

## Vigilance state fluctuations and performance using brain–computer interface for communication

Barry Oken, Tab Memmott, Brandon Eddy, Jack Wiedrick & Melanie Fried-Oken

To cite this article: Barry Oken, Tab Memmott, Brandon Eddy, Jack Wiedrick & Melanie Fried-Oken (2018) Vigilance state fluctuations and performance using brain–computer interface for communication, Brain-Computer Interfaces, 5:4, 146-156, DOI: [10.1080/2326263X.2019.1571356](https://doi.org/10.1080/2326263X.2019.1571356)

To link to this article: <https://doi.org/10.1080/2326263X.2019.1571356>



Published online: 04 Feb 2019.



Submit your article to this journal [↗](#)



Article views: 2



View Crossmark data [↗](#)

# Vigilance state fluctuations and performance using brain–computer interface for communication

Barry Oken<sup>a</sup>, Tab Memmott<sup>a</sup>, Brandon Eddy<sup>b</sup>, Jack Wiedrick<sup>c</sup> and Melanie Fried-Oken<sup>b</sup>

<sup>a</sup>Department of Neurology, Oregon Health & Science University, Portland, OR, USA; <sup>b</sup>Institute on Development and Disability, Oregon Health & Science University, Portland, OR, USA; <sup>c</sup>Biostatistics and Design Program, Oregon Health & Science University, Portland, OR, USA

## ABSTRACT

The effect of fatigue and drowsiness on brain–computer interface (BCI) performance was evaluated. Twenty healthy participants performed a standardized 11-min calibration of a Rapid Serial Visual Presentation BCI system 5 times over 2 h. For each calibration, BCI performance was evaluated using area under the receiver operating characteristic curve (AUC). Self-rated measures were obtained following each calibration including the Karolinska Sleepiness Scale and a standardized boredom scale. Physiological measures were obtained during each calibration including P300 amplitude, theta power, alpha power, median power frequency, and eye-blink rate. There was a significant decrease in AUC over the five sessions. This was paralleled by increases in self-rated sleepiness and boredom and decreases in P300 amplitude. Alpha power, median power frequency, and eye-blink rate also increased but more modestly. AUC changes were only partly explained by changes in P300 amplitude. There was a decrease in BCI performance over time that related to increases in sleepiness and boredom. This worsened performance was only partly explained by decreases in P300 amplitude. Thus, drowsiness and boredom have a negative impact on BCI performance. Increased BCI performance may be possible by developing physiological measures to provide feedback to the user or to adapt the classifier to state.

## ARTICLE HISTORY

Received 11 September 2018  
Accepted 15 January 2019

## KEYWORDS

Drowsiness; performance; boredom; vigilance; P300

## 1. Introduction

Brain–computer interface (BCI) systems for communication have relied on neurophysiologic signals for intent selection of letters and words [1]. With the ultimate goal of serving as an augmentative and alternative communication tool for people with severe speech and physical impairments, BCI systems are of particular interest for individuals with brainstem strokes or neurodegenerative disorders when traditional access methods have been ineffective due to limited voluntary motor function [2–4].

One paradigm referred to as the P300-based speller is popular for communication BCIs. Target letters are interspersed with non-targets in an oddball paradigm. Perception of the salient target results in an attentional event-related response (e.g. a P300). The stimuli used to generate the attentional potentials generally appear in various displays, from Rapid Serial Visual Presentation (RSVP) [5,6] to matrix speller paradigms [7,8] or nested presentations such as the Hex-O-Spell or Shuffle Speller interfaces [9,10]. More complex systems have merged P300 with steady-state visual evoked potentials or even used non-visual modalities [11,12]. The BCI system relies on various machine learning

approaches to detect the presence of the target from just several target stimuli.

Attentional potentials are known to be sensitive to the vigilance state of the individual. State-dependent alterations in attentional biomarkers have been identified, such as P300 amplitude [13–15] or event-related desynchronization [16].

Intra-individual performance variability associated with vigilance changes is related, in part, to variation in attention performance. Performance variability has been studied across multiple time frames, spanning from seconds to years, using methods ranging from single cell recordings [17] to fMRI [18]. Intra-individual variation in cognitive performance is greater with aging [19], clinical disorders including traumatic brain injury [20,21], sleep disturbances [22], vigilance decrements [23], and medication [24].

Intra-individual performance variability is in part related to changes in phasic and tonic activation states of cortex that modulate cortical signal processing. The terminology used to describe these activation states varies, depending on the specific field. Arousal level on the wake-sleep spectrum is one aspect of the variability. Drowsiness or stage-1 sleep with the eyes closed is well-described from

the clinical EEG perspective [25]. The clinical sleep field has established a scoring system for stage-1 sleep based on slower frequency EEG, decreased eye blinks, and increased slow lateral eye movements [26]. An individual who, by EEG criteria, is drowsy or in stage-1 sleep contrasted to wakefulness, will have poorer ability to process information, including BCI stimuli. Other conceptual states such as vigilance, mental fatigue, and boredom also will impact cognitive processing and produce intra-individual variability in performance. Vigilance decrement is a term used to describe the decline in performance, usually on a repetitive or boring task over time [27,28]. The vigilance decrement is sometimes related to sleepiness but is conceptually independent because it extends into boredom – another state of inattention related to lack of motivation, mental fatigue, and the vigilance decrement [29].

BCI research and development would benefit from a better understanding of performance variability related to state ranging from seconds to hours in both healthy populations and people with disabilities with the intention to maximize performance. As stated by Millan et al. [30], there is a need for optimized human-computer interactions in which the interface ‘... adjust[s] the dynamics of interaction as a function of the user’s control capabilities’. These control capabilities extend into both attentional and cognitive resource availability.

Automated drowsiness and vigilance detection systems have been used for driver and cognitive performance fatigue, including for BCI [31–35]. Some researchers have even begun to incorporate functional near-infrared spectroscopy to drowsy detection [36,37]. While many psychophysiology researchers have related non-BCI performance errors to EEG measures of vigilance, few researchers have specifically looked at how BCI performance changes relate to EEG vigilance changes. Recently, a study on BCI for communication with participants with locked-in syndrome showed that, although performance was generally low, there was a negative correlation between accuracy and theta-band power, possibly related to decreased vigilance [38]. A recent study took a novel approach to calculation of state via the use of a multimodal design [31]. Data were extracted from secondary sensors of eye activity and EEG which greatly improved the classification of reduced alertness. Increased blink rate was identified as the most sensitive among several metrics of eye activity associated with increased mental fatigue.

