Medical Engineering and Physics 68 (2019) 25–34

Contents lists available at ScienceDirect

Medical Engineering and Physics

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

Two-channel in-ear EEG system for detection of visuomotor tracking state: A preliminary study

Eunice Kuatsjah, Xin Zhang, Mahta Khoshnam, Carlo Menon*

Menrva Research Group, Schools of Mechatronic Systems Engineering and Engineering Science, Simon Fraser University, Metro Vancouver, British Columbia, Canada

ARTICLE INFO

Article history:
Received 26 October 2018
Revised 21 March 2019
Accepted 31 March 2019

Keywords:
Visuomotor tracking
In-ear EEG system
Mental workload

ABSTRACT

Evaluating an operator's mental workload during work activities is crucial to maintain safety and performance. By minimizing human error associated with work demands, especially in a hazardous environment, potentially serious errors may be avoided. This study aims to assess the feasibility of using an in-ear EEG system to classify the user's state in a visuomotor tracking task that may influence mental workload and motor action. A two-channel wireless in-ear EEG system was used to record EEG signals while subjects performed the task using a joystick to manipulate an object displayed on a monitor. A highly comparative time series analysis was employed on the processed signals to extract and select the top features for each subject individually. The features sets were trained and tested with support vector machines, random forest, linear discriminant analysis, subspace discriminant, and neural network to compare their performances. Models trained on two trials, each 14 minutes in duration and tested on the other trial were able to yield an accuracy of 79.30 ± 4.85% on average across the ten participants with an individualized moving average threshold filter and classifier. This proof-of-concept study demonstrates the feasibility of using a two-channel wireless in-ear EEG system as a viable solutions to develop wearable devices to detect mental workload associated with the execution of visuomotor tasks.

1. Introduction

The ability to assess an operator's mental workload continuously and autonomously whilst undertaking tasks could lower the incidence of industrial accidents and injuries [1], improve the usability of human-computer interfaces [2], and aid in the design of appropriate adaptive automation systems [3–7]. One of many ways this can potentially be achieved is through the integration of shared autonomy techniques, in which a closed-loop dynamics are formed between the user and the brain-actuated device in such a way that the task can be performed as easily and safely as possible [8]. For example, such a system may aid in monitoring the user's cognitive load while operating a heavy equipment, to provide alerts and feedback when high mental workload is detected.

Operator performance and workload have been previously investigated in several studies involving driving tasks [9,10], mental arithmetic tasks [11], n-back tasks, as well as during multitasking [12]. For example, the Multi-Attribute Task Battery (MATB-II) provides a set of tasks analogous to activities that aircraft crew members perform in flight, like tracking, monitoring, resource management, and communications [13]. The need for increased attentional effort and alertness for visuomotor control has been shown to induce a change in brain activity, especially in the right parietal lobe [14].

Several measures of mental workload have been studied in the past, including various behavioral, subjective, and physiological measures [15,16]. Physiological measures appeared to be more appropriate for practical applications, since they can provide continuous data over time and do not affect task performance [17]. Electroencephalography (EEG) provides excellent time resolution and can track changes as different behaviors are performed. Moreover, EEG can be collected using portable platforms, and thus can be used outside the laboratory environment, unlike other neuroimaging modalities such as functional magnetic resonance imaging [18]. Previous studies have found correlations between EEG and mental workload in visuomotor tasks [19–21].

The limited portability and long setup time of traditional EEG systems have hindered the integration of such brain monitoring platforms into daily lives [22]. Consequently, a more discreet, unobtrusive, and user-friendly approach has been developed in which EEG is recorded through an in-ear electrode (ear-EEG) with a device worn like an earbud [23]. The ear-EEG systems yielded similar performance to that of conventional on-scalp EEG systems, and reflected the same cortical activity as nearby temporal region scalp

* Corresponding author.
E-mail address: cmenon@sfu.ca (C. Menon).

https://doi.org/10.1016/j.medengphy.2019.03.016
1350-4533/

electrodes [24]. Several arrangements of the in-ear/ear EEG system have been used in past studies, such as two in-canal electrodes placed diametrically opposed to each other on a viscoelastic substrate [23], and multiple in-canal and concha electrodes on a custom-made earpiece made using the user's ear imprints [24]. In this study, one in-canal electrode on a viscoelastic material attached to a generic 3D printed enclosure was used for each ear.

The visuomotor tracking task used in this study involved visual object recognition, as well as working memory which refers to short-term maintenance and manipulation of items [25]. Visual recognition has been typically viewed as a bottom-up hierarchy process that deals with information sequentially with increasing complexity: lower-level cortical processors such as the visual cortex are at the bottom and higher-level cortical processors such as the inferotemporal cortex (IT) are at the top, where recognition is facilitated [26]. Within the inferior temporal cortex (ITC), several regions work together to process and recognize objects. One such example is the connectivity between the ITC and the parahippocampal place area, which helps differentiate between scenes and objects, while the hippocampus remembers the appearance of an object and compares it to others for future use [27,28]. A study investigating the neural mechanisms through which the medial temporal lobe (MTL), pre-frontal cortex (PFC), and ITC interact during working memory maintenance found that with increasing load, EEG phase synchronization increased between the IT cortex and anterior parahippocampal gyrus, as well as within the MTL [29]. The MTL has also been found to be an important junction in the planning and execution of motor actions, whether they were internally or visually driven [30].

Rolandic beta rhythms (13–30 Hz) are observed as spontaneous activity in healthy subjects over the sensory-motor strip and are modulated during various motor and cognitive tasks. Movement is associated with neuronal activation in the somatosensory area and rolandic beta rhythms were speculated to increase when the sensory-motor strip neuronal system is relaxing after an activation phase [31]. The higher frequency band gamma rhythm (30–80 Hz) is modulated by sensory input and internal processes such as working memory and attention [32]. Gamma power in the somatosensory cortex increases with sensory drive [33] and is related to visual attention during a cognitive task [34]; while in the higher cortex, gamma power is elevated during working memory [35] and learning [36].

The beta band (13–30 Hz) has been widely used for the measurement of mental workload when indicating engagement level during sustained attention [37], alertness [38], sensory processing in active wakefulness [39], and increasing memory load [40]. In addition, the beta band has also been associated with motor imagery, movement preparation, and movement [41,42] which relates to the visuomotor tracking task performed. Further, gamma frequencies >30 Hz are also closely associated with sensory processing [43]. Oscillation of neural activity within the gamma band (35–45 Hz) has been observed after visual stimuli or just before a movement task, and is associated with the binding of sensory information for sensorimotor integration [44]. A previous study investigating salient features to measure mental workload found that among the five common EEG bands (delta, theta, alpha, beta, gamma) and physiological signals (heart rate, blinks, and breath intervals), gamma band features in the 31–40-Hz range appeared to be the most salient, followed by the continuous eye-blink feature [45]. The beta band features in the 13–30-Hz range along with heart rate features have also been found to be consistently salient [45].

This study focused on assessing the feasibility of utilizing an unobtrusive two-channel in-ear wireless EEG system to classify the subject's state when performing a visuomotor tracking task using a joystick to manipulate an object displayed on a monitor. The in-ear EEG system was first assessed by performing an alpha attenuation paradigm test on one subject. Then, several classifier models were explored for each subject's data by individually training them on two trials and testing them on the other trial. Finally, the differences in the classifiers' accuracies across all participants were evaluated using a cross-validation testing procedure to observe the effects of different models on the testing accuracy.

Table 1
Demographic data of the participants.

Participants	Gender	Age
A00	F	25
A01	M	31
A02	F	41
A03	M	28
A04	M	54
A05	M	32
A06	F	25
A07	F	23
A08	M	23
A09	F	54

2. Methods

2.1. Participants

Ten healthy participants, five females and five males, mean age 33.6 ± 12.0 years participated in this study. All of them were right-handed. Their demographic data are presented in Table 1.

