Noname manuscript No.
(will be inserted by the editor)

Investigation of Different Classifiers and Channel Configurations of a Mobile P300-based Brain-Computer Interface

Simone A. Ludwig and Jun Kong

Received: date / Accepted: date

Abstract Innovative methods and new technologies have significantly improved the quality of our daily life. However, disabled people, for example those that cannot use their arms and legs anymore, often cannot benefit from these developments, since they cannot use their hands to interact with traditional interaction methods (such as mouse or keyboard) to communicate with a computer system. A Brain-Computer Interface (BCI) system allows such a disabled person to control an external device via brain waves. Past research mostly dealt with static interfaces, which limit users to a stationary location. However, since we are living in a world that is highly mobile, this paper evaluates a speller interface on a mobile phone used in a moving condition. The spelling experiments were conducted with 14 able-bodied subjects using visual flashes as the stimulus to spell 47 alphanumeric characters (38 letters and 9 numbers). This data was then used for the classification experiments. In particular, two research directions are pursued. The first investigates the impact of different classification algorithms, and the second direction looks at the channel configuration, i.e., which channels are most beneficial in terms of achieving the highest classification accuracy. The evaluation results indicate that the Bayesian Linear Discriminant Analysis algorithm achieves the best accuracy. Also, the findings of the investigation on the channel configuration, which can potentially reduce the amount of data processing on a mobile device with limited computing capacity, is especially useful in mobile BCIs.

Keywords P300 speller interface · Classification · Channel selection

Simone A. Ludwig and Jun Kong
Department of Computer Science
North Dakota State University
Fargo, ND, USA
E-mail: {simone.ludwig,jun.kong}@ndsu.edu

1 Introduction and Related Work

A brain-computer interface (BCI) allows paralyzed people to engage in independent activity. Different from traditional human-computer interaction that is based on muscular movement, BCI uses the brain signals collected through the electroencephalogram (EEG) to directly control an external device. Especially, the P300 component of the event-related potential (ERP), which indicates a long-lasting high-amplitude positive ERP component elicited around 300 ms after the onset of a stimulus [1], is commonly used to analyze a user's intention. P300 spellers have emerged as the most common application of BCI. The principle of P300 spellers is to elicit a P300 brain signal by flashing a target symbol, and the P300 component serves as the input to a classifier for detecting the information encoded by the visual flash.

The row-column paradigm (RCP) [2], being pioneering work, consisted of a 6-by-6 matrix of 36 alphanumeric characters displayed on a screen in front of a user. The rows and columns are flashed alternately in a randomized order, while the user's attention is focused upon the target character to silently count how many times the target character is flashed. Once the row or column containing a target character is flashed, the P300 component is elicited. A classification algorithm is then applied to the collected EEG data to determine the row and column that elicited the largest P300 amplitude, after which the target is identified as the intersection of that row and that column. Though RCP has achieved big success, this paradigm has faced several challenges such as the adjacency, crowding, and fatigue problems. In order to address those issues, various approaches have been proposed to improve the classification accuracy, such as the Single Character Paradigm [3], the Checkerboard Paradigm [4], the region-based BCI speller paradigm [5], and the Zigzag paradigm [6].

Since the processing of EEG data is time consuming, the above research approaches of P300 spellers were experimented with on a desktop, in front of which a user sits in a comfortable chair in a quiet environment. Since mobile devices nowadays are part of our daily life, mobile-brain interfaces have attracted a lot of attention recently. Especially, the rapid hardware development makes it feasible to directly collect and process EEG data on a mobile device. For example in [7], a wireless brain computer interface was developed to detect real-time driver's drowsiness; in [8] customized EEG recording and signal processing modules for a truly wearable BCI is presented; in [9] a small and wireless mobile EEG system was presented that performed on average as good as a high-quality wired EEG amplifier; [10] proved that reliable EEG data can be recorded through a mobile EEG system, which used a smartphone to deliver a stimulus and to collect and process EEG data. The validity of Mobile EEG systems was also evaluated in the condition of walking outdoors [11, 12] or in a gaming control application [13]. Several studies demonstrated that high-density EEG can be used to study brain dynamics during whole body movements [14, 15]. The above studies provide a solid foundation to justify the feasibility of wireless EEG systems that overcome the mobility limitation in natural environments.

Being a low-cost wireless device, the Emotiv EEG headset has been commonly used to develop mobile BCI systems [11,12,16]. Based on the Emotiv EEG device, this paper focuses on investigating how mobility and a small screen affect the classification accuracy in mobile P300 spellers. For example, NeuroPhone [16] provides a P300-speller similar interface that flashes a sequence of photos of contacts. A P300 component is elicited when the flashed photo matched with whom a user wishes to dial. Distinct from other P300 spellers, the Edges paradigm (EP) presents a square adjacent to each column/row in the outer boundary of a 6 by 6 matrix [17]. More specifically, a square is displayed in the bottom of an odd column and on the top of an even column, and to the left of an odd row and to the right of an even row. For the EP flashing technique, the EP's edge point flash replaces the row or column flash. The EP doubled the distance between flash objects and decreased the number of flash objects surrounding a target flash object. In other words, the visual interface design of the EP maximized the distance between edge squares within the dimensional constraints of a smartphone screen, and thus is especially suited to a mobile BCI interface. To validate the feasibility of a mobile P300 speller, EP was implemented on a simulated mobile device in the realistic setting of a moving wheelchair [18]. The above study proved the feasibility of a mobile EP speller.

This paper investigates two research questions. The first is whether using a mobile environment and interface has any influence on the classification accuracy of the different classifiers. We have experimented with classifiers that are commonly used in the BCI domain as well as applied classifiers that have achieved good accuracy in the data mining area. The second research question is concerned with the investigation of different channel selections. Researchers have shown quite recently that using different channels during the classification task for different subjects results in different classification accuracy [19]. In our study first of all we have used a different BCI device (Emotiv), and secondly the data was collected using a mobile interface simulated in a moving environment rather than using a static interface as had been studied in the past.

2 Experiments and Methods

2.1 Speller Interface and Experiments

14 able-bodied human subjects volunteered to participate in the study. The subject population consisted of all male participants between ages of 22 to 35 (mean: 28 years, std. dev.: 3.8 years). The participants reported normal or corrected vision, and none had a neurological disease. All participants gave informed consent, and the user study was approved by the Institutional Review Board of North Dakota State University.

