 3CVTs, IR/IIRs, and simulated driving.

4.1. Performance

In study 1, identical tests were conducted during the screening and the morning of the fully rested sessions. Both a one-way repeated measures ANOVA and Pearson's correlations were used to evaluate the test–retest reliability and stability of the neurocognitive tasks. Repeated measures ANOVA found no significant differences at any time point from screening and fully rested sessions, indicating that the 3CVT, SIR, IIR performances did not significantly change over time (2–6 weeks apart), or even with additional exposure. Reaction time p-values ranged from .08 to .23, indicating a potential trend toward slowed reaction times, however, inaccuracy and missed p-values ranged from .47 to .72. Pearson's correlations were used to additionally assess reliability and stability of these tasks for both inaccuracy/misses and reaction times. These analyses resulted in $rs = .56–.83$.

Sensitivity of reaction time and inaccuracy rates were assessed in both studies for each of the three primary neuropsychological

tests (3CVT, SIR, IIR). Study 1 compares morning and afternoon performance under rested and sleep deprived conditions, while study 2 compares performance over 48 h of sleep deprivation. In study 1, a 2 (rested, sleep deprived) × 2 (time of day) repeated measures ANOVA found no significant Condition × Time of Day interaction for any of the three neuropsychological tests ($p > .05$). A main effect of condition (fully rested vs. sleep deprived), however, was shown for all three neuropsychological tests on reaction time, $Fs (1, 63) \geq 9.68$, $ps < .0001$, and inaccuracy $Fs (1, 63) \geq 7.14$, $ps < .001$ (3CVT data shown in Fig. 2A and B, respectively). Post hoc analysis indicated that sleep loss was associated with poorer performance at each time point.

Study 2 repeatedly assessed performance in all tasks every 4 h, up to 48 h of sleep deprivation. Repeated measures ANOVA revealed a significant main effect of time for both reaction time, $Fs (9, 215) \geq 13.99$, $ps < .0001$; and inaccuracy $Fs (9, 215) \geq 10.58$, $ps < .0001$ for 3CVT, IIR and SIR. Data for 3CVT are shown in Fig. 2C. Post hoc analysis indicated that, overall, as sleep loss increased, performance was negatively impacted, though not in a linear manner.

4.2. Drowsiness

Drowsiness is defined as the summated probability of the SO and DIS classes. A 2-way repeated measures ANOVA (Condition × Time of Day) revealed no interaction of condition with time of day, however a main effect of condition was revealed for all three neuropsychological tasks (3CVT, SIR, and IIR), $Fs (2, 194) \geq 12.12$, $ps < .0001$. Data for 3CVT and SIR are presented in Fig. 3A and B, respectively.
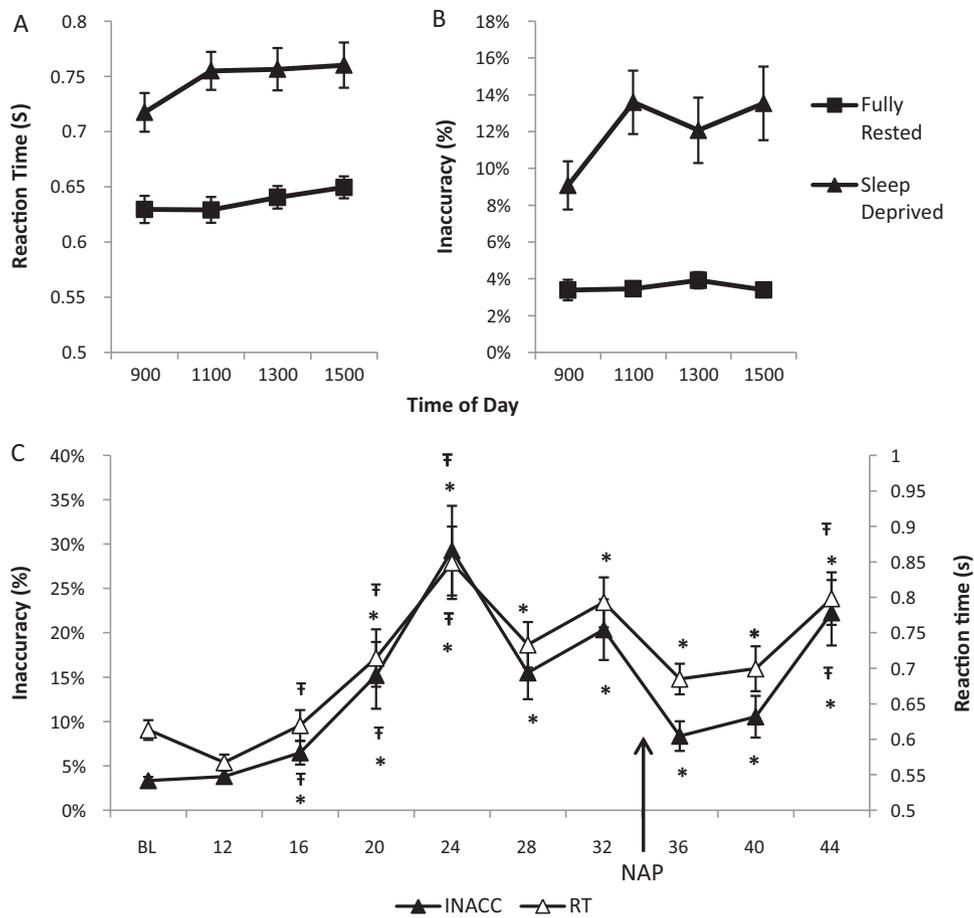
Fig. 2. Reaction time (A) and inaccuracy (B) are shown for 3CVT throughout a day when fully rested (■) and sleep deprived (24 hr, ▲), as well as across 48 h of sleep deprivation (C). The main effect of condition was found ($ps < .001$) for the 3CVT, indicating performance decrements occur with sleep loss. The main effect of time of sleep deprivation over 48 h was also found ($p > .001$, C). Significant post hoc differences are shown as asterisks (*) for change from baseline, and crosses (Ŧ) to indicate significant change from immediately prior time point. Similar effects were also found for IIR and SIR (data not shown).