2.2. In-ear EEG system setup

Twenty individual in-ear EEG earpieces were fabricated for the ten participants for use in each ear. The earpiece consisted of a 3D printed enclosure, memory foam earbud (Comply™ T-200, Hearing Components, Inc., USA), conductive silver fabric electrode (resistance of less than 1 ohm per foot), and a stainless-steel conductive thread. A 4 mm x 10 mm silver fabric was cut and sewn onto the memory foam with a conductive thread into which a standard wire was soldered. The end of the wire was then connected to a 1.5 mm touch-proof connector (Cadwell®, Cadwell Industries, Inc., USA) that connected to a BioRadio™ Wireless Physiology Monitor (Great Lakes NeuroTechnologies, Cleveland, OH, USA). The in-ear EEG electrode was located on the upper side of the ear canal when donned, as illustrated in Fig. 1. The total impedance measured from the twenty in-ear electrodes to the connector was $2.65 \pm 0.82\ \Omega$.

Two channels were placed in the left as well as the right ear using the earpieces. The reference electrode was placed on the left earlobe using a 10 mm diameter gold-plated cup ear clip electrode (Technomed Europe, NL), and the ground electrode was placed on the forehead using a 1×1 inch pre-gelled cloth electrode (MVAP Medical Supplies, Inc., CA, USA) connected to a snap connector (Advantage Medical electronics, LLC, FL, USA). To improve the electrical contact between the electrodes and the skin, a saline solution was applied on the in-ear electrodes and conductive paste (Ten20® Weaver and Company, Colorado, USA) was applied on the earlobe reference electrode. EEG data were recorded with a sampling rate of 250 Hz.

2.3. Data acquisition

All ten participants were seated in front of a monitor wearing the in-ear EEG system while performing a computer-based visuomotor tracking task as shown in Fig. 2. Each participant performed three trials of a 14-minutes task consisting of eight alternating 'resting' and 'tracking' states, each with a duration of

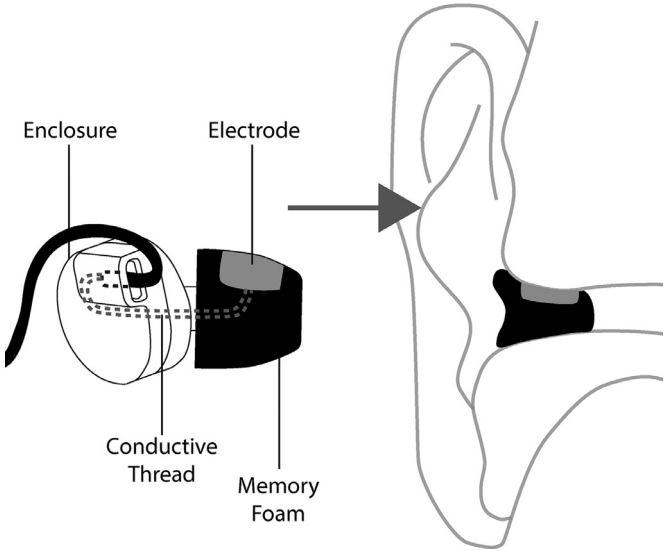


Fig. 1. In-ear EEG earpiece. The earpiece consisted of a 3D printed enclosure, memory foam earbud, conductive silver fabric electrode and stainless-steel conductive thread that connects to an EEG amplifier. The electrode was positioned on the upper side of the ear canal when the earpiece was worn.

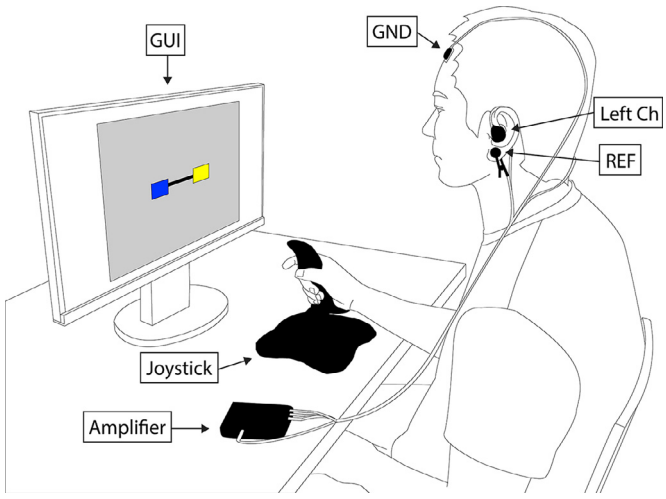


Fig. 2. Experiment setup. Participant performing a computer-based visuomotor tracking task using a joystick to control the yellow box on the right side with the goal to maintain the connecting black bar as horizontal as possible at all times, while wearing an in-ear EEG system with electrodes in both ears (left channel shown), ground electrode (GND) on the forehead and reference electrode (REF) on the left earlobe.

105 s. During the 'resting' states, the participant was asked to rest while the monitor screen showed a crosshair with a black background. During the 'tracking' states, the monitor screen showed a graphical user interface (GUI) in which the participant used a joystick (Extreme 3D Pro Joystick, Logitech, Switzerland) to vertically control a yellow box on the right side connected to a blue box on the left side. Both boxes moved only in the vertical direction and were connected to each other with a black bar. This interface partly resembled that of a previous study investigating the effects of Galvanic Vestibular Stimulation (GVS) on manual tracking behavior in Parkinson's disease [46]. The blue box oscillated vertically and followed a pre-generated trajectory. The participant was asked to imitate the trajectory of the blue box and keep the connecting black bar horizontal at all times.

In each trial, the tracking task required following trajectories which were produced by generating an amplitude modulated signal with a carrier frequency of 0.01 Hz, a modulation frequency of 0.03 Hz, and a modulation index of 0.4/0.95. The differences between the trajectories in the alternated tracking blocks were a phase shift and an inverted signal, this minor variation was introduced so that the participants did not expect the same trajectories in all the trials while also keeping a consistent level of difficulty between the trials. The overall task performed is illustrated in Fig. 3. The trajectories were scaled to have a peak-to-peak magnitude of 576 pixels, displayed at a resolution of 1920 × 1080 pixels on the 27″ monitor. The tracking task involved fine motor movements mainly generated from the wrist to control the displayed object on the screen using the joystick. Participants were instructed to not produce unnecessary body movements and to sit still throughout the test.

2.4. In-ear EEG processing and validation

EEG signals were filtered at a 12–50 Hz frequency band. This range was chosen as it was expected to indicate a change in brain activity due to the activation of the sensorimotor strip area during the fine wrist movements, as well as changes in mental workload as a result of somatosensory processing. The in-ear EEG was also situated close to the temporal lobe, which is associated with visual object recognition and working memory.

The reliability of the signals obtained through this in-ear EEG system was first assessed through an alpha attenuation paradigm. A previous study which characterized EEG signals recorded through the ear-EEG method found that ear-EEG signals have lower, but perfectly acceptable discriminatory power compared to those of a conventional on-scalp EEG setup [24]. In this study, the setup was first assessed by performing an Alpha Attenuation Response (AAR) test over 100 s, in which a subject opened and closed their eyes, holding each state for 10 s. During the transition, a short auditory stimulus was generated to instruct the subject to open or close their eyes. The signal was filtered using a 2–30 Hz bandpass filter and a short-time Fourier transform (STFT) was performed using a 1-second Hamming window with 50% overlap. The number of discrete Fourier transform points was set to 256, obtained from $\max(256,2^p)$, where $p=\lceil log_2\ N_w \rceil$ and N_w is the length of the window. An increase in alpha band power was expected during the closed eyes period.

2.5. Feature extraction and selection

As explained in the previous section, participants were asked to perform continuous 'resting' and 'tracking' states alternatingly, each lasting for 105-second as illustrated in Fig. 3. The middle 100-s of each block was used to exclude the between-state transition period. A 12–50 Hz bandpass FIR filter was applied to a moving 7.5-s window with $^1/_3$ overlap, and the last 5-s time series signal was used to extract the features, hence there was no overlap between each successive 5-s EEG segment. The length of the window chosen was obtained experimentally and was expected to show a stable representation of the state. This pre-processing stage resulted in each alternating state (block) comprising nineteen 5-s windows. Therefore, with three trials of alternating states, each subject produced a total of 3 trials × 8 blocks × 19 windows = 456 windows of time series signals for each channel. Each window was normalized by subtracting its mean and dividing by its standard deviation.