The EEG signals were recorded with a Emotiv EEG wireless headset (EPOC+ research edition) [21]. Fourteen saline electrodes recorded the EEG signals of the participants during the study. The fourteen electrodes of the

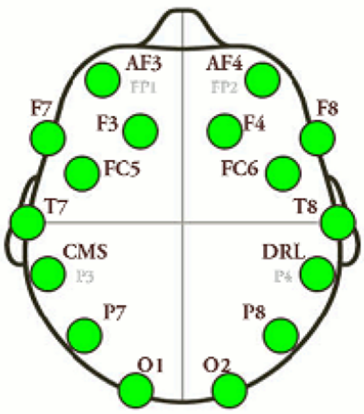


Fig. 1: EPOC electrode positions [20]

Emotiv EEG device were placed on the scalp of the participant based on the International 10-20 system locations [22] (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4). Figure 1 shows the electrode locations on the scalp of a user. The EEG signals were digitized at 128Hz and bandpass filtered (0.16-45Hz) before they were transmitted via Bluetooth to a Surface Pro 2 device.

Our approach applies Matlab to analyze the brain signals. Unfortunately, Matlab is only available on the desktop, not on mobile devices. Thus, without loosing generality, we simulated a Samsung Galaxy S4 on the Surface Pro 2 and implemented a mobile P300 speller with the size of 6.375×9.3 cm. During the experiments, the distance between a subject's head and the simulator is about 60 cm. Figure 2 shows the speller interface [18]. The subjects are being pushed in a wheelchair during the spelling experiments. A visual flash was used as the visual stimulus in this study. The flash objects are the row edge points and column edge points. For each target character there were two consecutive stages of random sequences of flashes [23] punctuated by 2-second intervals. During the first stage there were 72 row edge point flashes, whereas during the second stage there were 72 column edge point flashes. The stimulus-onset asynchrony duration was 120ms and consisted of 70ms for the flash duration, and 50ms for the inter-stimulus interval. During each stage, each target object flashed 12 times, and thus, 12 out of the 72 flashes were the target object flashes and the remaining 60 flashes were the non-target flashes. More details can be found in [18].

Each participant completed the spelling experiments over two days whereby one experiment was conducted on each day and lasted about 60 minutes. During the calibration phase, the Emotiv device was placed on the participant's head and the subjects were instructed to gaze at the target column edge point during the first stage, and then at the target row edge point at the second stage. The column edge points disappeared during the first stage whereas the row edge points disappeared during the second stage. Moreover, the participants

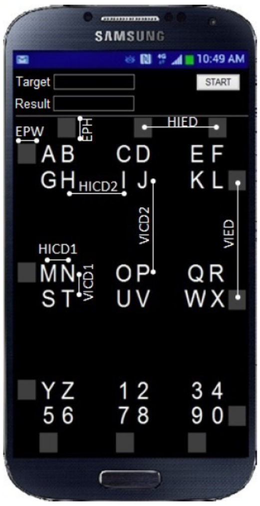


Fig. 2: Spelling interface on mobile device [18]

were asked to silently count the number of target object flashes. Afterwards, the experiments were conducted.

Two experiments were used with different sets of words and numbers. The first spelling experiment consisted of 29 alphanumeric characters and the participants were asked to spell 6 words and a number (LAP, ROD, BAND, FLAG, DRINK, MINUTE, and 9253). The second spelling experiment consisted of 18 alphanumeric characters and included 3 words and a number (FOX, LION, CRAFTS, 40150). More details can be found in [18]. The data obtained from the first spelling experiment was used as the training data, and the data from the second spelling experiment was used as the testing data for the classification. Please note that it is not possible to distinguish spelling errors of the subjects during the experiments since we did not provide a backspace button.

2.2 Preprocessing and Feature Extraction

Before a classification model can be learned and validated, the EEG signals need to undergo a series of preprocessing steps. These include referencing, filtering, downsampling, windsorizing, electrode selection, and feature vector construction. The referencing step is done using the two referencing electrodes (CMS and DRL) in order to average the multichannel EEG signals from the 14 electrodes. This is done in order to reduce the noise of the measurements by 30 percent [21]. Then, a bandpass filter is applied and the EEG signals is downsampled from 2048Hz to 128Hz. Afterwards, the EEG signals are winsorized to remove large amplitude noise signals such as eye blinks and movements as well as muscle activity [24]. All amplitude values below the 5th percentile or above the 95th percentile for each channel were replaced by the 5th percentile or the 95th percentile values, respectively. Furthermore, the EEG signals are also normalized applying the zero-mean normalization method by which each EEG data channel was normalized based on the mean and standard deviation of the same channel according to the calibration data. Following is the electrode selection stage. We have selected and experimented with different channel configurations, which are explained in detail in Section 2.3.2. The last preprocessing step is the feature vector construction. The EEG data were extracted based on the time domain. However, the EEG data for each single target character were epoched to 600ms from each flash onset. Thus, each epoch consisting of 600ms is a segment, and each segment corresponds to one visual flash. Each segment contained data recorded by the selection channels (14 when all were used) with 77 time points. Thus, for each target character with 144 flashes, there were 144 segments, i.e., one segment corresponded to one flash; 72 row edge point segments and 72 column edge point segments. Therefore, the samples from the selected electrodes were concatenated into feature vectors. The dimensionality of the feature vector is $N_e \times N_t$, where N_e represents the number of electrodes, and N_t represents the number of temporal samples in one trial.

2.3 Classification

The aim of the classification in a BCI system is to capture a user's intention on the basis of a feature vector that characterizes the brain activity. There are two possible methods by which this can be achieved; one is using a regression algorithm that uses the extracted features as independent variables in order to predict the user intentions, and the second is using a classification algorithm that use the extracted features as independent variable in order to define boundaries between the different targets in the feature space. Classification is the more popular approach for BCI systems [25]. In this study, several different classification algorithms have been applied to the data and are described in detail below.