We propose to explain the influence of attentional changes in BCI performance from both neurophysiologic and behavioral measures. First, we hypothesize that repeated RSVP BCI sessions will induce increased drowsiness and boredom. The task consists of detecting a target letter while viewing single letters sequentially

flashed on a display and thus is not cognitively demanding. While not cognitively demanding, it clearly requires sustained attention to the current moment that is discussed in varied contexts such as vigilance [23], flow theory [39], and even mindfulness meditation [40]. This increased drowsiness and boredom will be reflected in self-rated measures as well as physiological changes. The physiological changes will include decreased P300 amplitude (drowsiness and boredom P300 paper) and thus BCI performance. In addition, the increased drowsiness will be associated with increased eye blink rate [41], increased alpha and theta power [25] and likely increased median power frequency which is highly related to the high percentage of EEG power contained in the alpha frequency range. Within-subject comparisons will further analyze whether any observed performance declines are related to P300 amplitude reductions or other EEG changes. This may determine whether these measures would be useful as within- or between-subject predictors of BCI performance.

## 2. Materials and methods

### 2.1. Participants

A convenience sample of 22 healthy adults from the Portland, Oregon USA metropolitan area participated in this study. All participants passed a telephone cognitive screening and provided demographic information, including years of education. Exclusion criteria included significant visual impairment, photosensitive seizures, or current use of medications that would likely affect EEG or wakefulness such as neuroleptics, narcotic analgesics, and benzodiazepines. Participants signed informed consent prior to enrollment. This study was reviewed and approved by the Oregon Health and Science University Institutional Review Board (#15,331). After consenting, a brief health screening was conducted to document changes in medical status, and recent caffeine intake, alcohol intake, nicotine intake, and sleep. Participants also passed a brief vision screening, including visual acuity at 2 ft.

Data from 20 participants were analyzed. Two participants were excluded: one participant was unable to complete the study due to discomfort with the electrode cap, and the second participant’s data were excluded from analysis due to significant artifact caused by the participant manually moving sensors during calibrations. Seventeen participants reported they were Caucasian, one African-American and two Asian-American, and one Caucasian reported being Hispanic. The mean age (SD) was 34.3 years (13.4), the mean duration of education was 16.8 years (2.5).

The cohort included 14 women and 6 men. The averaged total reported sleep the night before test session was 425.1 min (83.7).

## 2.2. Procedures

The experimental design was a single study visit with each participant performing five BCI calibration sessions on the same day with physiological measures obtained during each calibration. The primary outcome measure was classifier accuracy measured by AUC. At the start of the study visit, participants completed the Profile of Mood States (POMS) focusing on the Vigor and Fatigue subscales [42], the Pittsburgh Sleep Quality Index [43], and reported how much sleep they had the previous night (reported as the difference between bedtime and rising time). The WAIS-IV Letter-Number Sequencing (LNS) [44] was administered to evaluate the effect of working memory capacity on the calibration performance. Two self-rated questionnaires were completed after each calibration: sleepiness was assessed with the Karolinska Sleepiness Scale (KSS [45]), and boredom was assessed with the 6-item Boredom Questionnaire [46].

The RSVP Keyboard™ BCI system was used, with code that was developed in MATLAB (MathWorks, Natick, MA) [6,47]. The RSVP Keyboard™ is a non-invasive, P300 speller that presents letters in a rapid serial visual presentation, and classifies letter selection based on EEG data. During generative spelling, which was not used in this experiment, the system fuses the EEG data with input from an integrated language model using probabilistic Bayesian fusion [48]. Twenty-eight characters including the English alphabet, < for backspace, and □ for space, were presented on a black background sequentially at 5 Hz.

A DSI-24 dry-electrode cap (Wearable Sensing, San Diego, CA), featuring 21 electrodes positioned at approximately 10–20 locations, with the ground at FpZ, was placed on the participant's scalp. Two ear-clip dry-electrodes were placed on the participant's earlobes as a reference. The ear-clip electrodes were built into the DSI-24 dry-electrode cap. The DSI-24 dry-electrode cap was connected directly via USB to the desktop. Each channel was visually inspected on the signal-monitoring graphical user interface (Wearable Sensing, San Diego, CA) that displayed EEG signals from the active electrodes and allowed signal quality to be verified before the user began the BCI calibration. We used Wearable Sensing's Diagnostic Tab to confirm good electrode contact on setup and between calibration sessions. Good contact indicates impedance less than 1 Mohm, voltage less than 5000 uV, and root-mean-square less than 15 uV in all channels. The EEG researcher

ensured good signal quality for alpha and other EEG activity and that there was electromyographic activity during jaw clenching. Electrodes were adjusted as necessary to achieve maximal signal quality. For example, the researcher adjusted electrodes if there was significant sweat artifact or cardioballistic activity. Signal quality was re-examined between each calibration session.

For the calibration task, participants were asked to sit stationary in a chair exactly 24 inches from a Dell 1704FPT 17-in monitor (Dell Inc., Round Rock, TX) positioned at eye level. Participants were introduced to the RSVP Keyboard™ via a video recording presented on a separate nearby external monitor prior to completing a calibration. Each calibration required the participant to visually attend to the center of the monitor. Each calibration consisted of 100 trials and lasted 11 min. In each trial, a target letter of about four degrees visual angle was presented at the center of the screen and then a red '+' was presented in its place to prepare the participant for the trial. A series of 10 letters (including the target letter) then was presented at a rate of five per second in the center of the screen. Participants were encouraged to select a metacognitive strategy to 'alter' their brain signal for the intended selection when they saw the target letter. A sample instruction was: 'When you see your target letter, you will need to do something to change your brain signal in order to "select". Some possible ways to change your brain signal might include thinking "BAM" when you see the target letter; pretend that you are clicking a mouse when you see the target letter, or; think about a sensation (like hot or cold) or a sound when you see the target letter.' Following the fifth and final calibration, participants were administered a second POMS questionnaire.

## 2.3. EEG processing

EEG was analog filtered by the DSI-24 from 0.003 to 150 Hz and sampled at 1200 samples per sec. The RSVP software used a Butterworth bandpass filter from 1.5 to 42 Hz and downsampled to 256 Hz [6,47]. Machine learning of RSVP data was trained on 500 msec of EEG following target and non-target letters using a regularized discriminant classifier following dimension reduction using a principal components analysis. The best parameters for the classification were obtained using a 10-fold leave-one-out cross-validation [6,47]. At each session, the Nelder-Mead simplex-reflection method [49] was used to optimize the free parameters such that a local maximizer of the area-under-the-ROC-curve (AUC) estimated using 10-fold cross-validation is achieved. The scores for each

validation fold are calculated and the 10 scores are concatenated to generate the session AUC. Other BCI performance metrics besides the AUC such as Information Transfer Rate [50] were not possible since only classifier calibration sessions were used.