The non-stationarity and inherent variability of EEG signals are among the challenges in developing a reliable BCI system [47]. Continuous changes over time between or within the recording, as well as mental/emotional state in different sessions and inherent differences between subjects, could contribute to signal variability [48]. One approach used in this study was to use

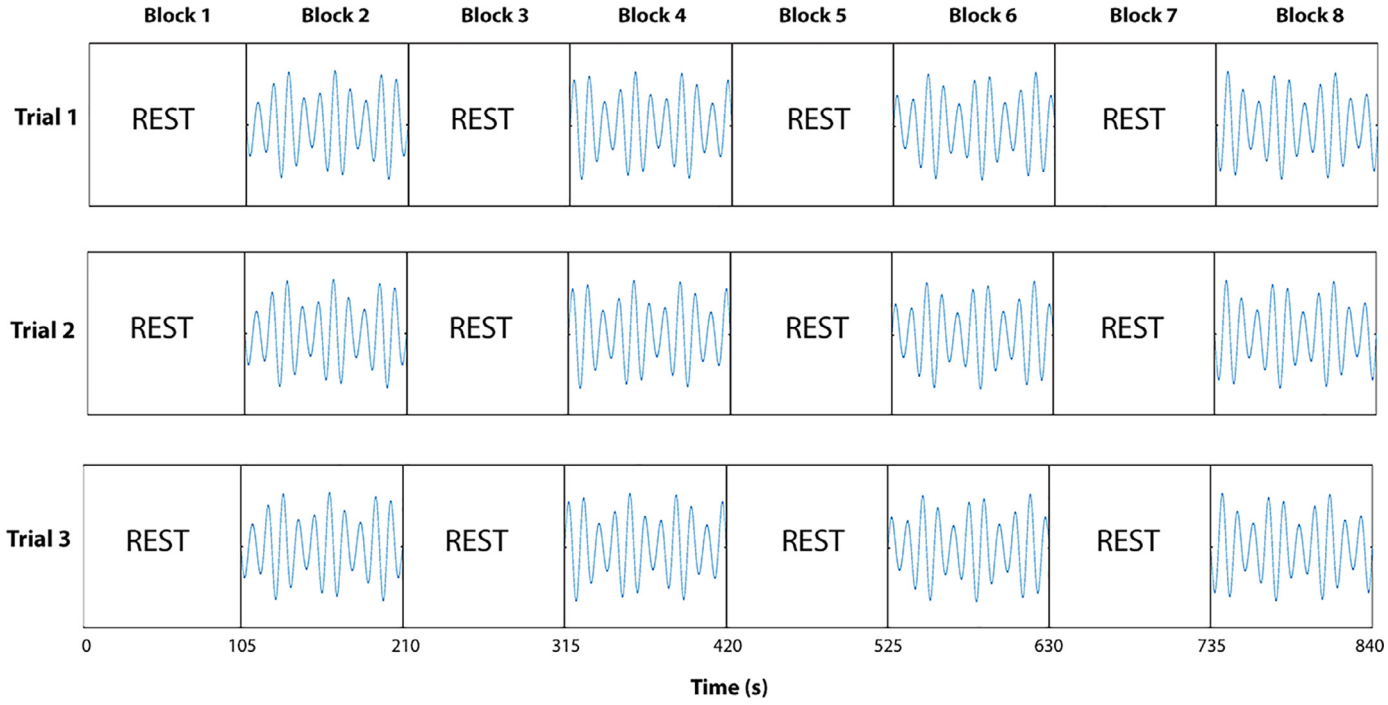


Fig. 3. Alternated task. Participants performed three trials of alternated resting and tracking states. Each trial duration was 14 min, consisting of eight 105-s blocks performed continuously.

subject-dependent measures and employ the Highly Comparative Time-Series Analysis (HCTSA) [49] software package within MATLAB to extract the filtered and normalized signals for each individual. HCTSA is a computational framework that uses ~7000 time series features to automatically select useful features for a given application [50]. The features were derived from the vast scientific time-series analysis literature, which includes summaries of time series in terms of their correlation structure, distribution, entropy, stationarity, scaling properties, and fits to a range of time-series models [51]. The computed features were normalized using a scaled quantile-based sigmoidal transform and ten clusters were generated from the top forty features that best separated the labeled time series using a linear classifier. The feature with the highest accuracy in each cluster was selected, yielding 20 features for both the left and right channel for each subject. This approach was applied to each subject's dataset individually.

2.6. Model training and testing

After selecting the features, each subject's data were trained individually with six classifiers: SVM with linear kernel, SVM with Gaussian kernel, random forest, LDA, subspace discriminant, and neural network (NN) using MATLAB. For both linear and Gaussian kernel SVM, the box constraint levels were set to 1 and the kernel scale was set to 'auto' to automatically compute the value using a subsampling procedure. For both ensemble learning classifiers, i.e.: subspace discriminant and random forest, the number of learning cycles was set to 64 and the number of predictors to sample was set to 4, obtained from $\sqrt{pred}$where *pred* is the number of features [52]. The random forest classifier was set to grow deep trees with a maximum number of splits set to the number of observations – 1. The NN was configured to have one hidden layer of 10 neurons with a scaled conjugate gradient backpropagation training algorithm and the performance was calculated using a cross entropy function. The training stopped if the network reached any of the following: validation error increase after 6 iterations, 1000 epochs, 0 performance goal, or $1e^{-6}$ performance gradient.

The primary measure to evaluate the performance of the system is the test accuracy obtained in an across-trial procedure, in which the classifier models were generated by training on two trials (2 trials × 8 blocks × 19 windows = 304 windows) and were tested on the third trial of the dataset (1 trial × 8 blocks × 19 windows = 152 windows). Each window was non-overlapping and represented one observation. This across-trial testing was performed three times for each training and test trial set combination, as illustrated in Fig. 4(a). A moving average was calculated over a sliding window of five data points centered around the current element position to smooth out noisy prediction output. The first and last two data points after averaging were excluded resulting in 148 windows. A moving average value less than or equal to a certain threshold value was classified as a resting state, while a moving average value greater than the threshold value was classified as a tracking state. In order to obtain the final output of the averaged predicted result and maximize accuracy, a grid search of threshold values with steps of 0.2 was performed, where 0 represented the resting state and 1 represented the tracking state.

Since the optimal test accuracies were achieved by different classifiers in the across-trial testing procedure, a cross-validation was performed to observe if any of the classifiers performed better in general compared to the rest. Each subject's entire three trials dataset (456 windows) were permutated and divided into 5-fold for training and testing to obtain the average across-fold accuracy, as illustrated in Fig. 4(b). No moving average filter was used in this procedure, since the observations were permutated. The process was repeated 10 times to obtain a grand average to represent the validation accuracy of each model. Here, 80% of the dataset was used for training compared to 67% used in the across-trial testing.