4.3. Drowsiness vs. Performance

The relationship of drowsiness probability and inaccuracy were examined for both studies, during both neuropsychological testing, as well as the driving simulator, using Pearson's correlation analysis. For both studies the neurocognitive tasks were pooled for these analyses, as well as separated by task.

In study 1, drowsiness probability was significantly associated with performance, both when the tasks were pooled ($r = .58$), and by task for 3CVT, SIR and IIR ($rs = .48, .22,$ and $.24$, respectively; 3CVT data is shown in Fig. 4A, pooled data are shown in Fig. 4B and C for

rested and sleep deprived states, respectively). Conversely, drowsiness was not significantly related to performance under the rested condition ($rs = .07, −.03,$ and $−.09$, respectively, $.02$ for pooled data).

In study 2, similar phenomena developed (3CVT data shown in Fig. 4D–F). The relationship between drowsiness and inaccuracy was examined using Pearson's correlation analysis. Once again, when subjects were sleep deprived, a significant relationship occurred for both the pooled data ($r = .55$), and for each task, 3CVT, SIR, and IIR tasks ($rs = .55, .46$ and $.52$). In contrast, the correlations found under rested conditions were not significant ($rs = .12, .08$ and $.18$, respectively; pooled $r = .15$).
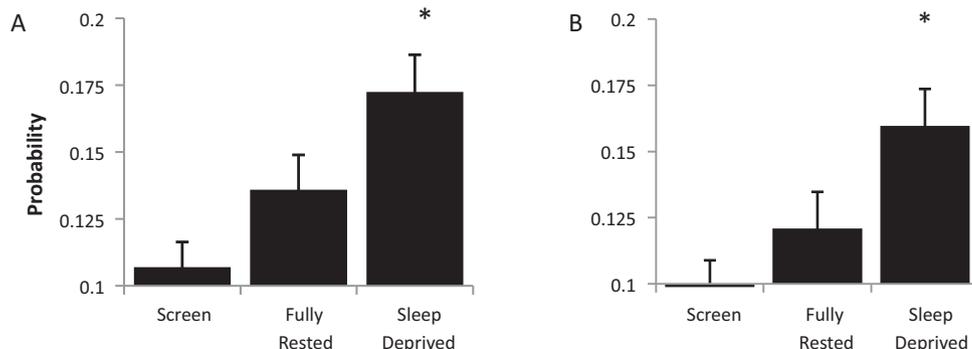