2.3.1 Classification Algorithms

Fisher's Linear Discriminant Analysis (LDA) is a widely applied classifier for P300 BCI systems [26,27] outperforming other classifiers [28]. The algorithm works by finding a projection from the N-dimensional feature space onto a one-dimensional space $w^T f$ for which the variance between the two classes (target and non-target) versus the variance within the classes is maximized. Let us assume that the target represents $+1$ and the non-target represents -1, then the optimal projection is estimated as $w = (\sum_{-1} + \sum_{+1})^{-1}(\mu_{+1} - \mu_{-1})$, where $\sum$ is the covariance, and μ is the mean of the two classes that are to be separated.

Stepwise Linear Discriminant Analysis (SWLDA) has also been used in relation to P300 BCI speller applications [29,30]. SWLDA is an extension to LDA with the addition of an incorporated filter feature selection. SWLDA works by adding and removing terms from the linear discriminant model based on their statistical significance in regression, and therefore producing a model that is adjustable to the training data. It has been shown that SWLDA performs equally well or even better than several other classification methods in P300 BCI applications [26].

Bayesian Linear Discriminant Analysis (BLDA) has also been applied to P300 BCI systems [31]. BLDA is based on a probabilistic regression network whereby the targets t_i (for binary classification they are $+1$ and -1) are linearly dependent on the observed features $f^i = [f_1^i, ..., f_N^i]^T$ with an additive Gaussian noise term $\epsilon_n : t_i = w^T f^i + \epsilon_i$. By assuming an independent generation of the examples from a data set, the likelihood of all data is $p(t|w, \sigma^2) = \prod_{i=1}^{N} (2\pi\sigma^2)^{-1/2} exp(-\frac{(t_i - w^T f^i)^2}{2\sigma^2})$. Also, the prior distribution over all weights with a zero-means Gaussian is $p(w|\alpha) = \prod_{j=1}^{n} (\frac{\alpha}{2\pi})^{1/2} exp(-\frac{\alpha}{2} w_j^2)$. Applying Bayes' rule, the posterior distribution is defined as $p(w|t, \alpha, \sigma^2) = \frac{p(t|w,\sigma^2)p(w|\alpha)}{p(t|\alpha,\sigma^2})$ which is a Gaussian with mean $\mu = (F^T F + \sigma^2 \alpha I)^{-1} F^T t$ and covariance matrix $\sum = \sigma^2 (F^T F + \sigma^2 \alpha I)^{-1}$, where I is the identity matrix, F represents a matrix with each row corresponding to a training example in the feature space, and t corresponds to a column vector of true labels for all corresponding training samples. Thus, the resulting hyperplane will be $\mu^T f$. The regression parameters σ and α can be tuned with an automatic, iterative procedure [31].

Support Vector Machine (SVM) is considered a good classifier for BCI applications [26,32]. The classification process of SVM is to find the separating hyperplane between two classes as to maximize the distance between the hyperplane and the closest points from both classes. This results in the maximization of the margin of the two classes. Given that it is not always the case that two classes are linearly separable, SVM is generalized to the case where the data points are allowed to fall within the margin by adding a regularization term. One way to achieve this is to use the method of linear least squares [33] to solve the optimization problem $min_{w,b,e}((1/2)w^T w) + \gamma \sum_{i=1}^{N} e_i^2$ with respect to $y_i(w^T f^i + b) = 1 - e_i, i = 1, ..., n$, where f_i corresponds to the training points in the feature space, and y_i is the associated output (+1 for the responses to the target stimulus, and -1 for the non-target stimulus). The regularization parameter is estimated via a line search on cross-validation results.

Multilayer Feed-forward Neural Network (NN) represent a system of interconnected neurons that exchange messages between each other. The learning capability is achieved by the adaptive ability of the neurons that is performed by changes to the connections via numeric weights [34]. As a classifier applied to BCI, NN have been used to classify different number of tasks (two [35], three [36], and five [37] as well as to design synchronous [38] and asynchronous [39] BCIs). NN with a single hidden layer and with sigmoid activation functions is a universal approximator [34]. Thus, the classifier is $y(f, w, b) = \sum_{i=1}^{M} w_i^2 F(\sum_{j=1}^{N} w_{ji}^1 f_j + b_i) + b$, where M is the number of neurons in the hidden layer, with sigmoidal activation function $F(t) = 1/(1 + exp(t))$, N is the number of observed features, $b = b_1, ..., b_N, b$

and $w = w_1^2, ..., w_M^2, w_{11}^1, ..., w_{NM}^1$ is the set of thresholds and weight coefficients, respectively. The weight coefficients are optimized using the Levenberg-Marquardt back propagation method based on the desired outcome (target and non-target). The implementation with a 10-neuron single hidden layer neural network is referred in this paper as NN-10. Furthermore, it has been proven that a multilayer architecture with several hidden layers gives neural networks the ability of being universal approximators [40]. In comparison, a fixed kernel SVM has a depth of 2 [41], and boosting methods (described below) have a depth of 3 [42]. Thus, several multi-layer neural network configurations (varying from 3 to 5 hidden layers) were experimented with and the best performing network with 3 hidden layers of size 60, 30, and 10 were chosen (referred to as NN-60-30-10; resulting in depth of 4).

Boosting is a method for improving the accuracy of any given learning algorithm, which is a widely used and powerful prediction technique. Boosting works by sequentially constructing an ensemble of weak classifiers. A weak classifier is a very simple model that usually only has a slightly better accuracy than a random classifier that achieves an accuracy of 50%. The weak classifiers are constructed iteratively from the training data over hundreds or thousands of iterations. During each iteration, the data points of the training data are reweighted according to their performance, i.e., the larger weights are given to misclassified data points. The weighted predictions from each weak classifier are combined using voting to compute a final prediction of the overall outcome [43].

Three different boosting methods were chosen:

- ***AdaBoost*** is a popular boosting algorithm for binary classification. Similar to the overall boosting method description above, the AdaBoost algorithm increases the weight for the observations that are misclassified by a learner and reduces the weights for observations that are correctly classified by the learner. The next learner is then trained on the data with the updated weights. After the training finishes, the AdaBoost algorithm computes the prediction for new unseen data [44].
- ***RUSBoost*** (RUS stands for Random Under Sampling) takes the number of members in the class with the fewest members in the training data as the basic unit for sampling. The classes with more members are undersampled by taking only a certain number of observations for every class. The boosting procedure follows the AdaBoost algorithm for reweighting and constructing the ensemble [45].
- ***LogitBoost*** works similar to AdaBoost with the exception that it minimizes the binomial deviance. The binomial deviance assigns less weight to badly misclassified observations. LogitBoot is said to give better average accuracy results compared to AdaBoost for data with poorly separable classes [46].