2.7. Statistical analysis

Statistical analyses were carried out using IBM SPSS Statistics to observe whether there was a significant difference in the achieved test accuracies for the across-trial procedure for different test trial sets. A significant difference may imply that the data are time sensitive and produced a substantial change during the protocol. The

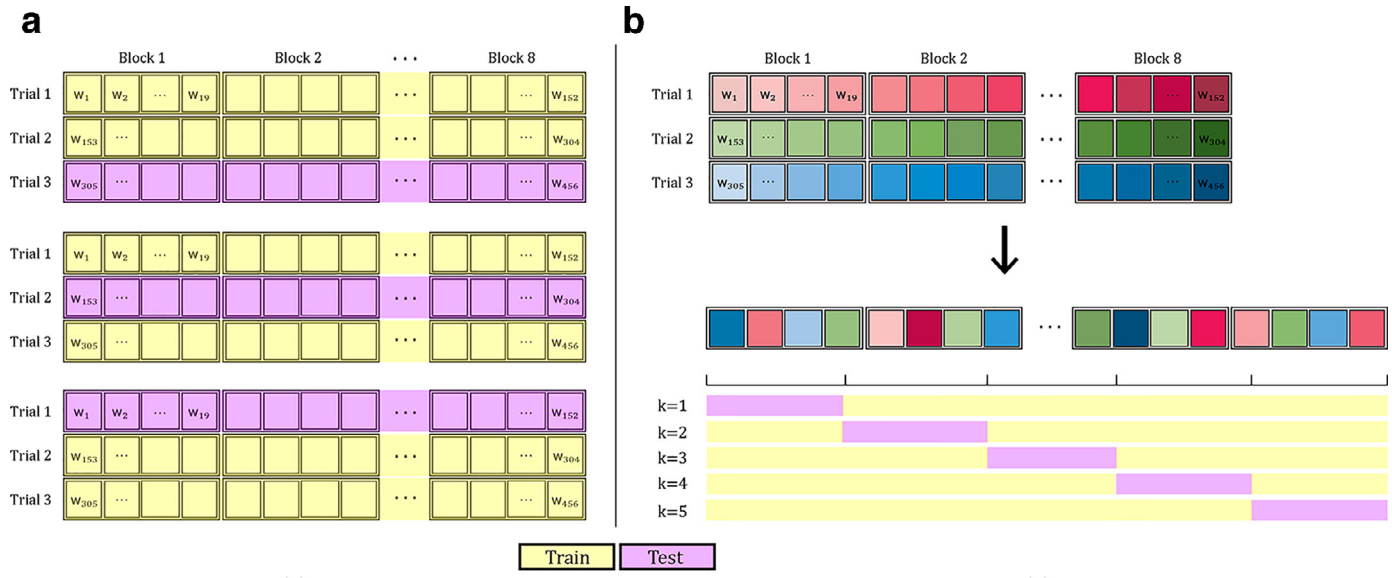


Fig. 4. Data splitting. (a) Across-trial testing: train on two trials and test on the other trial in all combinations. (b) Cross-validation testing: all windows from three trials were combined, permutated, and divided into 5-fold.

highest test accuracies obtained through one of the six classifiers for each subject were compared on each test trial to study the effects of time on the models' performances, i.e. test accuracy of test trials 1, 2, and 3. The average validation accuracy results from the cross-validation procedure were also compared to see if there was a significant difference in the accuracies produced by the different classifiers.

A repeated measures one-way ANOVA was used if the dataset met the following assumptions: no significant outliers, approximately normal distributions, and equal variance in differences between all possible pairs of within-subject conditions. Boxplots were used to detect the outliers in the group, the normality assumption was checked by performing a Shapiro-Wilk test as well as by visually observing the histogram and Q-Q plot of the data. The sphericity assumption was checked by performing Mauchly's sphericity test. Although the repeated measures ANOVA is quite robust to some violations of the assumptions, Friedman's non-parametric test was used instead if the assumptions were immensely violated. If a significant difference in the non-parametric test was obtained, a Wilcoxon signed rank test was used to observe the pairwise difference between the group with the highest mean rank compared to the rest.

3. Results

3.1. In-ear EEG alpha attenuation response (AAR)

As an example, this setup was tested on one of the subjects, and the time-frequency plots for the left and right in-ear EEG electrode are shown in Fig. 5(a) and (b). For every 10 s when the subject's eyes were closed, an increase of power in the alpha band (8–13 Hz) for both channels was observed. The power spectral densities (PSD) on the alpha band (8–13 Hz) during eyes opened and closed were compared by averaging the power over the middle 9 s of each state using the STFT. The measured alpha band powers on the left channel were 6.283×10^{-13} and 1.994×10^{-13}, and 5.272×10^{-12} and 2.129×10^{-12} on the right channel during eyes closed and opened, respectively as shown in Fig. 5(c) and (d). In this example, the Alpha Attenuation Coefficient (AAC), calculated as a ratio of the mean alpha power during eyes-closed to eyes-open was 3.129 and 2.476 on the left and right channels, respectively. This shows that the in-ear EEG system could distinguish the change in the alpha wave in the AAR paradigm.

3.2. Feature selection

Ten features for each left and right channels for each subject were obtained from the automated feature selection algorithm employing ~7000 feature operations using HCTSA [50]. Therefore, a total of 100 features for all ten subjects for each left and right channels were selected. These selected optimum features varied across subjects and were all unique for the left channel. On the right channel, two identical operations were used twice, i.e. standard deviation of standardized mean of FFT power spectrum split into four normalized bands and root mean square error of model prediction using a 6th order autoregressive model for one step ahead prediction [49,50]. A brief description showing an example of selected features for the left channel of subject A00 is provided in the supplementary material (Table S1). Details of the parameters algorithm can be found in the HCTSA documentation [49,50]. Consistent with HCTSA terminology [49,50], the exhaustive feature operations computed could be loosely categorized into several types (e.g. model fitting (MF), correlation (CO), stationarity (SY), distribution (DN)) that employ common algorithms but may use different input parameters and produce multiple outputs. Of the 200 features extracted from both channels across all ten participants, features obtained from correlation and model fitting parameters showed the highest occurrence in best separating the two class conditions (resting vs tracking state) as illustrated in Fig. 6.

3.3. Model testing

3.3.1. Across-trial

The data for each subject were trained on two trials (304 windows) and tested on the other trial (152 windows) in all combinations with all six classifiers. Fig. 7 illustrates the results of the classification for subject A00 trained on the first and second trials, and tested on the third trial, using a NN classifier with a threshold of 0.6. This procedure was applied to all subjects individually. The results from training on the first two trials and testing on the last trial with classifiers and thresholds that yielded the highest test accuracies are listed in Table 2.