Fig. 3. Drowsiness increases significantly with sleep deprivation (study 1). Drowsiness increased significantly for 3CVT (A) and SIR (B), as well as IIR (not shown). Significant differences are noted with (*).
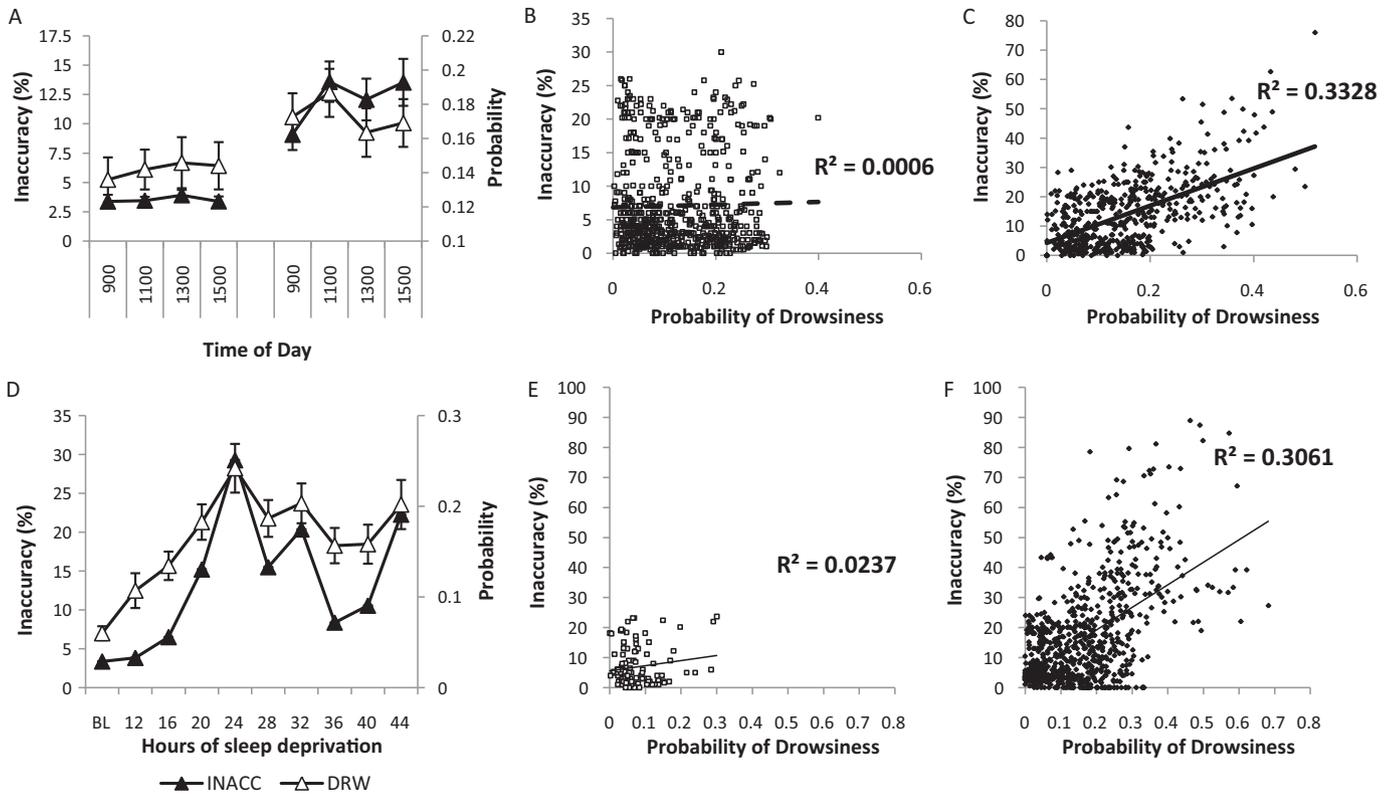
Fig. 4. The probability of drowsiness appeared to track drowsiness in both studies, in all neurocognitive tasks (data for 3CVT in studies 1 and 2 shown in A and D, respectively). Pearson's correlations confirmed a strong relationship for the sleep deprived state for both studies (C and F, respectively), but not when the subjects were rested (B and E, respectively). R^2 for each sleep deprived session are provided in the upper part of each graph, and rested in the lower part.

In addition to the neurocognitive test performance available in both studies, simulated driving performance was also assessed in study 1 at 2 time points under both rested and sleep deprived conditions. Pearson's correlation found that, as with the neurocognitive tests, drowsiness is significantly correlated with errors when subjects were sleep deprived ($r = .46$), but not when subjects were rested ($r = .17$, Fig. 5).

5. Discussion

The objective of these studies was to meet the following five criteria: (1) ensure maximal stability and inter-individual generalizability by using a large sample size, (2) accommodate individual