Outcome measures that have been shown to be sensitive to performance declines related to vigilance decrements or drowsiness were analyzed: P300 amplitude, alpha power, theta power, MPF, and eye-blink rates.

EEG for P300 amplitude analysis was imported to EEGLAB v 14.1. Independent Component Analysis was used to remove eye blinks and segmented  $-100$  to  $1000$  ms around target letters. Data then were semi-automatically artifact-rejected based on the presence of muscular artifact or other larger interference not removed by filtering. P300 amplitude was calculated in the Cz channel with EEGLAB using a peak-to-trough method with the highest positive potential in the  $350$ – $600$  msec range after letter presentation as the peak and the most negative potential preceding the P300 peak as the trough.

The EEG for frequency analysis was segmented into 4-s epochs since the stimulus presentation sequence including the initial target letter presentation lasted approximately 4 s. Semi-automatic artifact rejection as for the P300 was applied. Following application of a Hanning window using the MATLAB signal processing toolbox `hann()` function with periodic window sampling, frequency analysis was done by MATLAB's power spectral density function and averaged across all the epochs for each calibration session. Frequency band spectral density measures were calculated for theta ( $4$ – $7.75$  Hz) and alpha ( $8$ – $12.75$  Hz). MPF ( $2$ – $20$  Hz) was calculated using the MATLAB `medfreq()` function.

Eye-blink rates were obtained using an average of Fp1/2 channels and MATLAB's `findpeaks()` function with `MinPeakDistance = 75` samples and `MinPeakHeight = 50`  $\mu$ V. We used the time elapsed in seconds for 10 blinks to occur to determine blink rate and took the average of those values to calculate the session blink rate per second and per minute. To avoid pre- and post-trial artifacts, both blink rates, and the frequency measures were calculated from trial stimulus start time to end time as was done with the frequency measures.

### 3. Statistical analysis

A main goal of this study was to identify potentially useful predictors of BCI performance with respect to subject vigilance. A useful between-subject predictor could help with individualizing stimulus presentation and EEG classification; for example, alternative BCI

configurations could be applied to mitigate subject-specific deficits in vigilance. A useful within-subject predictor might enable real-time monitoring, adaptation, or feedback in response to vigilance decrements. Further, if a within-subject association does not hold cross-sectionally between subjects, that is an indication of (possibly unmeasured) intruding factors that differ from subject to subject. Given this broad goal to find associations that were predictive (high correlation), impactful on performance (meaningful effect size), and reliable across subjects (high statistical significance), we felt it was important to fully explore the space of correlations among the measures under consideration and report on all findings, regardless of strength or future promise. Due to the limited sample, highly interrelated measurements, and lack of an a priori causal model of the associations, we opted for one-by-one correlations as the most informative exploratory approach. We have not adjusted any of the p-values for multiple comparisons because such corrections are not meaningful when reporting all results; that is, the relative 'significance' among all findings is the same with or without adjustments, and since we are not making population-level claims at this juncture, the p-values are merely descriptive devices used to characterize reliability of associations within the sample at hand. We have not filtered any of the results presented here based on their p-values or other statistical summaries, and have attempted to present an appropriately tempered set of judgments concerning whether and to what degree certain variables appear to be important for the scientific question.

For between-subject analyses, we examined the relationships between session-static variables such as baseline sleep quality and across-session average AUC. These comparisons provide information about whether interindividual heterogeneities may be predictive of a subject's typical or future performance. Measures obtained only at baseline were Pittsburgh Sleep Quality Index (PSQI), self-reported minutes slept the night before, and LNS. POMS-fatigue and POMS-vigor were obtained at baseline and again at the end of the final calibration session. Mean AUC and mean P300 amplitude were obtained by averaging results from the five sessions. Pearson correlation coefficients were calculated for mean AUC versus age, years of education, LNS score, reported minutes of sleep the night before, PSQI, and the two POMS measures. To analyze the relation of these baseline measures to performance decrement, linearized within-subject change in AUC across sessions was estimated as the slope in a simple linear regression of the AUC on the session number, separately for each subject. The AUC slopes and the five individual session AUCs were also correlated to the baseline measures listed above

to examine whether the rate of decline or ‘eventual’ performance may be related to subject-specific characteristics. To estimate these correlations, we used Spearman correlation coefficients because of the likelihood that cumulative session effects would tend to differ by subject and create nonlinearities in the relationships between BCI performance and baseline measurements. (The Spearman statistic is more sensitive than Pearson to true correlations when the underlying relationship is nonlinear.)

For within-subject analyses, the primary outcome measure for each calibration session was classifier accuracy measured as AUC. The changing values across time for each participant were related to other measures of the same participant’s vigilance at the same timepoints. This analysis allows us to explore whether a subject’s changing internal state may be predictive of (and possibly causally related to) current BCI performance in real time. The other measures included self-report of sleepiness (KSS), self-report of boredom, and physiological measures potentially related to drowsiness: P300 amplitude, theta power, alpha power, MPF, and eye-blink rate. A complication here is that the five session values for a participant are not independent observations but rather show expected positive correlation by virtue of being close repeated measurements on the same subject; the actual amount of information available from the five sessions is called the ‘effective’ sample size (always less than five in this case), and is calculated as described below. Within-subject (i.e. session-by-session) correlations between AUC and P300 amplitude, KSS score, and boredom score were estimated using Bland and Altman’s method [51], implemented as a fixed-effects longitudinal model [52]; this method calculates the average correlation between the measures of interest observed in individual participants, while accounting for the expected correlation in all values due to the repeated measurements. Degrees of freedom for the significance calculations were taken as the number of subjects times the average effective sample size of a subject’s panel (calculated as  $5 / (1 + 4 * ICC)$ , where ICC represented the intraclass correlation coefficient estimated by the fixed-effects model for the outcome), minus 3. Between-subject correlations in these longitudinal measures (essentially, correlations of the across-session means) were estimated using the complementary ‘between-effects’ longitudinal estimator used in the calculation of the fixed-effects estimates [52].