2.3.2 Channel selection

Past research has shown that the selection of channel locations is an important factor in BCI systems. The selection impacts the system performance quite significantly [47] where the evaluation of a sensitivity analysis revealed that identifying an appropriate channel set for an individual was more important than factors such as features space, pre-processing, and classifier choice. Furthermore, empirical results summarized the benefits of channel selection [48–50].

Channel selection has been investigated for BCI paradigms such as for motor imagery tasks [51], and mutual information maximization for cognitive load classification [52]. In [49], the use of a channel selection procedure integrated within the training of a support vector machine classifier is done. Furthermore, a BLDA classifier in connection with Particle Swarm Optimization was used to select channels in a system that spelled Chinese characters [53]. As with reference to the P300 speller classification, [50] reported an active channel selection method that improved the classification performance comprising of a sequential reverse selection process. The most recent study published in 2014 introduced a jumpwise regression method that extended the SWLDA [19]. The findings were that the proposed method offers accuracy gains that are similar to the best-performing feature-selection methods and to be robust enough for online use.

All this past research work motivated the study of analyzing different combinations of the 14 channels available for the Emotiv BCI device. Extensive experiments were performed to identify the most important channels to be used for the classification of the spelling tasks for different subjects.

3 Evaluation and Results

Two different sets of experiments were conducted. The first set of experiments applied nine different classifiers based on all 14 channels and compared them in terms of accuracy and speed. The second set of experiments were done using different combinations of selected channels. Again the accuracy was evaluated and is reported below.

As mentioned, the first set of experiments involved the accuracy of the different classifiers listed in Section 2.3.1. All 14 channels were used and the spelling data was split into training data and test data. As mentioned before, the training data consisted of the spelling of six words and one number: LAP, ROD, BAND, FLAG, DRINK, MINUTE, and 9253. The testing data consisted of the spelling of three words and one number: FOX, LION, CRAFTS, 40150. The classifier was trained on the training data, and evaluated against the test data and averaged over all 14 subjects. Tables 1 and 2 show the results. Given that the data is split into 12 iterations, the accuracy values are listed for each classifier per iteration. The results highlight that the accuracy improved for all classifiers with increasing iterations. Furthermore, the classifiers that have been applied to BCI applications such as BLDA, SWLDA, NN-60-30-10, SVM, and NN-10 show the better results (above 68%) compared to the boosting methods (AdaBoost, RUSBoost, LogitBoost). Boosting methods are frequently applied for classification tasks within the data mining community achieving usually good classification accuracy. However, in our study the accuracy values obtained range between 61% and 62% for the boosting methods.

Table 1: Classification accuracy of different classifiers - Iterations 1 to 6

Iterations	**1**	**2**	**3**	**4**	**5**	**6**
LDA	0.2967	0.3302	0.3807	0.4363	0.4563	0.4670
BLDA	0.4932	0.5521	0.6158	0.6911	0.7481	0.7616
SWLDA	0.3799	0.4919	0.5244	0.6027	0.6218	0.6517
SVM	0.3813	0.4195	0.4748	0.5180	0.5699	0.6063
NN-10	0.3400	0.4110	0.4490	0.4635	0.5250	0.5743
NN-60-30-10	0.3655	0.4485	0.4857	0.5598	0.5988	0.6266
AdaBoost	0.3004	0.3798	0.4083	0.4241	0.4704	0.4916
RUSBoost	0.3015	0.3649	0.3949	0.4190	0.4592	0.4719
LogitBoost	0.2998	0.3628	0.3938	0.4295	0.4692	0.4715

Table 2: Classification accuracy of different classifiers - Iterations 7 to 12

Iterations	**7**	**8**	**9**	**10**	**11**	**12**
LDA	0.4838	0.5033	0.5207	0.5521	0.5665	0.5674
BLDA	0.7911	0.8332	0.8766	0.8854	0.9185	0.9220
SWLDA	0.6953	0.7142	0.7394	0.7666	0.7806	0.8018
SVM	0.6439	0.6614	0.6855	0.6979	0.7145	0.7324
NN-10	0.5841	0.5945	0.6284	0.6496	0.6653	0.6896
NN-60-30-10	0.6476	0.6722	0.7118	0.7430	0.7811	0.7857
AdaBoost	0.5303	0.5406	0.5675	0.5688	0.5833	0.6230
RUSBoost	0.5168	0.5400	0.5515	0.5786	0.5956	0.6225
LogitBoost	0.5058	0.5299	0.5198	0.5295	0.5567	0.6108

Figure 3 shows the results graphically for the five best-performing classifiers. The figure shows that BLDA by far is the best classifier scoring an accuracy of 92.2% followed by SWLDA scoring 80.2%, NN-60-30-10 with 78.6%, SVM achieving 73.2%, and NN-10 obtaining 69.0%. The "deeper" neural network (NN-60-30-10) outperforms the "shallow" neural network (NN-10) by almost 10%. The reason for this is that NN-60-30-10 with 3 hidden layers can learn the non-linearity of the data much better than NN-10.

In terms of classification speed, each classifier was evaluated in terms of execution time for the classification and testing for one subject, i.e., this consists of the time to train the classifier plus the time to test the classifier. Table 3 shows the execution time obtained. The fastest classifier is BLDA followed by NN-10 and SWLDA. The remaining classifiers have much longer execution times with AdaBoost having the longest with 1039 seconds (approx. 17 minutes).