variability, (3) be applicable across tasks (task generalizability), (4) be computationally accurate and efficient for use in a portable un-tethered hardware application, and (5) provide a hardware platform to apply the algorithm in the field. The current study utilized data from a large sample size ($n = 160$) to build an algorithm that begins to address generalizability. This data set is substantially larger than the largest sample size in previously published drowsiness algorithms ($n = 30$) (Subasi, 2005). Individual variability was accommodated through individualization of the model centroids based on a 15 min set of 3 "ID" tasks (EO, EC, and the first 5 min of the 3CVT). This individualization method makes cross-validation mathematically invalid; thus, generalizability and validation had to be assessed in an alternative manner. We were able to demonstrate
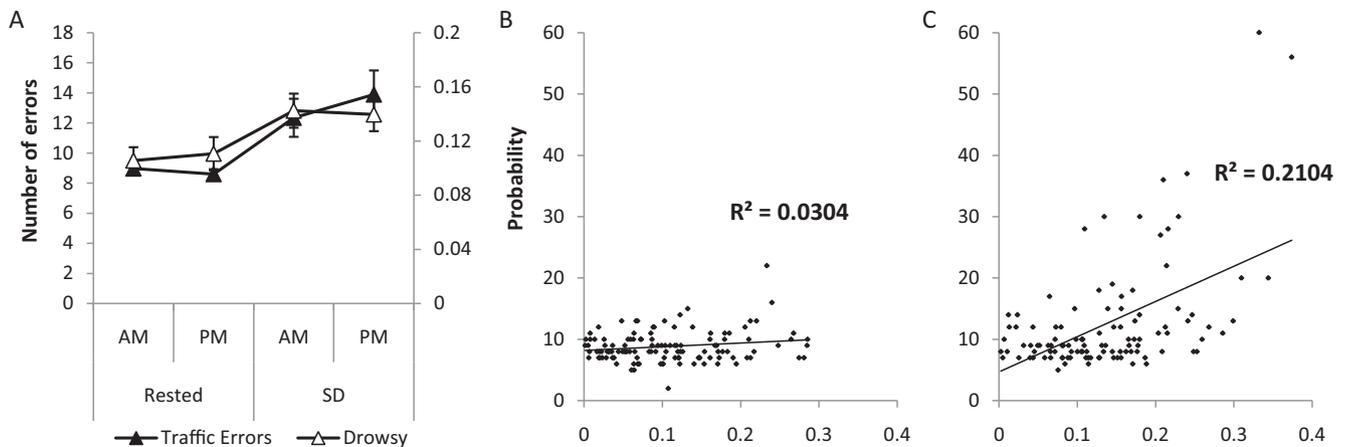
Fig. 5. As with the neurocognitive tasks, drowsiness and errors while driving track each other (A), and as with the neurocognitive tasks, drowsiness is significantly correlated with errors only when sleep deprived (C), not when rested (B). R^2 for each sleep deprived session are provided in the upper part of the graph, and rested in the lower part.

that the algorithm's designation of drowsiness (SO + DIS probability) tracked errors across multiple laboratory and simulation tasks, with significant correlation to error rates over time when subjects were sleep deprived. In order to accommodate computational efficiency, the algorithm was limited to data from Fz-POz and Cz-POz (as opposed to algorithms that required up to 32 channels). Additionally, as eye blink and muscle movement are inherently present in all EEG acquisition, the current system is able to identify a range of subtle and large artifacts (including eyeblinks) using EEG alone, removing the need to use EOG. The algorithms developed for EEG artifact decontamination can identify and remove contaminated signals automatically throughout acquisition, whereas past studies have relied upon offline post-hoc analysis (Gevins et al., 1990; Santamaria et al., 1987; Coenen, 1995; Horne and Reyner, 1995; Makeig and Jung, 1996; Huang et al., 2005).

Error prediction is the only acceptable outcome of a drowsiness detection application in the field to avoid both fatal and non-fatal errors. While the current method provides an excellent foundation, these data indicate that the current solution must address several shortcomings. First, the current method requires an acquisition PC (laptop, palm pilot, etc.) to be within 30 ft, however, the algorithm can be programmed on an ASCII chip, allowing for onboard data collection with the headset alone for up to 12 h. This step has been evaluated as a "proof of concept," and the current algorithm could be acceptable in such an application. Second, the current algorithm, while quite effective under sleep deprived conditions, was less effective when subjects were rested, leading to a great potential for false alarms. Such false alarms would highly diminish productivity, reduce adoption, and reduce compliance with policies based on drowsiness assessment. Third, though the current solution was found to be effective across multiple tasks, these tasks were all tested under laboratory conditions. For example, the simulated driving did not include scenarios reminiscent of rush hour type traffic, where little movement occurs for long periods of time, but the density and potential for errors is greater. There is no data available to compare how the drowsiness algorithm might perform under either simulated "rush hour" traffic or actual driving conditions that shift from low density to high density depending upon the path chosen, time constraints, skill level of driver, and other variables. If an algorithm is to be adopted, such a comparison would be required prior to adoption. Finally, in addition to the discrepancy in accuracy between tracking errors under rested conditions and sleep deprived conditions, the current algorithm does not predict performance decrements. These shortcomings will be addressed in future development projects that will include field studies to determine the applicability of the algorithm outside of the laboratory.