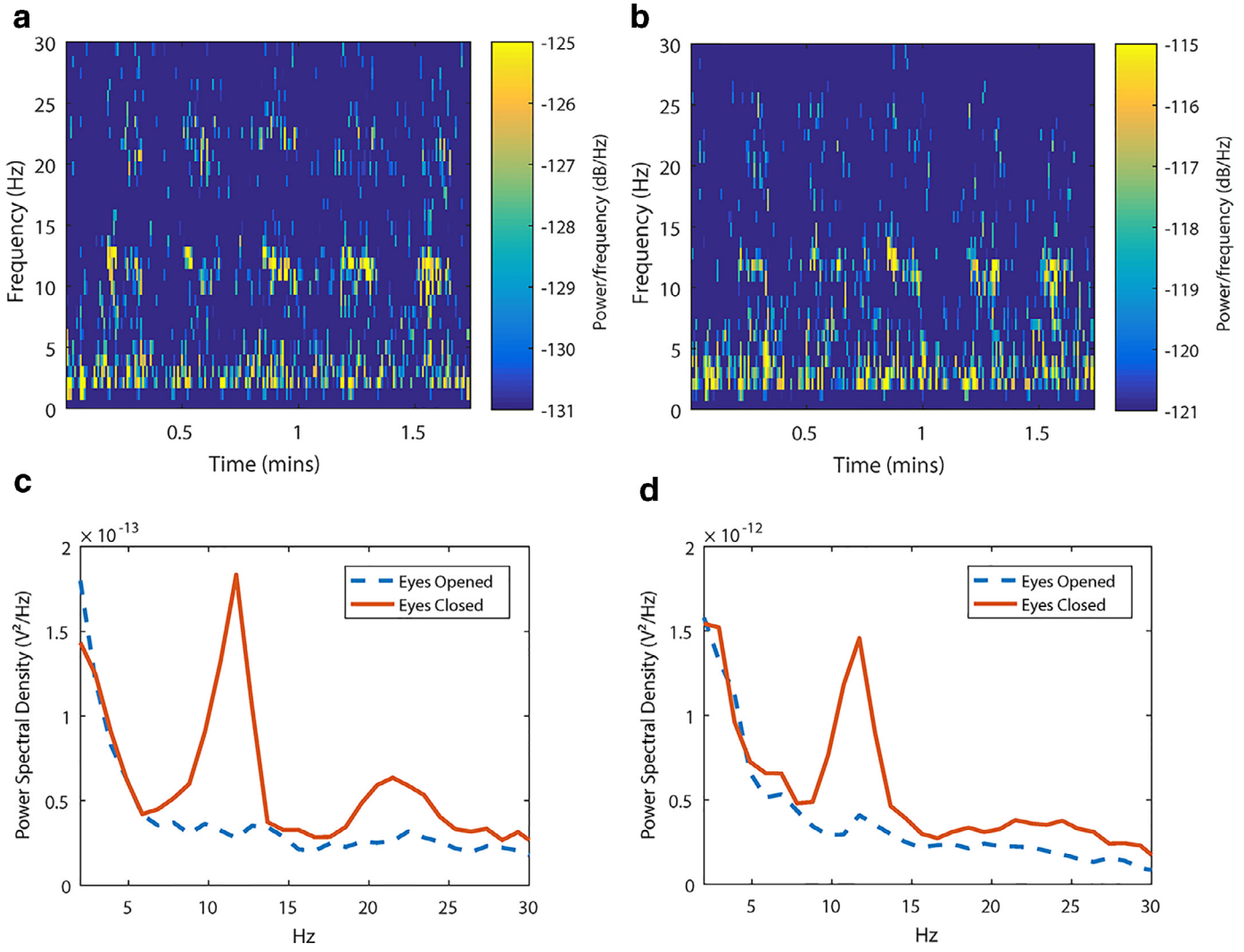
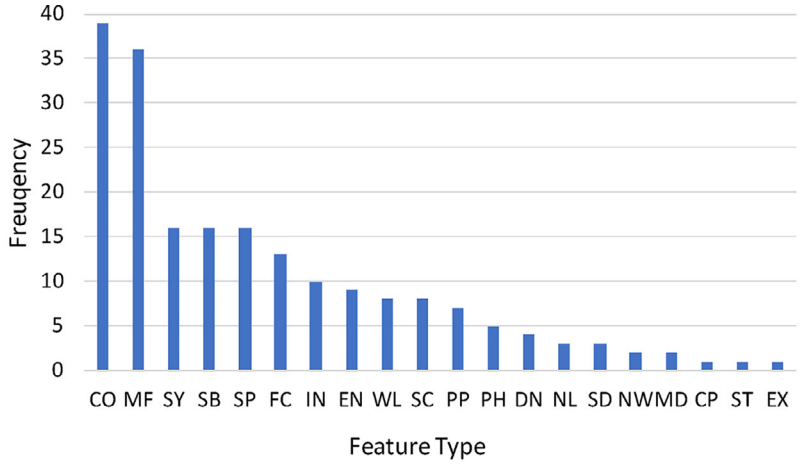


Fig. 5. AAR of in-ear EEG setup. Top panel: time-frequency plot of (a) left channel and (b) right channel of in-ear EEG signals showing an increase in alpha band (8–13 Hz) power in the 10 s interval during eyes closed compared to eyes opened. Bottom panel: power spectral density plot of (c) left channel (d) right channel showing a peak in the alpha band during the eyes closed condition.



CO:Correlation MF:Model Fitting SY:Stationarity SB:Symbolic Transformations SP:Power Spectrum FC:Forecasting
IN:Automutual Information EN:Entropy WL:Wavelet SC:Fluctuation Analysis PP:Pre-processing PH:Walker Simulation
DN:Distribution NL:Nonlinear Analysis SD:Surrogate Analysis NW:Visibility Graph MD:Heart Rate Variability Derived Analysis
CP:Step Detection ST:Statistics and Trend EX:Extreme Events Movng Threshold Model

Fig. 6. Types of selected features for the left and right channels across all participants. Ten features were selected for each channel for each participant, resulting in 200 features from both channels across ten participants. Features were broadly categorized based on HCTSA terminology [49,50].

Table 2
Across trial results of all subjects' data trained on trial 1 and 2, tested on trial 3.

Subject	Classifier	Train Acc.	Test Acc. Before Thres.	Mov. Avg. Thres.	Test Acc. After Thres.
A00	Neural network	78.95	78.29	0.4	89.86
A01	Subspace discriminant	66.78	68.42	0.4	78.38
A02	SVM Gaussian	91.78	64.47	0.6	72.30
A03	SVM linear	68.75	62.50	0.4	72.30
A04	LDA	75.00	77.63	0.4	87.84
A05	Subspace discriminant	70.07	72.37	0.4	84.46
A06	SVM Gaussian	92.76	66.45	0.4	75.68
A07	LDA	73.36	65.79	0.4	79.05
A08	Neural network	72.70	65.79	0.4	74.32
A09	Neural network	69.74	63.82	0.6	70.95
Mean			68.55		**78.51**
Standard deviation (SD)			5.65		**6.76**

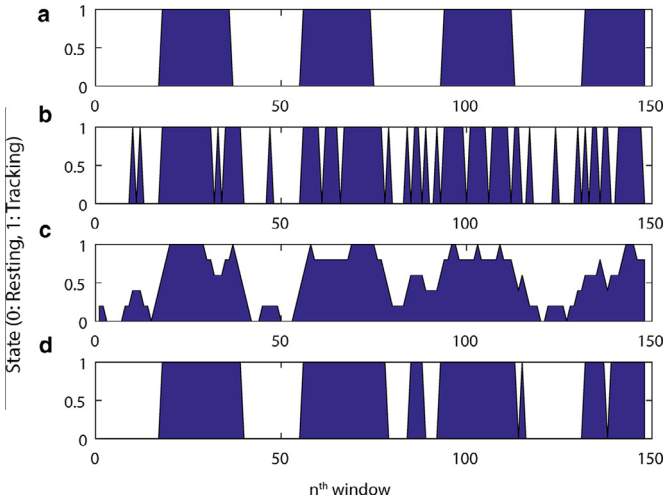


Fig. 7. Across-trial testing predicted output with moving average thresholding of Subject A00 with neural network classifier. (a) test label (0: resting, 1: tracking), (b) classifier's predicted value, (c) moving average of sliding window for 5 data points, (d) smoothed predicted output after thresholding (threshold = 0.6).

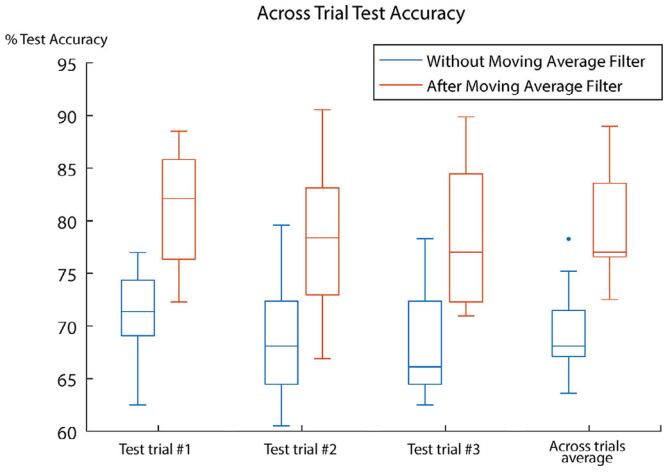


Fig. 8. Average across trial test accuracies for all subjects on: each test trial set and the average of all trials. Test accuracies were obtained from the highest performing classifiers trained on two trials and tested on the other one trial, a moving average filter was applied to smooth out the prediction output.

Fig. 8 summarizes the across-trial test results of each test set and on average across all subjects trained individually on two trials using the classifier and threshold value that yielded the highest accuracy, without (left) and with (right) the moving average filter for each group. Taking the average of all three trial set combinations individually produced an average test accuracy of 79.30 ± 4.85% after filtering. Testing on the first trial produced a slightly higher test accuracy at 80.74 ± 5.68% compared to the other two trials at 78.51 ± 6.76% and 78.65 ± 7.04%. The optimal test accuracies were achieved by different classifiers, and a cross-validation procedure was performed repeatedly with permutated data and more training samples to observe if any of the classifiers performed better in general.