As summaries of the changes in the longitudinal measures, we considered a marginal linearized change (i.e. the average of change rates across timepoints) as a characterization of effect size to indicate to what extent performance or vigilance was being affected by repeated trials of a boring task. The marginal linearized

change in AUC, P300 amplitude, and the KSS and boredom values across sessions was estimated using a mixed random-effects model of the outcome regressed on session number. This analysis allowed for subject-specific starting values and deviations from the average slope, with unstructured covariance between the random effects, assumed exponential decay of correlation in residual errors, and estimated covariance using the Huber-White robust (‘sandwich’) estimator [53]. In addition to marginal slope estimates from these models, effect sizes were calculated as change in outcome per standard deviation (SD) from the first session, and as the mean fraction of the previous session’s value (averaged over model predictions for the sessions).

Finally, marginal linear structural equation modeling [54] was used to estimate the fraction of variance in AUC explained by EEG values, adjusting for time (as session number), and to perform a mediation analysis for the hypothesis that P300 amplitude change mediates change in AUC. The assumed model for EEG relationship to AUC was:  $\text{time} \rightarrow \text{EEG} \rightarrow \text{AUC} \leftarrow \text{time}$ . Stata version 15.1 was used for all statistical analyses.

#### 4. Results

There was no indication of significant correlation of age with either the mean AUC or the AUC from the first session (largest correlation =  $-0.2$ ), but the correlation between age and the AUC decline was strongly negative ( $-0.4$ ,  $p = 0.00003$ ), indicating that older participants tended to show less performance decline. There was a positive correlation between years of education and mean AUC ( $0.5$ ,  $p = 0.014$ ), and the correlations with the session AUCs were consistently about  $0.4$  or  $0.5$  with generally low  $p$ -values (between  $0.015$  and  $0.15$ ); the AUC drop was also somewhat negatively correlated with years of education ( $-0.25$ ,  $p = 0.014$ ), suggesting that better-educated subjects may be both better-performing in general and able to sustain BCI performance for longer. LNS correlated highly significantly with the first-session AUC ( $0.7$ ,  $p = 0.0003$ ), but correlations with later sessions (and with the mean AUC) were negligible, and correlation with the AUC drop was positive and strong ( $0.3$ ,  $p = 0.002$ ). This pattern points to a relationship between working memory and peak performance. There is a faster loss of performance for those with better working memory as they become bored or sleepy.

Perhaps counterintuitively, reported minutes slept the night before was negatively correlated with both first-session AUC and mean AUC ( $-0.4$ ,  $p = 0.07$ ) but not AUC slope (magnitude only  $0.1$ ). Correlations

between mean AUC and PSQI, POMS-vigor (either at baseline or exit testing), and POMS-fatigue (either at baseline or exit testing) were all small (less than 0.2 in magnitude) with large p-values. Similarly, neither first-session AUC nor AUC drop was significantly correlated with PSQI or POMS-fatigue either at baseline or end of testing (all correlations 0.1 or less in magnitude). For POMS-vigor at baseline and exit testing, however, while the correlation with first-session AUC was extremely weak ( $-0.2$ ,  $p = 0.4$ ), the correlation with AUC drop was moderately suggestive at approximately the same magnitude ( $-0.2$  with  $p = 0.1$  at baseline,  $-0.3$  with  $p = 0.01$  at exit testing). This pattern is interesting as a hint that livelier subjects probably experience less decline in future BCI performance than less lively subjects and confirms the intuition that those who were more alert at the end of testing tended to be the same as those who had sustained their BCI performance better.

The between-subject correlation of mean AUC with mean P300 amplitude was positive but weak ( $0.3$ ,  $p = 0.2$ ), and the correlations with session AUCs declined in magnitude (from  $0.3$  to  $0.1$ ) as the sessions progressed, hinting that factors other than P300 amplitude became increasingly influential on AUC as boredom and sleepiness set in.

The key analysis for the primary aim demonstrated that performance significantly decreased over the five successive BCI sessions with AUC averages decreasing from  $0.88 \pm 0.05$  to  $0.79 \pm 0.10$ . The typical rate of drop was  $2 \pm 0.5$  points per session ( $z = -3.8$ ), which is roughly 2/5 of the first-session SD. See Table 1 and Figure 1. Self-rated measures of drowsiness (KSS) and boredom significantly increased: drowsiness  $z = 7.0$  and boredom  $z = 5.7$ . The increase in KSS was by 3/5 of the first-session SD and the increase in boredom was by 1/3 of the first-session SD, each increasing about 15% per session on average. P300 amplitudes significantly decreased over time, at an average rate of 11% decline per session ( $z = -4.7$ ).

There was an increase in alpha power ( $z = 2.6$ ) across the five sessions but this increase was small, by about 1/12 of the first-session SD. A comparably small increase in MPF was observed (1/11 of the first-session SD,  $z = 3.1$ ). Blink rate also increased, from 25.5 - per minute in the first session to 34.5 per minute in the last session ( $z = 3.1$ ). We observed a transient increase in mean theta power at session two (Table 1) due to a single subject extremely large outlier, but there was no apparent change in theta power over the remaining sessions and overall ( $z = -1.1$ ).

Within-subject correlations from the fixed-effects regressions showed that AUC was negatively correlated

**Table 1.** Data from all five sessions are presented as the mean  $\pm$  standard deviation in the first row and range [low, high] in the second row.

Session	AUC	KSS	Boredom	P300 Amplitude ( $\mu V$ )	Theta ( $\mu V^2/Hz$ )	Alpha ( $\mu V^2/Hz$ )	MPF (Hz)	Eye blink rate (blinks/min)
1	$0.876 \pm 0.051$ [0.763, 0.938]	$3.55 \pm 1.47$ [1,6]	$13.30 \pm 7.26$ [6,36]	$9.805 \pm 4.429$ [3.519, 20.790]	$1.214 \pm 0.797$ [0.456, 3.460]	$1.249 \pm 1.459$ [0.198, 6.940]	$7.960 \pm 1.296$ [5.288, 10.608]	$25.5 \pm 12.6$ [4.3, 53.4]
2	$0.841 \pm 0.084$ [0.622, 0.946]	$5.10 \pm 1.77$ [2,8]	$17.80 \pm 7.78$ [8,31]	$7.397 \pm 2.962$ [2.830, 12.629]	$2.390 \pm 4.320$ [0.440, 19.900]	$1.680 \pm 1.775$ [0.339, 7.980]	$7.997 \pm 1.578$ [4.892, 10.590]	$32.9 \pm 14.0$ [5.7, 71.6]
3	$0.816 \pm 0.102$ [0.555, 0.976]	$6.25 \pm 2.10$ [2,9]	$21.20 \pm 9.89$ [8,38]	$7.027 \pm 3.226$ [2.374, 13.243]	$1.274 \pm 0.659$ [0.451, 2.820]	$1.617 \pm 1.657$ [0.332, 7.680]	$8.184 \pm 1.370$ [5.222, 10.569]	$33.5 \pm 14.7$ [9.9, 72.5]
4	$0.812 \pm 0.100$ [0.659, 0.967]	$6.80 \pm 2.07$ [3,9]	$22.80 \pm 10.24$ [9,40]	$6.781 \pm 2.868$ [3.792, 13.266]	$1.191 \pm 0.582$ [0.400, 2.350]	$1.692 \pm 1.826$ [0.276, 7.600]	$8.337 \pm 1.354$ [5.294, 10.637]	$33.5 \pm 14.7$ [12.0, 72.6]
5	$0.794 \pm 0.096$ [0.608, 0.939]	$7.10 \pm 2.13$ [3,10]	$23.65 \pm 10.60$ [9,41]	$6.371 \pm 2.221$ [3.177, 11.998]	$1.227 \pm 0.604$ [0.414, 2.870]	$1.873 \pm 2.121$ [0.321, 9.490]	$8.385 \pm 1.397$ [5.427, 10.544]	$34.5 \pm 17.4$ [15.0, 75.1]