The algorithm was applied in a recent field study using professional truck drivers driving on a 37 km closed driving track located in rural Germany. This study presented a very limited number of stimuli (18 in a 7 h protocol) that drivers were instructed to identify as they drove 6 laps around a 37 km track at approximately 40 km/h. These data found that fewer errors occurred during the morning session, when the drivers were fully rested, and the few errors that did occur were not related to the drowsiness metric. On the other hand, a relationship began to appear in the afternoon, when errors were associated with elevated drowsiness levels in the final two laps (although no significant correlation was found) (Bingham and Kincses, 2008). These data are consistent with the data presented herein, whereby the drowsiness metric does not track errors for fully rested persons, and only begins to correlate as participants grow fatigued near the end of the day. These data support the need for further development to avoid false alarms and misses in the field, as well as the need for further field studies.

If any drowsiness algorithm is to prove useful in preventing accidents, it must meet the criterion discussed above. This is a shared shortcoming with most other drowsiness detection algo-

rithms reported thus far (with the exception being the actigraph solution proposed by researchers at Walter Reed) (Balkin et al., 2002). A predictive solution is under investigation at this time. Any such solution must have a limited false alarm rate, as well as a very low miss rate, to ensure that it is useful in improving public health and safety environments.

One potential explanation for the failure of the current algorithm to correlate with performance under rested conditions is individual variability. Individual variability occurs in performance, with some persons learning a task more quickly and accurately than others. While the current strategy of using an individual's coefficient matrix to build an individualized model may work for many subjects, further individualization may be required. For instance, the current solution relies on a single underlying model, while recent studies indicate that three or more "phenotypes" may exist, and thus a general model for each might be more effective (Doran et al., 2001; Van Dongen and Dinges, 2001; Rajaraman et al., 2008; King et al., 2009). Our data supports this hypothesis as well. In study 2 (n = 25), we found that some individuals (n = 4, 16%) were impervious to sleep deprivation and performed within normal parameters even after 40+ h without sleep. On the other hand, we also identified a small number of individuals (n = 6, 24%) that were already significantly impaired when only 12–18 h had passed since they last slept (i.e., at 7–11 pm in study 2). Thus, a single general model may not be sufficient, even with individualization. The underlying model may vary based on vulnerability to sleep deprivation or other variables (such as age or gender). These concerns will be further evaluated in future development projects.

In addition, the current algorithm relies on the assumption that a person's basal EEG and performance in the rested condition remains stable over a period of time. We were able to evaluate stability over a 24 h period: from the screening to the rested days in study 1. Further assessment is required, however, to determine if fully rested EEG changes over a period of days, weeks, or months.

The current data are only the first step in developing a drowsiness detection system that can be implemented to prevent industrial and/or vehicular errors associated with drowsiness, and additional work is required to have a fully filed validated and deployable algorithm. At this time the algorithm has shown to be robust in tracking inaccuracy and errors associated with sleep loss across multiple tasks (3CVT, IR, IIR, and DRIVE). In order to provide a useful tool for broad adoption in vehicular and/or industrial settings, the algorithm must further be developed in order to identify states that predict the onset of increased errors, with enough lead time to allow for an appropriate intervention to occur, to reduce false alarms associated with rested conditions, and ensure stability of the algorithm over time. Current investigations are underway to develop such a predictive algorithm. While the current algorithm has not demonstrated this level of utility, it has proven useful in multiple applications, some with an emphasis on drowsiness, others with no such emphasis. The algorithm has been utilized in altering information flow to increase productivity without overloading the user (Berka et al., 2004, 2007a,b; Berka, 2008). The algorithm has proven useful in other research environments as well, particularly in field applications (Stevens et al., 2006, 2007a; Berka et al., 2007b; Bingham and Kincses, 2008).

The demands of the global society for round-the-clock operations are likely to continue to increase the incidence of sleep deprivation worldwide. Although the deleterious effects of fatigue as a result of sleep deprivation or untreated sleep disorders on public safety are as well documented as those of alcohol intoxication (Jones et al., 2006; Howard et al., 2007), these findings have not yet been operationalized as policy or legal initiatives. One reason frequently cited for the lack of policy or regulation of driver drowsiness is the ongoing debate over whether an accurate and reliable assay for drowsiness can be developed for routine use in field appli-

cations. An equally important practical concern is to determine what level of fatigue causes performance impairment sufficient to result in motor vehicle and other accidents. Alternatively, when implemented in a closed-loop alarm feedback system, a drowsiness detection device can empower individuals to better monitor their levels of fatigue and make informed decisions regarding their level of risk to themselves and others. The system described herein (deemed the "B-Alert" method) holds potential to address these unmet needs for objective quantification of drowsiness to assist healthcare providers and other officials tasked with ensuring public safety. Utilization of the B-Alert system/method in industrial risk mitigation also has great potential. It is task independent, i.e., drowsiness detection does not rely on task specific metrics, and thus the system can potentially be applied in many different real-world environments (Stevens et al., 2006, 2007b; Berka, 2008).