3.3.2. Cross-validation

The performance of each classifier was compared using a 5-fold cross-validation trained and tested individually per subject and repeated over 10 iterations. Fig. 9 shows the average performance of the classifiers across all subjects. Overall, NN produced the highest mean accuracy across folds at 71.50±4.79% and median at 69.40% [interquartile range (IQR) = 69.07–72.36%]. The lowest mean accuracy was produced by random forest at 67.33 ± 4.29% and median at 65.82% [IQR = 64.58–67.94%].

3.4. Statistical analysis

In the across-trial procedure, test accuracies after moving average filtering in the three trial test sets were compared using a three-factor repeated measures one-way ANOVA. There were no outliers in either group, the residuals' normality assumptions were met through Shapiro-Wilk tests, and the sphericity assumption was met through Mauchly's sphericity test. No significant difference was obtained through testing on different trials ($F_{2,\ 18} = 0.548$, $p = 0.587$), this may suggest that the data were quite stable within the ~50 min recording sessions.

In the cross-validation procedure, average validation accuracies for each classifier were the dependent variables and boxplots showed 1 or 2 outliers in the random forest, subspace discriminant, and NN groups. Further, Shapiro-Wilk's tests on each group's residuals rejects the null hypothesis of the SVM linear, subspace discriminant, and NN groups having a normal distribution ($p < 0.05$). Based on these results, it was determined that the nonparametric test may be more appropriate, hence Friedman's test was used instead of the repeated measures one-way ANOVA. There was a statistically significant difference in the validation accuracy obtained with different classifiers ($\chi^2(5) = 23.886$, $p < 0.0005$), the NN having the highest mean rank at 6.00 compared to the SVM Linear, SVM Gaussian, random forest, LDA and subspace discriminant at 3.60, 300, 2.60, 2.50, and 3.30, respectively. All paired difference groups between NN and each of the remaining classifiers showed normal distributions based on visual inspection of the histogram and had a Shapiro-Wilk's test p-value > 0.05. Therefore, the Wilcoxon signed rank test was deemed appropriate. A Bonferroni correction was applied on the five comparisons performed, therefore adjusting the significance level to 0.01. All five comparisons

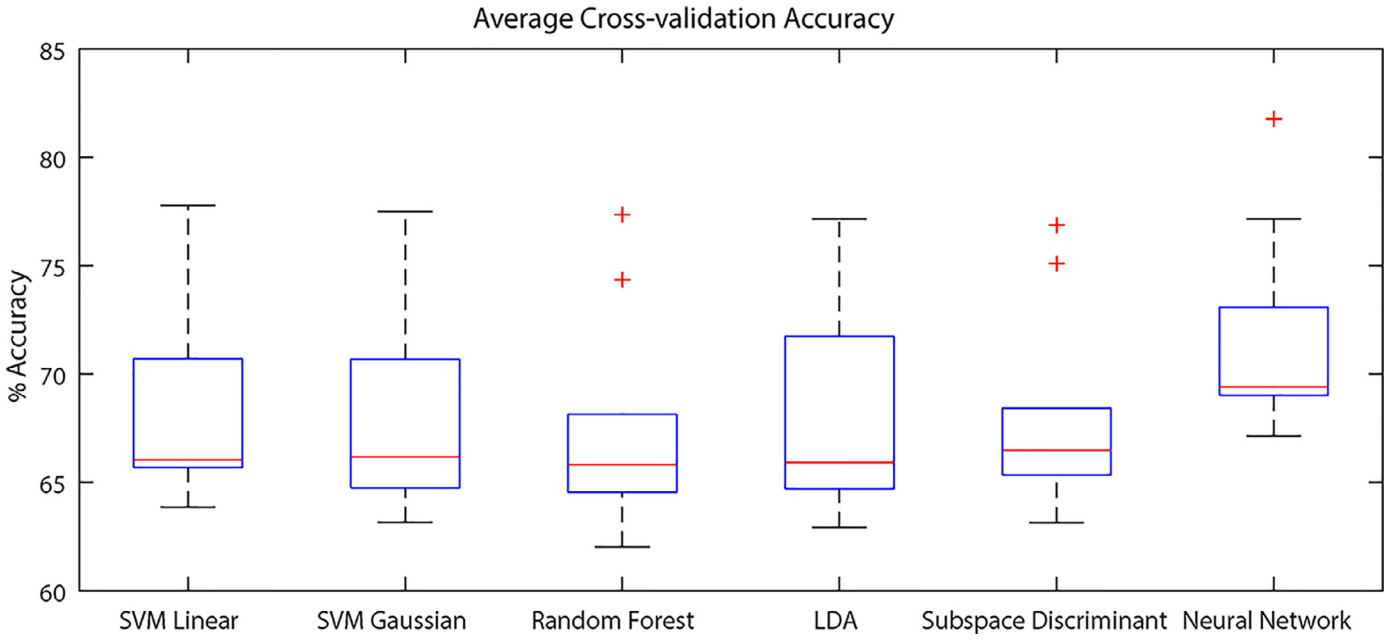


Fig. 9. Average 5-fold cross-validation accuracy over 10 iterations for all subjects. The performance of SVM linear, SVM Gaussian, random forest, LDA, subspace discriminant, and neural network classifiers were compared using a 5-fold cross validation over 10 iterations of training and testing within subject.

produced p-value < 0.005, rejecting the null hypothesis that the median differences between them were zero.

4. Discussions

In ear-EEG previously showed lower amplitudes compared to on-scalp EEG, but exhibited similar signal-to-noise ratio with diminished electro-oculogram and motion artifacts, as there are no muscle fibers in the ear canal and the face and eye muscles are located far away [22]. The tight fit of the viscoelastic foam earbud has also been shown to have an excellent resistance to motion artifacts, especially from the pulsatile ear canal wall movement [23]. In this preliminary study, a blink artifact removal algorithm was not applied during the data processing, since only two channels were used and blink artifacts are more prominent in the forehead area. The diminished blink artifacts, on the other hand, might also influence the distinction between the two states, since mental workload has been associated with blink duration, rate, and latency [53–55].

NN outperformed the other classifiers for all subjects in the cross-validation, but this was not always the case in the across-trial testing. This might be because of the larger training size (80%) used in the cross-validation, which is more appropriate for NN. On the other hand, in the across-trial method, only 67% of the data were used for training. Although the results of NN are not consistently stable, having a larger training size in the across-trial method is expected to boost the accuracies to reach what was obtained in the cross-validation method. Random forest produced the lowest mean accuracy among all, which might be due to overfitting as the classifier was configured as a deep tree with a maximum number of splits set to the maximum number of observations – 1. This showed in the high training accuracy achieved, however based on visual observation of the boxplots as shown in Fig. 9, random forest showed comparable accuracy to the rest of the other non-NN classifiers. Regardless, this study roughly compared different classifiers performance in such a task. Future work may focus on fine-tuning the parameters of a certain optimum classifier (i.e. NN based on the cross-validation testing in this study) to further improve its performance.

The average test accuracies before moving average filter obtained in the across-trial method (69.32±4.48%) were closely represented by the cross-validation accuracies of non-NN classifiers (67.83±4.50%), although a different train/test proportion was used between the two methods. In the cross-validation, the non-overlapping 5-second epochs were permutated randomly across all three trials and divided into 5-fold. The EEG data is not entirely independent where nearby consecutive windows may be classified similarly, hence this could possibly skew the performance level in the across-trial procedure. However, this did not appear to affect our results, as observed from the similar accuracies obtained between the across-trial procedure without moving average filter (non-permutated, non-overlapping epochs) and the cross-validation procedure (permutated, non-overlapping epochs). For example, using a LDA classifier, there was no statistical difference between across-trial accuracy on test trial 3 without moving average filter (mean = 65.92%, SD = 6.63%) and the cross-validation accuracy (mean = 67.81%, SD = 4.89%); $t(9) = 1.55$, $p = 0.16$. NN classifier, however, showed a significantly higher accuracy in the cross-validation (median = 69.40%, IQR = 5.19%) compared to the across-trial testing on trial 3 (median = 65.13%, IQR = 7.07%); $p = 0.005$ using the Wilcoxon signed rank test, which could be due to the previously mentioned larger training set used in the cross-validation being more appropriate for NN. Based on this, the across-trial accuracy is not expected to significantly improve compared to the cross-validation accuracy, although the EEG data is not entirely independent over time. However, these comparisons may not fully capture the potential limitation of skewing the performance level, and it may still exist, though no significant improvement in the across-trial accuracy was observed compared to the cross-validation accuracy, as the two methods used a different train and test proportions.