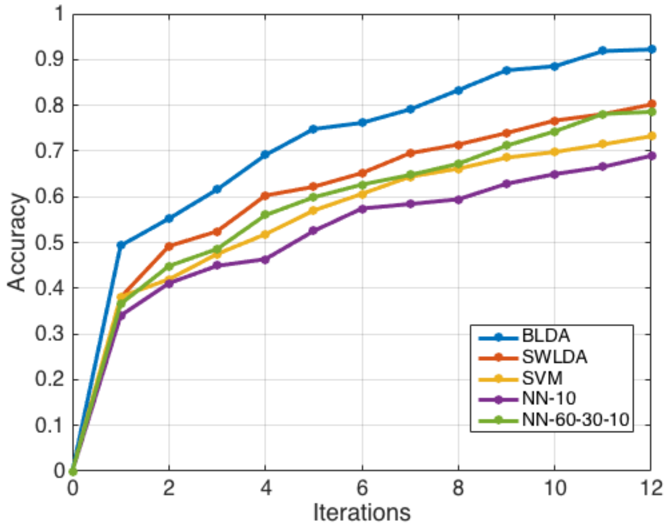


Fig. 3: Accuracy of the four best performing classifiers

Table 3: Execution time in seconds of classifiers

Classifier	BLDA	SWLDA	SVM	NN-10	NN-60-30-10	Ada	RUS	Logit
Execution time	7.13	30.12	440.89	12.43	207.58	1039.06	567.11	492.29

Table 4 shows the classification and misclassification results in the form of a confusion matrix for Subject 1 using the best performing classifier BLDA. The matrix consists of 11 columns and 11 rows. Column 6 and row 6 represent the number of correct classifications, whereas the other numbers in the matrix represent misclassification. For example in row 6, the "1" in column 1 represents one misclassification. This misclassification represents a wrongly spelt word that was identified in the correct row, but in the fifth left adjacent column compared to the actual character in the speller display. The "8" in column 6 and row 5 represents 8 misclassified characters for which a character was selected that was adjacently located one row above the target character. The BLDA classifier identified 213 characters correctly and misclassified 38 characters, thus, achieving a classification accuracy of 84.5%. Furthermore, it can be seen that the misclassifications are mostly concerned with either choosing the wrong adjacent column within a row, or the other way around, choosing the wrong adjacent row within a column.

The second sets of experiments were conducted in order to evaluate the impact of the different channel selections on the subjects with the Emotiv BCI device. In order to identify the most important channels based on each of the 14 subjects, the following procedure for each subject was performed. First, each channel was singly selected and the one achieving the highest accuracy was selected and kept. Then in the next round, out of the remaining

Table 4: Confusion matrix showing the correct and incorrect spellings of the experiments of one subject (BLDA classifier)

	-5	-4	-3	-2	-1	0	1	2	3	4	5
-5	0	0	0	0	0	0	0	0	0	0	0
-4	0	0	0	0	0	2	0	0	0	0	0
-3	0	0	0	0	0	1	0	0	0	0	0
-2	0	0	0	0	0	3	0	1	0	0	0
-1	0	0	0	0	0	8	0	0	0	0	0
0	1	2	4	2	2	213	4	1	0	2	1
1	0	0	0	0	0	1	0	0	0	0	0
2	0	0	0	0	0	2	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	1	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0

13 channels again each channel in combination with the already selected channel was evaluated and again the channel achieving the highest accuracy was kept. Then again, the same procedure was performed choosing the next best performing channel until all channels were used.

Table 5: Selected channels achieving highest accuracy

	Channels
S1	8,9,1,2,3,12,7,4,6
S2	12,6,9,8,10,13,2,7
S3	8,6,3,10
S4	7,2,3,5,10
S5	6,12,3,2,8,5,9,1,7
S6	12,1,9,4,10,7,2,3
S7	8,7,3,5,1
S8	7,9,3,5,10,1,4
S9	7,3,6,9,12
S10	3,11,8,13,14,6,7,2
S11	3,14,7,8,2,10
S12	7,9,12
S13	3,6,8
S14	7,9,1,3,4,6,2,5,8,12

Table 5 shows the selected channels for each subject until the best accuracy was achieved. It can be seen that different numbers of channels are responsible to achieve the best accuracy for each subject. For example, for Subject 13 only 3 channels were necessary to achieve the best accuracy, whereas for Subject 14, 10 channels were necessary. In addition, we can see that different channel combinations have an impact on the classification accuracy for each subject.

Based on the 14 subjects (see Table 6), the most important channels, i.e., the ones that are selected for most subjects, are Channel 3, 6, 7, 8, and 9 which correspond to electrode locations F3, P7, O1, O2, and P8 according to the international numbering system. F3 and O1 has the highest number of occurrences with 12 out of 14 followed by O2 with 9, and P7 and P8 are

Table 6: Number of occurrences for selected channels achieving the highest accuracy values

Channel	1	2	3	4	5	6	7	8	9	10	11	12	13	14
# of occurrences	5	7	12	4	5	8	12	9	8	6	1	6	2	2

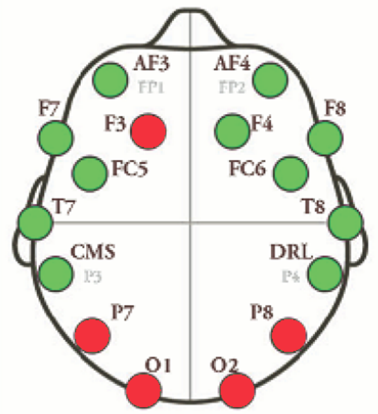


Fig. 4: EPOC electrode positions with overall most important electrodes marked in red [20]

selected in 8 out of the 14 subjects. The most important channel locations as with regards to the scalp locations are given in Figure 4.

Tables 7 and 8 show the actual accuracy values achieved with the different channel selections. As mentioned before, each subject needs a different number of selected channels for achieving the best classification results. The numbers in bold indicate the best accuracy achieved for each subject. Furthermore, we can see that an accuracy of 100% is achieved by five subjects. It is also interesting to observe that for certain subjects the accuracy decreases when more channels are added (Subjects 2, 5, 6, 7, 8, 9, 10, 11, 13 and 14).