with KSS ( $-0.46$ ,  $p = 0.009$ ) and boredom ( $-0.47$ ,  $p = 0.015$ ), and positively correlated with P300 amplitude ( $0.64$ ,  $p = 0.00035$ ). These strong within-subject correlations indicate a tendency for all of these measures to change together over time for the same subject (see Figure 1). This may hint at causal connections among them, with implications for future experimental manipulations.

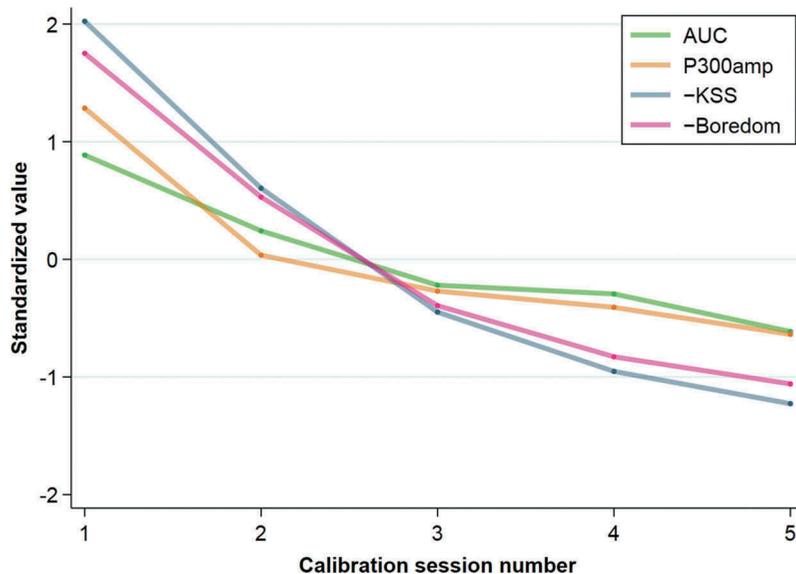
There was roughly a 3.3-point drop in each subject's AUC per unit SD drop in that subject's P300 amplitude. This trend was highly reliable across subjects ( $z = 5.3$ ). However, as already noted, there was only a weak between-subject correlation between mean P300 amplitude and mean AUC ( $r < 0.3$  and  $z < 1$ ) that actually declines (approaches zero in magnitude) as the sessions progress. This suggests there are factors other than P300 that become increasingly influential on AUC; there is considerable time- and subject-level confounding in the marginal relationship. That is, just knowing the subject's current or average P300 as a single summary is not good enough; one also needs to know how long the subject has been tested, whether their AUC is trending up or down, and other factors that have not yet been defined or measured. Further analysis using structural equation modeling revealed the AUC decrease over time was not fully explained by the P300 amplitude

decrease. The direct effect of time on AUC was  $-1.2$ /session ( $z = -2.0$ ), and the indirect effect through declining P300 amplitude was  $-0.7$ /session ( $z = -2.5$ ). Both the direct and indirect effects are significant, suggesting incomplete mediation of the AUC decline by the P300 decline (likelihood ratio test  $p = 0.0003$ ).

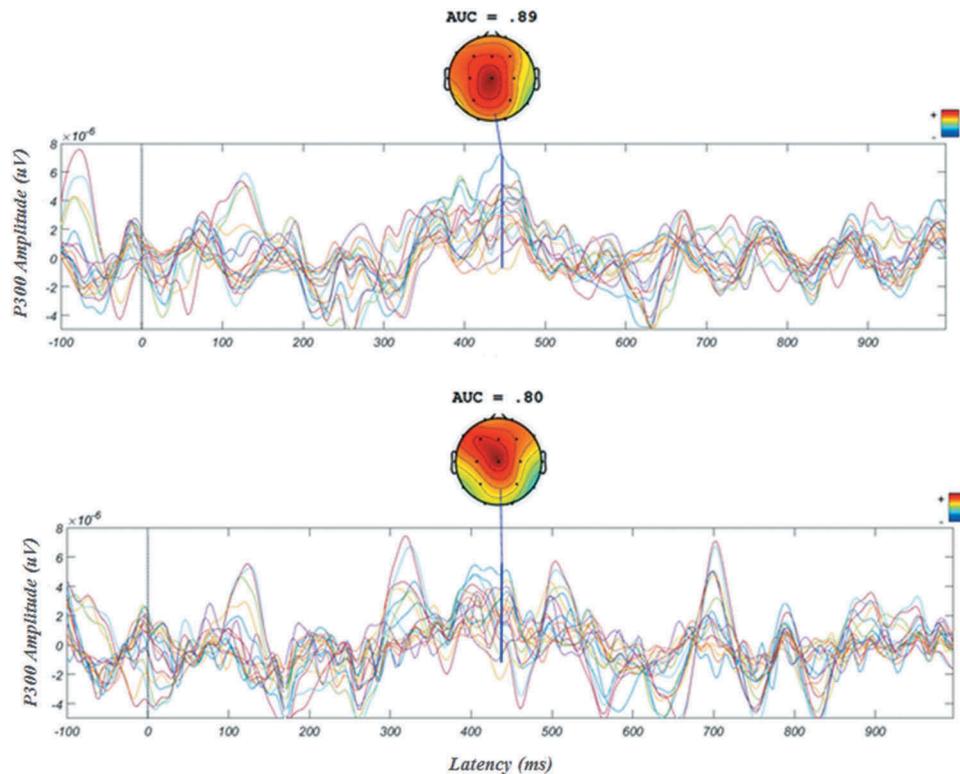
We found no evidence that the EEG measures of theta power, alpha power, or MPF (as within-session averages of 4-s epochs) were predictive of AUC. Associations with AUC were uniformly small, typically only about 1/2 point of AUC per SD change in the measure.