The optimal threshold of the moving average filter to smooth out the classifiers' predictions was obtained through a grid search by comparing the test accuracy results, and was also subject and trial dependent. The brain is non-stationary, adaptive (e.g. aging and learning factor), and dynamic (e.g. changing along with physiological variations) [56]. In this study, moving average thresholding boosted the test accuracies by about 10%. By applying such a

method, few subjects at different test trial sets (A02, A03, A04 on test trial set #2) produced a disproportionately high false negative requiring very low threshold values (0, 0.2, 0.2 respectively). However, analyzing the data by testing on the last trial in the across-trial method (train on trials 1 & 2, test on trial 3) among the participants yielded optimized threshold values mostly centered halfway, at either 0.4 or 0.6, which can be used as a baseline threshold.

On the software side, the offline analysis aspect was the main limitation, as HCTSA was performed on each subject's entire dataset in order to obtain the set of top features. In a practical implementation, an unseen test set would not be considered to obtain the set of top features for each individual. On the other hand, the automated feature selection algorithm used in this study was performed on each subject, and the optimal features selected varied and were mostly unique across subjects. This was expected since in finding the best-separating features, HCTSA employed ~7000 feature extractions computed based on several common operations that can receive different input parameters and produce multiple outputs [49,50]. Despite this, several types of operations produced the most selected features used across all subjects (i.e. correlation and model fitting parameters) in reference to HCTSA categorization [49,50]. Future works may consider extracting features using these types of operations, finding the general best feature set for all subjects, and investigating the trade-off between personalized feature selection and performance loss, since running massive feature extractions may not be feasible for the hardware in a wearable device application. A decrease in performance is expected when using a fixed set of operations to extract the features that are common for all subjects. Future research may also consider focusing on fine-tuning the optimum classifier coupled with the selected general optimum features.

Although the required low impedance of a similar viscoelastic in-ear EEG setup has been verified in another study [23], an impedance check was not performed as part of our common EEG setup procedure due to the amplifier's hardware limitations. A one-size viscoelastic memory foam earbud was used in this experiment, which might not be optimal for all subjects, as the size and shape of the ear canal vary among individuals. Adding an impedance measurement and using a customized viscoelastic earbud for each subject may be considered to improve the system's reliability and robustness in future studies.

To further investigate the feasibility of implementing the system in practice, more studies could be conducted in which participants are strictly controlled and involved in multiple sessions in order to assess the robustness of the trained model, and compare the features obtained through rigorous massive feature extraction processes to specific, predetermined features sets (power spectral density, wavelet-based, etc.). Other physiological signals such as heart rate have also been linked to mental demands [57]. Therefore, an electrocardiogram (ECG) might also be incorporated alongside the in-ear EEG setup to enhance performance while keeping the system discreet and portable.

5. Conclusions

This study demonstrated that a two-channel in-ear EEG system used as a BCI application can distinguish between resting and visuomotor tracking states with an average accuracy of 79.30 ± 4.85% across ten subjects. In achieving this accuracy, the training data of each subject were limited to 25.3 min of EEG recordings and tested on the other 12.6 min of the recorded signals. An extensive feature extraction algorithm was employed to select the top performing features for each individual. As a personal brain monitor to detect a user's mental workload, the overall system proposed herein may be more portable and discreet, as well as easier and quicker to set up than conventional scalp EEG systems. The results of this preliminary study show potentials for future application of in-ear EEG to improve operators' safety and performance in daily activities.

Competing interests

None.

Ethics statement

The protocol was approved by the Office of Research Ethics at Simon Fraser University and written informed consents were obtained from all participants.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canadian Institutes of Health Research (CIHR), and the Canada Research Chair (CRC) program.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.medengphy.2019.03.016.

References

[1] Rolfe JM. Ergonomics and air safety. Appl Ergon 1972;3:75–81. doi:10.1016/0003-6870(72)90056-7.
[2] Raskin J. The humane interface: new directions for designing interactive systems. Boston: Addison-Wesley Professional; 2000.
[3] Kaber DB, Onal E, Endsley MR. Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. Hum Factors Ergon Manuf 2000;10:409–30. doi:10.1002/1520-6564(200023)10:4⟨409::AID-HFM4⟩3.0.CO;2-V.
[4] Kantowitz BH, Casper PA. Human workload in aviation. Hum Factors Aviat, Elsevier; 1988. p. 157–87. doi:101016/B978-0-08-057090-750012-6.
[5] Parasuraman R. Application of human performance data and quantitative models to design of automation. In: Harris D, editor. Engineering psychology cognitive Ergon Vol. 5 aerospace transportation system. Aldershot: Ashgate; 2017. p. 3–13.
[6] Parasuraman R, Mouloua M, Molloy R. Effects of adaptive task allocation on monitoring of automated systems. Hum Factors J Hum Factors Ergon Soc 1996;38:665–79. doi:10.1518/001872096778827279.
[7] Wiener EL. Beyond the sterile cockpit. Hum Factors J Hum Factors Ergon Soc 1985;27:75–90. doi:10.1177/001872088502700107.
[8] Millán JDR, Rupp R, Müller-Putz GR, Murray-Smith R, Giugliemma C, Tangermann M, et al. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. Front Neurosci 2010;4:1–15. doi:10.3389/fnins.2010.00161.
[9] Silva FPda. Mental workload, task demand and driving performance: what relation? Procedia - Soc Behav Sci 2014;162:310–19. doi:10.1016/j.sbspro.2014.12.212.
[10] Dijksterhuis C, de Waard D, Brookhuis KA, Mulder BLJM, de Jong R. Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns. Front Neurosci 2013;7. doi:10.3389/fnins.2013.00149.
[11] Galy E, Mélan C. Effects of cognitive appraisal and mental workload factors on performance in an arithmetic task. Appl Psychophysiol Biofeedback 2015;40:313–25. doi:10.1007/s10484-015-9302-0.
[12] Ke Y, Qi H, He F, Liu S, Zhao X, Zhou P, et al. An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. Front Hum Neurosci 2014;8. doi:10.3389/fnhum.2014.00703.
[13] Santiago-Espada Y, Myer RR, Latorella KA, Comstock JR. The multi-attribute task battery II (MATB-II) software for human performance and workload research: a user's guide. Hampton, Virginia: NASA Langley Research Center; 2011.
[14] Hosseini SMH, Bruno JL, Baker JM, Gundran A, Harbott LK, Gerdes JC, et al. Neural, physiological, and behavioral correlates of visuomotor cognitive load. Sci Rep 2017;7:1–9. doi:10.1038/s41598-017-07897-z.
[15] Hill SG, Iavecchia HP, Byers JC, Bittner AC, Zaklade AL, Christ RE. Comparison of four subjective workload rating scales. Hum Factors J Hum Factors Ergon Soc 1992;34:429–39. doi:10.1177/001872089203400405.
[16] Wierwille WW, Williges RC, Wierwille WW. Behavioral measures of aircrew mental workload. Hum Factors J Hum Factors Ergon Soc 1979;21:575–93. doi:10.1177/001872087902100503.