Acknowledgements

References

Balkin, T.J., Belenky, G.L., et al., 2002. System and method for predicting human cognitive performance using data from an actigraph. 340/573.1. USPO. US. US 2002/0005784 A1: 21.

Baulk, S.D., Biggs, S.N., et al., 2008. Chasing the silver bullet: measuring driver fatigue using simple and complex tasks. Accident Analysis and Prevention 40 (1), 396–402.

Berka, C., 2008. Assessing Students Mental Representatives of Complex Problem Spaces with EEG Technologies. HFES, NY, New York.

Berka, C., Davis, G., Johnson, R., Levendowski, D., Whitmoyer, M., Fatch, R., Murray, S., Ensign, W., Yanagi, M., Olmstead, R., 2007a. Psychophysiological profiles of sleep deprivation and stress during marine corps training. Sleep 30, A132.

Berka, C., Levendowski, D., et al., 2004. Real-time analysis of EEG indices of alertness, cognition, and memory acquired with a wireless EEG headset. International Journal of Human–Computer Interaction 17 (2), 151–170.

Berka, C., Levendowski, D., et al., 2007b. EEG correlates of task engagement and mental workload in vigilance, learning and memory tasks. Aviation Space and Environmental Medicine 78 (5), B231–B244.

Bingham, R., Kincses, W.E., 2008. Final Report: Augmented Cognition – Phase 4 Cognitive Assessment and Task Management (CAT-M) D. o. Defense. Office of Naval Research, Arlington, VA.

Blood, M.L., Sack, R.L., et al., 1997. A comparison of sleep detection by wrist actigraphy behavioral response and polysomnography. Sleep 20 (6), 388–395.

Broughton, R., et al., 1988. Ambulatory 24 hour sleep-wake monitoring in narcolepsy-cataplexy compared to matched controls. Electroencephalography and Clinical Neurophysiology 70 (6), 473–481.

Calvin, A.D., Albuquerque, F.N., et al., 2009. Obstructive sleep apnea, inflammation, and the metabolic syndrome. Metabolism Syndrome and Related Disorder 7 (4), 271–278.

Chiou, J.C., Ko, L.W., et al., 2006. Using novel MEMS EEG sensors in detecting drowsiness application. In: Biomedical Circuits and Systems Conference, London.

Coenen, A.M., 1995. Neuronal activities underlying the electroencephalogram and evoked potentials of sleeping and waking: implications for information processing. Neuroscience and Biobehavioral Reviews 19 (3), 447–463.

Dinges, D., Kribbs, N.B., 1990. Comparison of the effects of alcohol and sleepiness on simple reaction time performance: enhanced habituation as a common process. Alcohol Drugs Driving 6 (1), 329–339.

Dinges, D.F., Kribbs, N.B., 1991. Performing while sleepy: effects of experimentally-induced sleepiness. In: Monk, T.H. (Ed.), Sleep Sleepiness and Performance. John Wiley and Sons Ltd., Chichester, U.K..

Dinges, D.F., Pack, F., et al., 1997. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. Sleep 20 (4), 267–277.

Dinges, D.F., Powell, J.W., 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behavior Research Methods Instruments and Computers 17, 652–655.

Dinges, D.F., 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 Medicine 4 (5), 393–402.

Doran, S.M., Van Dongen, H.P., et al., 2001. Sustained attention performance during sleep deprivation: evidence of state instability. Archives Italiennes de Biologie 139 (3), 253–267.

Edd, E.M., Flores, S., 2009. Sleepiness or excessive daytime somnolence. Geriatric Nursing 30 (1), 53–60.

Fletcher, A., McCulloch, K., et al., 2005. Countermeasures to driver fatigue: a review of public awareness campaigns and legal approaches. Australian and New Zealand Journal of Public Health 29 (5), 471–476.

Fournier, P.S., Montreuil, S., et al., 2007. Fatigue management by truck drivers in real life situations: some suggestions to improve training. Work 29 (3), 213–224.

Fu, J., Li, M., et al., 2008. Detecting drowsiness in driving simulation based on EEG. In: Proceedings of the 8th International Workshop, Shanghai, Springer, Netherlands.

Gevins, A.S., Bressler, S.L., et al., 1990. Effects of prolonged mental work on functional brain topography. Electroencephalography and Clinical Neurophysiology 76 (4), 339–350.

Gundel, A., et al., 1995. Sleepiness of civil airline pilots during two consecutive night flights of extended duration. Biological Psychology 40 (1–2), 131–141.

Higuchi, S., Liu, Y., et al., 2001. Diurnal variations in alpha power density and subjective sleepiness while performing repeated vigilance tasks. Clinical Neurophysiology 112 (6), 997–1000.

Horne, J.A., Reyner, L.A., 1995. Driver sleepiness. Journal of Sleep Research 4 (S2), 23–29.