## 5. Discussion

While individuals with severe speech and physical impairments value BCI systems for communication, there are many variables that may affect their performance and challenge the functional use of this new technology [4]. This study demonstrated that BCI performance for calibration of a P300 speller by healthy adults is significantly affected by attentional factors as measured by both neurophysiologic and behavioral variables. There was a significant decline in BCI performance as measured by AUC over a 2-h period. This was accompanied by an increase in sleepiness and



**Figure 1.** Standardized values for AUC, P300 amplitude, KSS, and boredom scales plotted over calibration session number. Because KSS and boredom increase with time, their scores were inverted to better demonstrate the relationship of their trends to AUC and P300 amplitude. All significantly changed over the five sessions ( $p < 0.00001$ ). Individual values were standardized by the residual standard deviation estimate from fixed-effects regression on a polynomial in session number and averaged by session using the same framework. An approximate 2-standard-error lower bound for comparing the first session to any subsequent session is 0.7 units below the first session's value for any of these standardized measures, so horizontal error bars can be visualized accordingly. It is easy to see that every session is not just lower in value than the previous one but also falls more than 0.7 units below the first session; by the final session, all measures have dropped by at least 1.5 standard units (AUC) and by as much as 3.3 standard units (KSS).



**Figure 2.** Evoked potential to target letters from a single subject from the first and best (top) and last and worst (bottom) calibration session as assessed by AUC. The chosen participant had one of the biggest changes from first to last session. These figures display all 21 channels, and a heat map shows distribution of the P300. One can see the typical P300 maximal Cz topographical distribution and the significantly more consistent and higher amplitude P300 in the first compared with the last session.

boredom. The AUC had a strong within-subject negative correlation with both of these self-rated measures.

The P300 amplitude also significantly decreased over the five sessions, and the P300 amplitude had a significant within-subject correlation with AUC. Whatever someone's mean P300 amplitude happens to be, that person will likely demonstrate a decrease in AUC as their P300 amplitude decreases. While not unexpected, this confirms that the P300 is a major contributor to the BCI RSVP Keyboard classifier. However, our mediation analysis (of the hypothesis that P300 amplitude change completely mediates change in AUC) reveals that the AUC decrease over the five sessions was not solely attributable to the P300 amplitude decrease since the AUC decreases even after accounting for the partial mediation of AUC by the P300 amplitude. And despite the very strong within-subject correlation between AUC and P300 amplitude, there was no large between-subject correlation between these two measures. The only partial mediation of AUC decrease by P300 amplitude decrease and the lack of intersubject correlation between AUC and P300 amplitude suggests that the classifier was able to use time-varying subject-level information beyond the

target stimulus P300 amplitude. This information could be the amplitude difference between target and non-target responses, the topography of the P300, or other aspects of the target evoked response, such as event-related desynchronization [55,56] or synchrony [57].

Consistent with prior reports, alpha, MPF, and eye-blink rate all increased over sessions with increased fatigue (e.g. [31]). While eye-blink rate decreases in stage 1 sleep with eyes closed [25], blinking increases with drowsiness in eyes open in this and prior visual cognitive studies [31,58]. Blinking in an eyes-open context may be used as an intentional way to increase wakefulness. Increased alpha has been associated with early stage-1 sleep and this likely contributes to the increase in MPF before there is a significant increase in theta or other slow wave activity. Our surprising lack of association of performance with theta overall, which has been observed in other studies dating back 25 years [59], might be due to the overlap of this frequency band with our stimulation frequency (5 Hz). These EEG frequency measures are therefore highly task- and stimulation-dependent. They may also be highly individualized, as there were no apparent between-subject associations with performance even

though each subject's changes over the course of the experiment were related to performance.

### 5.1. Limitations and future directions

There are several limitations to this study. The outcome measures were obtained during BCI calibration sessions which likely have higher mental fatigue compared with a more natural communication task. The BCI performance and physiological measures were calculated over an entire calibration session (about 11 min duration) and shorter-term fluctuations almost definitely contribute to performance variability. Assessment of these shorter-term fluctuations in attention and drowsiness will be necessary in order to integrate attention and drowsiness detection into the real-time BCI use with the goal of providing feedback to the participant or adapting the classifier to state. The participants in this study are healthy adults, relatively young, 85% Caucasian, and highly educated. Any findings from this study relevant to other populations including those with clinical disorders will need to be confirmed. Related to this issue, performance was correlated to years of education and, for the first session, highly correlated to a conventional working memory measure, LNS. This may in part explain the discrepancies across BCI studies in performance between healthy controls and clinical populations. The decline in performance over 2 h from an AUC of 0.88–0.79, while not dramatic, is sufficient to have an impact on actual communication and likely would be greater in clinical populations who present with challenges that affect vigilance systems. Individuals with various stages of locked-in syndrome do not have the ability to control schedules for sleep-wake cycles; they often must rely on caregivers whether they are tired or not, and do not set their daily activities. Additionally, brain injuries or medications may affect attention. Lack of daily stimulation may induce levels of inattention and drowsiness that is not considered within this experiment. There was a significant effect of session number on many of the measures. Given the experimental design, the relatively small number of participants, and the clear indications of hidden subject-level factors influence on AUC, it was not possible to reliably disentangle the effects of these related measures on the AUC independently of (or even adjusting for) time. An additional limitation of this study was that calibration sessions were unsupervised. Researchers allowed participants to complete the calibration without supervision to reduce peer influence on drowsiness but this may have resulted in increased

drowsiness, and one unsupervised subject adjusted the cap producing excessive artifact.

### 5.2. Conclusion

BCI performance with RSVP Keyboard showed a decline over time. Self-reports of increased sleepiness and boredom significantly correlated with decrements in P300 amplitude and AUC. While within-subject BCI RSVP performance was very dependent on P300 amplitude which decreased across time, the lack of intersubject correlation between AUC and P300 amplitude implies there were aspects of the physiological signal other than the P300 amplitude that contributed to classifier performance.

### Abbreviations

BCI	brain–computer interface
AUC	area under the (receiver operating characteristic) curve
ICC	intraclass correlation coefficient
LNS	letter-number sequencing
MPF	median power frequency
RSVP	rapid serial visual presentation
POMS	profile of mood states PSQI, Pittsburgh Sleep Quality Index

### Highlights

- BCI performance decreases over time and parallels increases in boredom and drowsiness.
- P300 amplitude decreases over time and significantly correlates with within-subject performance.
- Decrement in BCI performance over time is not fully explained by P300 amplitude decrease.