Table 7: Accuracy for optimized channel selections - Subjects 1 to 7

# of channels	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**
1	0.8500	0.5958	0.8188	0.9125	0.6583	0.7875	0.7063
2	0.9479	0.7583	0.9688	0.9375	0.8188	0.8604	0.8250
3	0.9479	0.8875	0.9750	0.9375	0.8750	0.9375	0.8813
4	0.9479	0.9438	**1.0000**	0.9375	0.9167	0.9375	0.9021
5	0.9479	0.9438	1.0000	**0.9792**	0.9167	0.9375	**0.9229**
6	0.9479	0.9438	1.0000	0.9792	0.9375	0.9375	0.9229
7	0.9479	0.9438	1.0000	0.9792	0.9333	0.9375	0.9229
8	0.9479	0.9688	1.0000	0.9792	0.9333	**0.9583**	0.9021
9	**0.9792**	0.9375	1.0000	0.9792	**0.9542**	0.9583	0.9021
10	0.9792	0.9229	1.0000	0.9792	0.9542	0.9583	0.9229
11	0.9792	0.9479	1.0000	0.9792	0.9542	0.9583	0.9021
12	0.9792	0.9688	1.0000	0.9792	0.9542	0.9583	0.9021
13	0.9792	**1.0000**	1.0000	0.9792	0.9333	0.9583	0.9021
14	0.9792	0.9479	**1.0000**	0.9792	0.8917	0.9375	0.8708

Table 8: Accuracy for optimized channel selections - Subjects 8 to 14

# of channels	S8	S9	S10	S11	S12	S13	S14
1	0.6083	0.6604	0.6500	0.7000	0.9063	0.8604	0.8604
2	0.7625	0.7583	0.7792	0.8396	0.9688	0.9583	0.9479
3	0.8188	0.8458	0.8458	0.8396	**1.0000**	**1.0000**	0.9479
4	0.8708	0.8563	0.8875	0.8646	1.0000	1.0000	0.9688
5	0.9083	**0.9021**	0.7542	0.8646	1.0000	1.0000	0.9688
6	**0.9333**	0.9021	0.8042	**0.8854**	1.0000	1.0000	0.9688
7	0.9333	0.9021	0.9125	0.8854	1.0000	1.0000	0.9688
8	0.9333	0.9021	**0.9375**	0.8854	1.0000	1.0000	0.9688
9	0.9333	0.8813	0.9375	0.8646	1.0000	1.0000	0.9688
10	0.9125	0.8500	0.9167	0.8646	1.0000	1.0000	**1.0000**
11	0.9083	0.8292	0.9125	0.8854	1.0000	1.0000	1.0000
12	0.9083	0.8500	0.9375	0.8854	1.0000	1.0000	1.0000
13	0.9083	0.8500	0.9375	0.8396	1.0000	0.9792	1.0000
14	0.8771	0.7563	0.8813	0.8604	1.0000	0.9583	0.9688

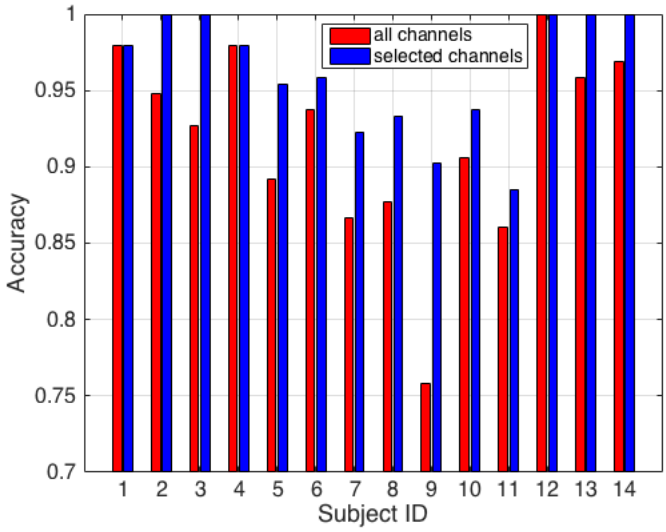


Fig. 5: Accuracy of different classifiers comparing all channels with the selected channels setup

To further analyze this, the accuracy achieved by a selected set of channels versus using all channels is shown in Figure 5. The red bars show the accuracy achieved when all 14 channels are used during the classification task, whereas the blue bars show the best accuracy achieved when only a selected set of channels is used. It seems that in some cases if more channels are used more redundancy and noise seem to be introduced and this in turn leads to lower accuracy values.

4 Discussion

The difference of this particular study was that the spelling experiments were conducted in a mobile environment since the subjects were sitting in a wheelchair being pushed around. In addition, the interface was simulated on a Samsung Galaxy S4 on the Surface Pro 2. We investigated whether this mobile environment would have an impact on the classification ability of the different classifiers. Furthermore, the study regarding the different channel selection configurations confirmed latest research stating that each individual achieves best spelling accuracy when different channels are selected. Thus, the novelty of this paper was to investigate whether a mobile environment and spelling interface would influence the classification accuracy and the channel selection as compared to a static environment.

In particular, spelling experiments were conducted with 14 able-bodied subjects. The spelling experiments consisted of visual flashes that were used as the stimulus using flash objects representing the row edge points and column edge points. 47 alphanumeric characters were used as the training and test data. The data collected from these spelling experiments were then used for our classification experiments.

The first set of experiments conducted experimented with different classification algorithms applied to the spelling task. Classifiers commonly used in the BCI area as well as classifiers applied in the data mining area were compared with. The results revealed that the BLDA classifier (accuracy of 92.2%) is the best classifier for mobile interfaces followed by SWLDA, NN-60-30-10, SVM, and NN-10. The boosting methods that usually achieve good results when applied in the data mining area did not perform very well (only achieving around 62%).

Another experiment looked at the classification results of one subject when the BLDA classifier was applied to the spelling task. The confusion matrix revealed that the classifier identified 213 characters correctly, and misclassified 38 characters, thus achieving an accuracy of 84.5%. The confusion matrix furthermore showed that the misclassifications were caused mostly by the subject selecting the wrong adjacent column or row on the speller interface. These findings are consistent with related work [18].

The next set of experiments investigated the channel selection for the different subjects. The results revealed that different numbers of channels lead to different accuracy values for the different subjects. Even though the channel selection is different for each subject, however, there are channels that seem to contribute more when comparing all the subjects. The electrode locations that most often lead to a higher accuracy for most of the subjects are P3, P7, O1, O2, and P8. This proves what past literature, e.g., [19] already had concluded that for different subjects different electrode locations influence the performance of the spelling task. In addition, when the selected channel configuration is compared with using all 14 channels, the results show that the classification accuracy using the selected channel configuration is higher. For example, for one subject accuracy values of 75.6% and 90.2% were measured for all 14 channels versus the selected channel configuration, respectively. A possible explanation for this is that in certain cases adding more channels adds noise to the classification task due to redundancy of the data. Thus, this will lead to lower predictive accuracy values due to overfitting.

