Biological Psychology 87 (2011) 241–250

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

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

Article history:
Received 16 June 2010
Accepted 10 March 2011
Available online 17 March 2011

Keywords:
Drowsy
Alertness
Electroencephalography

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.

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).

0301-0511/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.biopsycho.2011.03.003

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).