### Acknowledgments

Preliminary results were presented at the 2017 annual meeting of the Society for Neuroscience in Washington, DC and 2018 Annual Meeting of the International Congress of Clinical Neurophysiology.

We acknowledge Michelle Kinsella, Betts Peters, and Deniz Erdogmus for their support and theoretical considerations to this experiment. We also acknowledge Sina Dabiri and Dan Klee for assistance with calculations.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This work was supported by the National Institutes of Health [DC009834].

**ORCID**

Barry Oken  <http://orcid.org/0000-0002-3781-4273>  
 Tab Memmott  <http://orcid.org/0000-0001-6143-5057>

**References**

- [1] Wolpaw JR, Wolpaw EW, editors. Brain-computer interfaces: principles and practice. USA: Oxford University Press; 2012.
- [2] Fager S, Beukelman DR, Fried-Oken M, et al. Access interface strategies. *Assistive Technol.* 2011;24(1):25–33.
- [3] Brumberg JS, Pitt KM, Mantie-Kozlowski A, et al. Brain-computer interfaces for augmentative and alternative communication: a tutorial. *Am J Speech Lang Pathol.* 2018 Feb 6;27(1):1–12. PubMed PMID: 29318256; PubMed Central PMCID: PMC5968329.
- [4] Wolpaw JR, Bedlack RS, Reda DJ, et al. Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis. *Neurology.* 2018 Jul 17;91(3):e258–e267. PubMed PMID: 29950436; PubMed Central PMCID: PMC6059033.
- [5] Lees S, Dayan N, Cecotti H, et al. A review of rapid serial visual presentation-based brain-computer interfaces. *J Neural Eng.* 2018 Apr;15(2):021001. PubMed PMID: 29099388.
- [6] Orhan U, Hild KE, Erdogmus D, et al. RSVP keyboard: an EEG based typing interface. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP); 2012. p. 645–648.
- [7] Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol.* 1988;70:510–523.
- [8] Rezeika A, Benda M, Stawicki P, et al. Brain-computer interface spellers: a review. *Brain Sci.* 2018 Mar 30;8(4). PubMed PMID: 29601538; PubMed Central PMCID: PMC5924393.
- [9] Blankertz B, Dornhege G, Krauledat M, et al., editors. The Berlin brain-computer interface presents the novel mental typewriter Hex-o-Spell. Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course; 2006; Graz: Verlag der Technischen Universität Graz.
- [10] Higger M, Quivira F, Akcakaya M, et al. Recursive Bayesian coding for BCIs. *IEEE Trans Neural Syst Rehabil Eng.* 2017 Jun;25(6):704–714. PubMed PMID: 27416602; PubMed Central PMCID: PMC5536189.
- [11] Yin E, Zeyl T, Saab R, et al. A hybrid brain-computer interface based on the fusion of P300 and SSVEP scores. *IEEE Trans Neural Syst Rehabil Eng.* 2015 Jul;23(4):693–701. PubMed PMID: 25706721.
- [12] Yin E, Zeyl T, Saab R, et al. An auditory-tactile visual saccade-independent P300 Brain-computer interface. *Int J Neural Syst.* 2016 Feb;26(1):1650001. PubMed PMID: 26678249.
- [13] Broughton R, Low R, Valley V, et al. Auditory evoked potentials compared to performance measures and EEG in assessing excessive daytime sleepiness in narcolepsy-cataplexy. *Electroencephalogr Clin Neurophysiol.* 1982 Nov;54(5):579–582. PubMed PMID: 6181981.
- [14] Broughton R, Valley V, Aguirre M, et al. Excessive daytime sleepiness and the pathophysiology of narcolepsy-cataplexy: a laboratory perspective. *Sleep.* 1986;9(1 Pt 2):205–215. PubMed PMID: 3704444.
- [15] Koshino Y, Nishio M, Murata T, et al. The influence of light drowsiness on the latency and amplitude of P300. *Clin EEG.* 1993 Jul;24(3):110–113. PubMed PMID: 8403441.
- [16] Pfurtscheller G, Lopes Da Silva FH. EEG event-related desynchronization (ERD) and event-related synchronization (ERS). In: Schomer DL, Lopes Da Silva FH, editors. *Niedermeyer's electroencephalography.* New York (NY): Oxford University Press; 2018. p. 935–948.
- [17] Cohen MR, Maunsell JH. When attention wanders: how uncontrolled fluctuations in attention affect performance. *J Neurosci.* 2011 Nov 2;31(44):15802–15806. PubMed PMID: 22049423; PubMed Central PMCID: PMC3579494.
- [18] Leber AB, Turk-Browne NB, Chun MM. Neural predictors of moment-to-moment fluctuations in cognitive flexibility. *Proc Natl Acad Sci USA.* 2008 Sep 9;105(36):13592–13597. PubMed PMID: 18757744; PubMed Central PMCID: PMC2527350.
- [19] Allaire JC, Marsiske M. Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychol Aging.* 2005 Sep;20(3):390–401. PubMed PMID: 16248699; PubMed Central PMCID: PMC2908894.
- [20] Hill BD, Rohling ML, Boettcher AC, et al. Cognitive intra-individual variability has a positive association with traumatic brain injury severity and suboptimal effort. *Arch Clin Neuropsychol.* 2013 Nov;28(7):640–648. PubMed PMID: 23832096.
- [21] MacDonald SW, Nyberg L, Backman L. Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. *Trends Neurosci.* 2006 Aug;29(8):474–480. PubMed PMID: 16820224.
- [22] Schneider C, Fulda S, Schulz H. Daytime variation in performance and tiredness/sleepiness ratings in patients with insomnia, narcolepsy, sleep apnea and normal controls. *J Sleep Res.* 2004 Dec;13(4):373–383. PubMed PMID: 15560772.
- [23] Oken BS, Salinsky MC, Elsas SM. Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin Neurophysiol.* 2006;117:1885–1901.
- [24] Salinsky MC, Binder LM, Oken BS, et al. Effects of gabapentin and carbamazepine on the EEG and cognition in healthy volunteers. *Epilepsia.* 2002;43(5):482–490.
- [25] Santamaria J, Chiappa K. The EEG of drowsiness in normal adults. *J Clin Neurophysiol.* 1987;4:327–382.
- [26] Iber C, Ancoli-Israel S, Chesson AL, et al. The AASM manual for the scoring of sleep and associated events. In: *Medicine AAoS.* Westchester (IL); 2007.
- [27] Mackworth JF. Performance decrement in vigilance, threshold, and high-speed perceptual motor tasks. *Can J Psychol.* 1964;18:209–223.
- [28] Parasuraman R, Warm JS, See JE. Brain systems of vigilance. In: Parasuraman R, editor. *The attentive*