5 Conclusions

A Brain-Computer Interface (BCI) system is currently the only way that allows disabled people, for example those that cannot move their arms or legs anymore, to control an external device via brain waves. BCI systems have been very successfully applied using speller interfaces. A speller interface allows a disabled person to issue commands to the computer system and thereby controlling it.

This paper investigated two research directions related to a mobile speller interface used in a mobile environment. The first direction evaluated the classification accuracy of the spelling task applying different classification algorithms, and the second direction investigated the influence of different channel selections.

Overall, the research conducted in this paper confirmed that the best classifier for BCI systems including using a mobile interface used in a mobile environment is the BLDA classifier. Furthermore, the experiments revealed that the channel selection is different for different subjects using a mobile interface as is the case in a static environment.

Future work will involve the investigation of deep learning techniques applied to BCI speller systems since deep learning techniques have shown to outperform conventional multi-layer neural networks applied in other areas. However, since deep learning techniques can be time-consuming to train, state-of-the-art technology such as GPUs need to be leveraged.

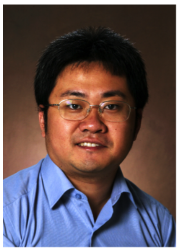
In addition, another interesting topic to investigate is to add a backspace button to the speller. This would allow to investigate genuine spelling mistakes made during the experiments.

Furthermore, it would be interesting to see if there are significant variations influencing the classification accuracy and best channel selection when subjects are performing the same task on different days. In other words, a person can be in different emotional (such as happy or angry) and physical (such as moving fast or slow) states on different days. Though calibration is performed to assure each sensor is placed at the correct location, however the calibration cannot alleviate the emotional and physical changes during different days. Thus, it is worth to investigate how the daily emotional and physical variations affect the channel selection.

Acknowledgements The authors would like to thank Qasem Obeidat for performing the spelling experiments on which this research study is based on.

Biographies

Simone A. Ludwig is an Associate Professor of Computer Science at North Dakota State University, USA. She received her PhD degree and MSc degree with distinction from Brunel University, UK, in 2004 and 2000, respectively. Before starting her academic career she worked several years in the software industry. Her research interests lie in the area of computational intelligence including swarm intelligence, evolutionary computation, neural networks, and fuzzy reasoning with application in data mining and image processing.



Jun Kong received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1998, the M.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 2001, and the Ph.D. degree from the University of Texas at Dallas, Richardson, USA, in 2005, all in computer science. He has been an Associate Professor of computer science with North Dakota State University, Fargo, USA, since 2012. His research and teaching interests include humancomputer interaction, visual languages, software modeling and design, brain-computer interface, and pervasive computing.

References

1. S. Sutton, M. Braren, J. Zubin, E. R. John, Evoked-potential correlates of stimulus uncertainty. Science, 150(3700), 1187-1188, 1965.
2. L. A. Farwell, E. Donchin, Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr. Clin. Neurophysiol., 70(6), 510-523, 1988.
3. C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, How many people are able to control a P300-based brain-computer interface (BCI)? Neurosci. Lett., 462(1), 94-98, 2009.
4. G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. E. Frye, C. K. Hauser, A novel P300-based braincomputer interface stimulus presentation paradigm: moving beyond rows and columns. Clin. Neurophysiol. 121(7), 1109-1120, 2010.
5. R. Fazel-Rezai, K. Abhari, A region-based P300 speller for brain-computer interface. Can. J. Elect. Comput. E., 34(3), 81-85, 2009.
6. Q. Obeidat, T. Campbell, J. Kong, J. The zigzag paradigm: a new P300-based brain computer interface. In Proc. of the 15th ACM on ICMI (pp. 205-212). ACM, 2013.
7. C. T. Lin, Y. C. Chen, T. Y. Huang, T. T. Chiu, L. W. Ko, S. F. Liang, Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver's drowsiness detection and warning. IEEE T. Bio-Med. Eng., 55(5), 1582-1591, 2008.
8. Y. T. Wang, Y. Wang, T. P. Jung, A cell-phone-based brain-computer interface for communication in daily life. J. Neural. Eng., 8(2):025018, 1-5, 2010.
9. M. De Vos, M. Kroesen, R. Emkes, S. Debener, P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. J. Neural. Eng., 11(3):036008, 1-8, 2014.
10. S. Debener, R. Emkes, M. De Vos, M. Bleichner, Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. Scientific Reports, 5:16743, 1-11, 2015.