[17] Wickens CD, Hollands JG, Banbury S, Parasuraman R. Engineering psychology and human performance. Englewood Cliffs: Prentice-Hall; 2012.
[18] Gevins A, Smith ME, Leong H, McEvoy L, Whitfield S, Du R, et al. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Hum Factors J Hum Factors Ergon Soc 1998;40:79–91. doi:10.1518/001872098779480578.
[19] Pfurtscheller G, Klimesch W. Event-related desynchronization during motor behavior and visual information processing. Electroencephalogr Clin Neurophysiol Suppl 1991;42:58–65.
[20] Sterman MB, Mann CA, Kaiser DA, Suyenobu BY. Multiband topographic EEG analysis of a simulated visuomotor aviation task. Int J Psychophysiol 1994;16:49–56.
[21] Valentino DA, Arruda JE, Gold SM. Comparison of QEEG and response accuracy in good vs poorer performers during a vigilance task. Int J Psychophysiol 1993;15:123–33.
[22] Looney D, Kidmose P, Park C, Ungstrup M, Rank M, Rosenkranz K, et al. The in-the-ear recording concept: user-centered and wearable brain monitoring. IEEE Pulse 2012;3:32–42. doi:10.1109/MPUL.2012.2216717.
[23] Goverdovsky V, Looney D, Kidmose P, Mandic DP. In-ear EEG from viscoelastic generic earpieces: robust and unobtrusive 24/7 monitoring. IEEE Sens J 2016;16:271–7. doi:10.1109/JSEN.2015.2471183.
[24] Mikkelsen KB, Kappel SL, Mandic DP, Kidmose P. EEG recorded from the ear: characterizing the ear-EEG method. Front Neurosci 2015;9:1–8. doi:10.3389/fnins.2015.00438.
[25] Baddeley A.D., Hitch G. Working memory, 1974, p. 47–89. doi:10.1016/S0079-7421(08)60452-1.
[26] Bar M. A cortical mechanism for triggering top-down facilitation in visual object recognition. J Cogn Neurosci 2003;15:600–9. doi:10.1162/089892903321662976.
[27] Spiridon M, Fischl B, Kanwisher N. Location and spatial profile of category-specific regions in human extrastriate cortex. Hum Brain Mapp 2006;27:77–89. doi:10.1002/hbm.20169.
[28] Denys K. The processing of visual shape in the cerebral cortex of human and nonhuman primates: a functional magnetic resonance imaging study. J Neurosci 2004;24:2551–65. doi:10.1523/JNEUROSCI.3569-03.2004.
[29] Axmacher N, Schmitz DP, Wagner T, Elger CE, Fell J. Interactions between medial temporal lobe, prefrontal cortex, and inferior temporal regions during visual working memory: a combined intracranial eeg and functional magnetic resonance imaging study. J Neurosci 2008;28:7304–12. doi:10.1523/JNEUROSCI.1778-08.2008.
[30] Tankus A, Fried I. Visuomotor coordination and motor representation by human temporal lobe neurons. J Cogn Neurosci 2012;24:600–10. doi:10.1162/jocn_a_00160.
[31] Kropotov JD. Beta rhythms. Quantitative EEG, event-related potentials neurotheraphy. Elsevier; 2009. p. 59–76. doi:101016/B978-0-12-374512-500003-7.
[32] Jia X, Kohn A. Gamma rhythms in the brain. PLoS Biol 2011;9:e1001045. doi:10.1371/journal.pbio.1001045.
[33] Henrie JA, Shapley R. LFP power spectra in V1 cortex: the graded effect of stimulus contrast. J Neurophysiol 2005;94:479–90. doi:10.1152/jn.00919.2004.
[34] Fries P, Reynolds JH, Rorie AE, Desimone R. Modulation of oscillatory neuronal synchronization by selective visual attention. Science 2001;291:1560–3. doi:10.1126/science.291.5508.1560.
[35] Pesaran B, Pezaris JS, Sahani M, Mitra PP, Andersen RA. Temporal structure in neuronal activity during working memory in macaque parietal cortex. Nat Neurosci 2002;5:805–11. doi:10.1038/nn890.
[36] Bauer EP, Paz R, Paré D. Gamma oscillations coordinate amygdalo-rhinal interactions during learning. J Neurosci 2007;27:9369–79. doi:10.1523/JNEUROSCI.2153-07.2007.
[37] Siegel M, Donner TH, Oostenveld R, Fries P, Engel AK. Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. Neuron 2008;60:709–19. doi:10.1016/j.neuron.2008.09.010.
[38] Chapotot F, Gronfier C, Jouny C, Muzet A, Brandenberger G. Cortisol secretion is related to electroencephalographic alertness in human subjects during daytime wakefulness. J Clin Endocrinol Metab 1998;83:4263–8. doi:10.1210/jcem.83.12.5326.
[39] Grønli J, Rempe MJ, Clegern WC, Schmidt M, Wisor JP. Beta EEG reflects sensory processing in active wakefulness and homeostatic sleep drive in quiet wakefulness. J Sleep Res 2016;25:257–68. doi:10.1111/jsr.12380.
[40] Pesonen M, Hämäläinen H, Krause CM. Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. Brain Res 2007;1138:171–7. doi:10.1016/j.brainres.2006.12.076.
[41] Pfurtscheller G, Lopes da Silva FH. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 1999;110:1842–57.
[42] Yuan H, Liu T, Szarkowski R, Rios C, Ashe J, He B. Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: an EEG and fMRI study of motor imagery and movements. Neuroimage 2010;49:2596–606. doi:10.1016/j.neuroimage.2009.10.028.
[43] Jensen O, Kaiser J, Lachaux J-P. Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci 2007;30:317–24. doi:10.1016/j.tins.2007.05.001.
[44] Sanei S, Chambers JA. EEG signal processing. Chicester: John Wiley & Sons, Ltd; 2007.
[45] Laine TI, Bauer KW, Lanning JW, Russell CA, Wilson GF. Selection of input features across subjects for classifying crewmember workload using artificial neural networks. IEEE Trans Syst Man Cybern Part ASystems Humans 2002;32:691–704. doi:10.1109/TSMCA.2002.807036.
[46] Lee S, Kim DJ, Svenkeson D, Parras G, Oishi MMK, McKeown MJ. Multifaceted effects of noisy galvanic vestibular stimulation on manual tracking behavior in Parkinson's disease. Front Syst Neurosci 2015;9. doi:10.3389/fnsys.2015.00005.
[47] Rao R, Scherer R. Brain-computer interfacing [In the Spotlight]. IEEE Signal Process Mag 2010;27:150–2. doi:10.1109/MSP.2010.936774.
[48] Abdulkader SN, Atia A, Mostafa MSM. Brain computer interfacing: applications and challenges. Egypt Inf J 2015;16:213–30. doi:10.1016/j.eij.2015.06.002.
[49] Fulcher BD, Little MA, Jones NS. Highly comparative time-series analysis: the empirical structure of time series and their methods. J R Soc Interface 2013;10 20130048 20130048. doi:10.1098/rsif.2013.0048.
[50] Fulcher BD, Jones NS. hctsa : a computational framework for automated time-series phenotyping using massive feature extraction. Cell Syst 2017;5:527–31 e3. doi:10.1016/j.cels.2017.10.001.
[51] Fulcher BD, Jones NS. Highly comparative feature-based time-series classification. IEEE Trans Knowl Data Eng 2014;26:3026–37. doi:10.1109/TKDE.2014.2316504.
[52] Oshiro TM, Perez PS, Baranauskas JA. How many trees in a random forest?. Machine Learning and Data Mining in Pattern Recognition. Berlin, Heidelberg: Springer; 2012. p. 154–68. doi:101007/978-3-642-31537-4_13.
[53] Marquart G, Cabrall C, Winter J De. Review of eye-related measures of drivers' mental workload. Procedia Manuf 2015;3:2854–61. doi:10.1016/j.promfg.2015.07.783.
[54] Gao Q, Wang Y, Song F, Li Z, Dong X. Mental workload measurement for emergency operating procedures in digital nuclear power plants. Ergonomics 2013;56. doi:10.1080/00140139.2013.790483.
[55] Veltman JA, Gaillard AW. Physiological workload reactions to increasing levels of task difficulty. Ergonomics 1998;41:656–69. doi:10.1080/001401398186829.
[56] Lance BJ, Kerick SE, Ries AJ, Oie KS, Mcdowell K. Brain-computer interface technologies in the coming decades. Proc IEEE 2012;100:1585–99. doi:10.1109/JPROC.2012.2184830.
[57] Fallahi M, Motamedzade M, Heidarimoghadam R, Soltanian AR, Miyake S. Assessment of operators' mental workload using physiological and subjective measures in cement, city traffic and power plant control centers. Heal Promot Perspect 2016;6:96–103. doi:10.15171/hpp.2016.17.