- brain. Cambridge (MA): The MIT Press; 1998. p. 221–256.
- [29] Eastwood JD, Frischen A, Fenske MJ, et al. The unengaged mind: defining boredom in terms of attention. *Perspect Psychol Sci.* 2012 Sep;7(5):482–495. PubMed PMID: 26168505.
- [30] Millan JD, Rupp R, Muller-Putz GR, et al. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Front Neurosci.* 2010;4. PubMed PMID: 20877434; PubMed Central PMCID: PMC2944670. doi:10.3389/fnins.2010.00161
- [31] Laurent F, Valderrama M, Besserve M, et al. Multimodal information improves the rapid detection of mental fatigue. *Biomed Signal Process Control.* 2013;8:400–408.
- [32] Papadelis C, Chen Z, Kourtidou-Papadeli C, et al. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clin Neurophysiol.* 2007 Sep;118(9):1906–1922. PubMed PMID: 17652020.
- [33] Zhang Z, Luo D, Rasim Y, et al. A vehicle active safety model: vehicle speed control based on driver vigilance detection using wearable EEG and sparse representation. *Sensors.* 2016 Feb 19;16(2):242. PubMed PMID: 26907278; PubMed Central PMCID: PMC4801618.
- [34] Lin CT, Chuang CH, Huang CS, et al. Wireless and wearable EEG system for evaluating driver vigilance. *IEEE Trans Biomed Circuits Syst.* 2014 Apr;8(2):165–176. PubMed PMID: 24860041.
- [35] Li J, Yan C, Xia B, et al., editors. EEG-base brain computer interface for vigilance analysis and estimate. 2nd APSIPA Annual Summit and Conference; 2010 14–17 Dec; Singapore: Bipolis.
- [36] Khan MJ, Hong KS. Passive BCI based on drowsiness detection: an fNIRS study. *Biomed Opt Express.* 2015 Oct 1;6(10):4063–4078. PubMed PMID: 26504654; PubMed Central PMCID: PMC4605063.
- [37] Nguyen T, Ahn S, Jang H, et al. Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Sci Rep.* 2017 Mar 7;7:43933. PubMed PMID: 28266633; PubMed Central PMCID: PMC5339693.
- [38] Chaudhary U, Xia B, Silvoni S, et al. Brain-computer interface-based communication in the completely locked-in state. *PLoS Biol.* 2017 Jan;15(1):e1002593. PubMed PMID: 28141803; PubMed Central PMCID: PMC5283652.
- [39] Csikszentmihaly M. *Flow: the psychology of optimal experience.* New York, (NY): Harper; 1990.
- [40] Lutz A, Slagter HA, Rawlings NB, et al. Mental training enhances attentional stability: neural and behavioral evidence. *J Neurosci.* 2009 Oct 21;29(42):13418–13427. [pii].
- [41] Barbato G, De Padova V, Paolillo AR, et al. Increased spontaneous eye blink rate following prolonged wakefulness. *Physiol Behav.* 2007 Jan 30;90(1):151–154. PubMed PMID: 17070878.
- [42] McNair DM, Lorr M, Droppleman LF. *Manual for the profile of mood states.* San Diego: EdITS/Educational and Industrial Testing Service; 1992.
- [43] Buysse DJ, Reynolds CF, Monk TH, et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatric Res.* 1989;28(2):192–213.
- [44] Wechsler D. *Wechsler Adult Intelligence Scale.* Fourth ed. San Antonio (TX): Pearson; 2008.
- [45] Gillberg M, Kecklund G, Akerstedt T. Relations between performance and subjective ratings of sleepiness during a night awake. *Sleep.* 1994;17(3):236–241.
- [46] Markey A, Chin A, Vanepps EM, et al. Identifying a reliable boredom induction. *Percept Mot Skills.* 2014 Aug;119(1):237–253. PubMed PMID: 25153752.
- [47] Hild KE, Orhan U, Erdogmus D, et al. An ERP-based brain-computer interface for text entry using rapid serial visual presentation and language modeling. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations; Portland (OR): Association for Computational Linguistics; 2011.* p. 38–43.
- [48] Oken BS, Orhan U, Roark B, et al. Brain-computer interface with language model-electroencephalography fusion for locked-in syndrome. *Neurorehabil Neural Repair.* 2014 May;28(4):387–394. PubMed PMID: 24370570; PubMed Central PMCID: PMC3989447.
- [49] Nocedal J, Wright SJ. *Numerical optimization.* 2nd ed. Springer; 2006.
- [50] Speier W, Arnold C, Pouratian N. Evaluating true BCI communication rate through mutual information and language models. PubMed PMID: 24167623; PubMed Central PMCID: PMC3805537 *PLoS One.* 2013;8(10):e78432.
- [51] Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: part 2—correlation between subjects. *BMJ.* 1995 Mar 11;310(6980):633. PubMed PMID: 7703752; PubMed Central PMCID: PMC2549010.
- [52] Allison PD. *Fixed effects regression models.* Thousand Oaks (CA): Sage; 2009.
- [53] Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal data analysis.* 2nd ed. Hoboken (NJ): John Wiley & Sons, Inc; 2011. p. 627–653.
- [54] Kline RB. *Principles and practice of structure equation modeling.* 4th ed. New York (NY): Guilford Press; 2015.
- [55] Luu P, Tucker DM, Makeig S. Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clin Neurophysiol.* 2004 Aug;115(8):1821–1835. PubMed PMID: 15261861.
- [56] Yordanova J, Kolev V, Polich J. P300 and alpha event-related desynchronization (ERD). *Psychophysiology.* 2001 Jan;38(1):143–152. PubMed PMID: 11321615.
- [57] Varela F, Lachaux J-P, Rodriguez E, et al. The brainweb: phase synchronization and large scale integration. *Nat Rev Neurosci.* 2001;2:229–239.
- [58] Bodala IP, Li J, Thakor NV, et al. EEG and eye tracking demonstrate vigilance enhancement with challenge integration. *Front Hum Neurosci.* 2016;10:273. PubMed PMID: 27375464; PubMed Central PMCID: PMC4894919.
- [59] Makeig S, Inlow M. Lapses in alertness: coherence of fluctuations in performance and EEG spectrum. *Electroencephalogr Clin Neurophysiol.* 1993 Jan;86(1):23–35. PubMed PMID: 7678388.