11. S. Debener, F. Minow, R. Emkes, K. Gandras, M. De Vos, How about taking a lowcost, small, and wireless EEG for a walk? Psychophysiology, 49(11), 1617-1621, 2012.
12. M. De Vos, K. Gandras, S. Debener, Towards a truly mobile auditory braincomputer interface: exploring the P300 to take away. Int. J. Psychophysiol., 91(1), 46-53, 2014.
13. L. D. Liao, C. Y. Chen, I. J. Wang, S. F. Chen, S. Y. Li, B. W. Chen, Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors. J. Neuroeng. Rehabil., 9(1), 1-11, 2012.
14. J. T. Gwin, K. Gramann, S. Makeig, D. P. Ferris, Removal of movement artifact from high-density EEG recorded during walking and running. J. Neurophysiol., 103(6), 3526-3534, 2010.
15. K. Gramann, J. T. Gwin, N. Bigdely-Shamlo, D. P. Ferris, S. Makeig, Visual evoked responses during standing and walking. Front. Hum. Neurosci., 4(202), 1-12, 2010.
16. A.Campbell, T. Choudhury, S. Hu, H. Lu, M. K. Mukerjee, M. Rabbi. NeuroPhone: brain-mobile phone interface using a wireless EEG headset. In Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds (pp. 3-8). ACM, 2010.
17. Q. Obeidat, T. Campbell, J. Kong, Introducing the Edges Paradigm: A P300 Brain-Computer Interface for spelling written words. IEEE Trans. Hum. Mach. Syst., 45(6), 727-738, 2015.
18. Q. Obeidat, Towards improving P300-based brain-computer interfaces: From desktop to mobile, PhD Dissertation, North Dakota State University, 2014.
19. K. A. Colwell, D. B. Ryan, C. S. Throckmorton, E. W. Sellers, L. M. Collins, Channel selection methods for the P300 Speller, Journal of Neuroscience Methods, Volume 232, Pages 6-15, 2014.
20. M. Lang, Investigating the Emotiv EPOC for cognitive control in limited training time, Honors Report, University of Canterbury, 2012.
21. EmotivSystem, Emotiv: Brain-Computer Interface Technology, `http://www.emotiv.com`, [Online; accessed May 5, 2016].
22. F. Sharbrough, G. E. Chatrian, R. P. Lesser, H. Luders, M. Nuwer, and T. W. Picton, American electroencephalographic society guidelines for standard electrode position nomenclature, J. Clin. Neurophysiol. 8, no. 2, 200-202, 1991.
23. H. Serby, E. Yom-Tov, and G. F. Inbar, An improved p300-based braincomputer interface, IEEE T Neur Sys Reh. 13, no. 1, 89-98, 2005.
24. U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, An efficient p300-based brain-computer interface for disabled subjects, J Neurosci Meth. 167, no. 1, 115-125, 2008.
25. F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for EEG-based brain-computer interfaces. J. Neural Eng. 2007.
26. D. J. Krusienski, E. W. Sellers, F. Cabestaing, A comparison of classification techniques for the P300 Speller, Journal of Neural Engineering, vol. 3, no. 4, pp. 299-305, 2006.
27. R. C. Panicker, S. Puthusserypady, and Y. Sun, Adaptation in P300 braincomputer interfaces: a two-classifier cotraining approach, IEEE Transactions on Biomedical Engineering, vol. 57, no. 12, pp. 2927-2935, 2010.
28. H. Mirghasemi, R. Fazel-Rezai, and M. B. Shamsollahi, Analysis of P300 classifiers in brain computer interface speller, in Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'06), pp. 6205-6208, September 2006.
29. F. Nijboer, E. W. Sellers, J. Mellinger, A P300-based brain-computer interface for people with amyotrophic lateral sclerosis, Clinical Neurophysiology, vol. 119, no. 8, pp. 1909-1916, 2008.
30. E. W. Sellers and E. Donchin, A P300-based brain-computer interface: initial tests by ALS patients, Clinical Neurophysiology, vol. 117, no. 3, pp. 538-548, 2006.
31. U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, An efficient P300-based brain-computer interface for disabled subjects, Journal of Neuroscience Methods, vol. 167, no. 1, pp. 115-125, 2008.
32. M. Thulasidas, C. Guan, and J. Wu, Robust classification of EEG signal for brain-computer interface, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 14, no. 1, pp. 24-29, 2006.
33. J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vanderwalle, Least Square Support Vector Machines, World Scientific, Singapore, 2002.

34. G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals, and Systems, vol. 2, no. 4, pp. 303-314, 1989.
35. R. Palaniappan, Brain Computer Interface Design Using Band Powers Extracted During Mental Tasks. In Proceedings of the 2nd International IEEE EMBS Conference on Neural Engineering (NER'05), Arlington, VA, USA, March 2005.
36. J. del R. Millan, J. Mourino, M. Franze, F. Cincotti, M. Varsta, J. Heikkonen, F. Babiloni, A local neural classifier for the recognition of EEG patterns associated to mental tasks. IEEE Transaction of Neural Networks, 2002.
37. K. Nakayaman, K. Inagaki, A Brain Computer Interface Based on Neural Network with Efficient Pre-Processing. In Proceedings of the International Symposium on Intelligent Signal Processing and Communications (ISPACS'06), Yonago, Japan, 2006.
38. E. Haselsteiner, G. Pfurtscheller, Using time-dependent neural networks for EEG classification. IEEE Trans. Rehabil. Eng. 2000.
39. J. R. Millan, J. Mourino, Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project. IEEE Trans. Neural Syst. Rehabil. Eng., 2003.
40. K. Hornik, Approximation Capabilities of Multilayer Feedforward Networks Neural Networks, 4(2), 251-257, 1991.
41. Y. Bengio, Y. LeCun, Scaling learning algorithms towards AI, Large-Scale Kernel Machines, L. Bottou, O. Chapelle, D. DeCoste, J. Weston (eds), MIT Press, 2007.
42. Y. Bengio, O. Delalleau, C. Simard, Decision trees do not generalize to new variations. Computational Intelligence 26(4), 449-467, 2010.
43. Y. Freund and R. E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of Computer and System Sciences, 55(1):119-139, 1997.
44. R. E. Schapire, Boosting the margin: A new explanation for the effectiveness of voting methods. Annals of Statistics, Vol. 26, No. 5, pp. 1651-1686, 1998.
45. C. Seiffert, T. Khoshgoftaar, J. Hulse, and A. Napolitano, RUSBoost: Improving classification performance when training data is skewed, 19th International Conference on Pattern Recognition, 2008.
46. J. Friedman, T. Hastie, and R. Tibshirani, Additive logistic regression: A statistical view of boosting, Annals of Statistics, Vol. 28, No. 2, pp. 337-407, 2000.
47. D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, J. R. Wolpaw. Toward enhanced P300 speller performance? J Neurosci Methods, 167(1):15-21, 2008.
48. M. Schroeder, T. N. Lal, T. Hinterberger, M. Bogdan, N. J. Hill, N. Birbaumer. Robust EEG channel selection across subjects for braincomputer interfaces. EURASIP J Adv Signal Process, 2005(19):3103-12, 2005.
49. A. Rakotomamonjy, V. Guigue. BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller. IEEE Trans Biomed Eng, 55(3):1147-54, 2008.
50. H. Cecotti, B. Rivet, M. Congedo, C. Jutten. A robust sensor selection method for P300 braincomputer interfaces. J Neural Eng, 8(1):1-21, 2011.
51. T. N. Lal, M. Schroeder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer. Support vector channel selection in BCI. IEEE Trans Biomed Eng, 51(6):1003-10, 2004.
52. T. Lan, D. Erdogmus, A. Adami, S. Mathan, M. Pavel. Channel selection and feature projection for cognitive load estimation using ambulatory EEG. Comput Intell Neurosci, 2007:74895, 2007.
53. J. Jin, B. Z. Allison, C. Brunner, B. Wang, X. Wang, J. Zhang. P300 Chinese input system based on Bayesian LDA. Biomed Tech (Berl), 55:5-18, 2010.