Howard, M.E., Jackson, M.L., et al., 2007. The interactive effects of extended wakefulness and low-dose alcohol on simulated driving and vigilance. Sleep 30 (10), 1334–1340.

Huang, R.S., Jung, T.P., et al., 2005. Analyzing event-related brain dynamics in continuous compensatory tracking tasks. In: Conf Proc IEEE Eng Med Biol Soc 6, pp. 5750–5753.

Ives, J.R., Mainwaring, N.R., inventors; Beth Israel Hospital Association, assignee. Ambulatory electroencephalography system. US patent 522503. 1993.

Jones, C.B., Dorrian, J., et al., 2006. Self-awareness of impairment and the decision to drive after an extended period of wakefulness. Chronobiology International 23 (6), 1253–1263.

Jung, T., Makeig, S., 1994. Estimating levels of alertness from EEG. IEEE, 1103–1104.

Kamdar, B.B., Kaplan, K.A., et al., 2004. The impact of extended sleep on daytime alertness, vigilance, and mood. Sleep Medicine 5 (5), 441–448.

Karis, D., Fabiani, M., et al., 1984. P300 and memory: individual differences in the von Restorff effect. Cognitive Psychology 16, 177–216.

King, A.C., Belenky, G., et al., 2009. Performance impairment consequent to sleep loss: determinants of resistance and susceptibility. Current Opinion in Pulmonary Medicine.

Krieger, A.C., Ayappa, I., et al., 2004. Comparison of the maintenance of wakefulness test (MWT) to a modified behavioral test (OSLER) in the evaluation of daytime sleepiness. Journal of Sleep Research 13 (4), 407–411.

Lal, S.K., Craig, A., 2001. A critical review of the psychophysiology of driver fatigue. Biological Psychology 55 (3), 173–194.

Levy, P., Bonsignore, M.R., et al., 2009. Sleep, sleep-disordered breathing and metabolic consequences. European Respiratory Journal 34 (1), 243–260.

Liang, S.F., Lin, C.T., et al., 2005. Monitoring driver's alertness based on the driving performance estimation and the EEG power spectrum analysis. In: Conf Proc IEEE Eng Med Biol Soc 6, pp. 5738–5741.

Lim, C.L., Gordon, E., et al., 1999. Dynamics of SCR, EEG, and ERP activity in an oddball paradigm with short interstimulus intervals. Psychophysiology 36 (5), 543–551.

Lin, C.T., Ko, L.W., et al., 2006. Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks. IEEE Transactions on Circuits and Systems 53 (11), 2469–2476.

Makeig, S., 1993. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalography and Clinical Neurophysiology 86 (4), 283–293.

Makeig, S., Elliott, F.S., et al., 1994. First Demonstration of an Alertness Monitoring Management System. Naval Health Research Center, San Diego, CA, 22.

Makeig, S., Inlow, M., 1993. Lapses in alertness: coherence of fluctuations in performance and EEG spectrum. Electroencephalography and Clinical Neurophysiology 86 (1), 23–35.

Makeig, S., Jung, T.P., 1995. Changes in alertness are a principal component of variance in the EEG spectrum. Neuroreport 7 (1), 213–216.

Makeig, S., Jung, T.P., 1996. Tonic, phasic, and transient EEG correlates of auditory awareness in drowsiness. Brain Research: Cognitive Brain Research 4 (1), 15–25.

Mallis, M., 1999. Ocular measures as an index of fatigue and as the basis for alertness management: experiment on performance-based validation of technologies. In: Conference on Ocular Measures of Driver Alertness, Herndon, Virginia, Office of Motor Carrier and Highway Safety/Federal Highway Administration.

Miller, J.C., 1995. Batch processing of 10 000 h of truck driver EEG data. Biological Psychology 40 (1–2), 209–222.

NTSB, 1990. Fatigues, Alcohol, Other Drugs and Medical Factors in Fatal to the Driver Heavy Truck Crashes. S. Study. Washington, DC. NTSB/SS-90/01 and NTSV/SS-90/02.

NTSB, 1999. Evaluation of U.S. Department of Transportation Efforts in the 1990s to Address Operator Fatigue. NTSB. Washington, DC. SR-99/01.

Pal, N.R., Chuang, C.Y., et al., 2008. EEG-based subject- and session-independant drowsiness detection: an unsupervised approach. In: EURASIP J Adv in Sig Proc 2008.

Powell, N.B., Schechtman, K.B., et al., 2001. The road to danger: the comparative risks of driving while sleepy. Laryngoscope 111 (5), 887–893.

Rajaraman, S., Gribok, A.V., et al., 2008. Individualized performance prediction of sleep-deprived individuals with the two-process model. Journal of Applied Physiology 104 (2), 459–468.

Randerath, W.J., Gerdesmeyer, C., et al., 2000. A test for the determination of sustained attention in patients with obstructive sleep apnea syndrome. Respiration 67 (5), 526–532.

Richardson, G.S., et al., Wyatt, J.K., et al., 1996. Objective assessment of sleep and alertness in medical house staff and the impact of protected time for sleep. Sleep 19 (9), 718–726.

Sambeth, A., Maes, J.H., et al., 2004. Enhanced re-habituation of the orienting response of the human event-related potential. Neuroscience Letters 356 (2), 103–106.

Santamaria, J., et al., 1987. The EEG of drowsiness in normal adults. Journal of Clinical Neurophysiology 4 (4), 327–382 (review available).

Sigurdson, K., Ayas, N.T., 2007. The public health and safety consequences of sleep disorders. Canadian Journal of Physiology and Pharmacology 85 (1), 179–183.

Sing, H.C., Kautz, M.A., et al., 2005. High-frequency EEG as measure of cognitive function capacity: a preliminary report. Aviation Space and Environmental Medicine 76 (7 Suppl.), C114–135.

Smith, M.E., McEvoy, L.K., et al., 2002. The impact of moderate sleep loss on neurophysiologic signals during working-memory task performance. Sleep 25 (7), 784–794.

Stevens, R., Galloway, T., et al., 2006. Integrating EEG models of cognitive load with machine learning models of scientific problem solving. In: Schmorrow, D., Stanney, K., Reeves, L. (Eds.), Augmented Cognition: Past, Present and Future. Strategic Analysis Inc., Arlington, VA, pp. 55–65.

Stevens, R., Galloway, T., et al., 2007a. EEG-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills. In: User Modeling Conference, Athens, Greece.

Stevens, R., Galloway, T., et al., 2007b. Exploring neural trajectories of scientific problem solving skill acquisition. In: Schmorrow, D., Stanney, K., Reeves, L. (Eds.), Augmented Cognition: Past, Present and Future. Springer-Verlag Berlin, Arlington, VA, p. 16.

Subasi, A., 2005. Application of classical and model-based spectral methods to describe the state of alertness in EEG. Journal of Medical Systems 29 (5), 473–486.

Subasi, A., Ercelebi, E., 2005. Classification of EEG signals using neural network and logistic regression. Computer Methods and Programs in Biomedicine 78 (2), 87–99.

Surani, S., Aguillar, R., et al., 2009. Influence of Hispanic ethnicity in prevalence of diabetes mellitus in sleep apnea and relationship to sleep phase. Postgraduate Medicine 121 (5), 108–112.

Tabachnick, B.G., Fidell, L.S., 1983. Using Multivariate Statistics. Harper & Row, New York.

Tabachnick, B.G., Fidell, L.S., 2007. Using Multivariate Statistics. Pearson/Allyn & Bacon, Boston.

Torsvall, L., Akerstedt, T., 1987. Sleepiness on the job: continuously measured EEG changes in train drivers. Electroencephalography and Clinical Neurophysiology 66 (6), 502–511.

Valencia-Flores, M., Bliwise, D.L., et al., 1996. Cognitive function in patients with sleep apnea after acute nocturnal nasal continuous positive airway pressure (CPAP) treatment: sleepiness and hypoxemia effects. Journal of Clinical and Experimental Neuropsychology 18 (2), 197–210.

Van Dongen, H., Dinges, D., 2001. Modeling the effects of sleep debt: on the relevance of inter-individual differences. SRS Bulletin 7 (2), 69–72.

Van Dongen, H.P., Baynard, M.D., et al., 2004. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep 27 (3), 423–433.

Vuckovic, A., Radivojevic, V., et al., 2002. Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. Medical Engineering and Physics 24 (5), 349–360.

Weinstein, M., Silverstein, M.L., et al., 1999. Sustained attention on a continuous performance test. Perceptual and Motor Skills 89 (3 Pt 1), 1025–1027.

White, D.P., Gibb, T.J., et al., 1995. Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home. Sleep 18 (2), 115–126.

Wilkinson, R.T., 1990. Response-stimulus interval in choice serial reaction time: interaction with sleep deprivation, choice, and practice. Quarterly Journal of Experimental Psychology A 42 (2), 401–423.

Wilkinson, R.T., Houghton, D., 1982. Field test of arousal: a portable reaction timer with data storage. Human Factors 24 (4), 487–493.

Wilson, G.F., Russell, C.A., 2003. Operator funcitonal state classificaiton using multiple psychophysiological features in an air traffic control task. Human Factors 45 (3), 381–389.

Wong, K.K., Marshall, N.S., et al., 2008. Comparing the neurocognitive effects of 40 h sustained wakefulness in patients with untreated OSA and healthy controls. Journal of Sleep Research 17 (3), 322–330.