Learning Discriminative Spatiospectral Features of ERPs for Accurate Brain–Computer Interfaces

Berdakh Abibullaev, *Member, IEEE*, and Amin Zollanvari, *Member, IEEE*

***Abstract*—Constructing accurate predictive models is at the heart of brain–computer interfaces (BCIs) because these models can ultimately translate brain activities into communication and control commands. The majority of the previous work in BCI use spatial, temporal, or spatiotemporal features of event-related potentials (ERPs). In this study, we examined the discriminatory effect of their spatiospectral features to capture the most relevant set of neural activities from electroencephalographic recordings that represent users' mental intent. In this regard, we model ERP waveforms using a sum of sinusoids with unknown amplitudes, frequencies, and phases. The effect of this signal modeling step is to represent high-dimensional ERP waveforms in a substantially lower dimensionality space, which includes their dominant power spectral contents. We found that the most discriminative frequencies for accurate decoding of visual attention modulated ERPs lie in a spectral range less than 6.4 Hz. This was empirically verified by treating dominant frequency contents of ERP waveforms as feature vectors in the state-of-the-art machine learning techniques used herein. The constructed predictive models achieved remarkable performance, which for some subjects was as high as 94% as measured by the area under curve. Using these spectral contents, we further studied the discriminatory effect of each channel and proposed an efficient strategy to choose subject-specific subsets of channels that generally led to classifiers with comparable performance.**

***Index Terms*—Brain-computer interfaces, ERPs, P300, EEG, signal processing, machine learning.**

I. Introduction

BRAIN-COMPUTER Interface (BCI) systems have attracted extensive attention in the past decade, because of their potential in improving human life, especially for those who are affected by severe motor disorders such as the amyotrophic lateral sclerosis (ALS) [1]–[3]. ALS is a neurological condition which affects neurons responsible for voluntary movements [4], [5]. As a result, patients rapidly lose the ability to move their arms, legs and face muscles, and gradually become totally locked-in and unable to communicate, although their brain is fully functional. The only means of communication for those patients remains to be a BCI speller [6]–[9].

A BCI speller is a system that measures brain activity of an individual and translates it into communication commands, thereby bypassing the natural link between the central and peripheral nervous system [10]. BCI spellers mainly use event-related potentials (ERPs)—the components of EEG signals—to infer brain activity patterns [11], [12]. These ERPs are the neural responses time-locked to an external sensory stimulus event such as a visual [13], [14], auditory [15], haptic or internal event associated with a motor task execution [16].

A typical ERP waveform is composed of several subcomponents, each of which related to a positive or a negative voltage deflection correlating with a distinct neural process. In the BCI speller context, one is particularly interested in decoding P300 components ERPs that reflect processing of a physical, auditory, and visual stimulus. P300 is a positive wave evoked 300-400 milliseconds (ms) after the stimulus onset when a user mentally attends to a rare target sensory event. Most P300 based BCI spellers still use the row-column paradigm introduced by Farwell and Donchin [17], [18] in which a rectangular grid of alphanumeric characters (arranged in rows and columns) are intensified for visual stimulus presentation. Depending on user's mental intent, a BCI system infers the presence of a P300 wave after row/column stimulus intensification and selects characters successively for spelling and communication. After experimental design, which is extensively discussed in [19]–[21], perhaps the most salient aspects of BCI are: 1) development of signal processing algorithms to extract sensitive signatures of ERPs with which human mental intents are best encoded [22], [23]; and 2) applications of machine learning techniques for robust decoding of extracted signatures [24]–[27].

From a machine learning point of view, the accurate decoding of ERPs is challenging due to the poor signal-to-noise ratio, the presence of non-task relevant features, enormous variability, and high-dimensionality (i.e., the number of channels $\times$ the number of time points [28], [29]). Significant variation of ERPs within and between sessions and subjects, or even across trials renders most BCI technologies inaccurate [30], [31]. Such variability may be caused by the neurophysiological state of a user such as the level of attention, drowsiness, tiredness, stress, and even the level of caffeine or medicine intake [32], [33]. At the same time, specific experimental setups such as the difficulty and task relevance of a stimulus event, its emotional meaning, different

Manuscript received June 22, 2018; revised September 27, 2018; accepted November 17, 2018. Date of publication January 16, 2019; date of current version September 4, 2019. This work was partially supported by Nazarbayev University Faculty Development under Grant SOE2018008. *(Corresponding author: Berdakh Abibullaev.)*

B. Abibullaev is with the Department of Robotics and Mechatronics, Nazarbayev University, Astana 010000, Kazakhstan (e-mail: berdakh.abibullaev@nu.edu.kz).

A. Zollanvari is with the Department of Electrical and Computer Engineering, Nazarbayev University, Astana 010000, Kazakhstan (e-mail: amin.zollanvari@nu.edu.kz).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author.

Digital Object Identifier 10.1109/JBHI.2018.2883458

presentation rate, as well as inter-stimulus interval, including stimulus intensity and matrix size can cause great variation in ERPs' feature distributions. This can consequently lead to the failure of a previously functioning BCI system.

The advent of BCI technologies in the past few decades has brought a host of data-driven techniques to extract discriminative features from P300 waveforms [12], [25], [28], [34]–[38]. For example, Mowla *et al.* used wavelet transforms to extract invariant temporal features by taking into account a latency jitter and variation in P300 waves and showed an improved BCI performance [36]. Bostanov has shown that decomposing P300 waves via t-statistics based continuous-wavelet transforms extracts reliable spectral and temporal ERP features [39]. In another study, Yoon *et al.* proposed three features that composed of the raw samples and the amplitude and negative area of P300, and achieved an improved performance using a multiple kernel classifier [35]. The design of spatial feature extractors has been widely studied as well. This includes spatial whitening [30], common spatial-temporal patterns [40], [41], nonlinear principal-component analysis [42] and advanced spatiotemporal beamforming techniques [43]–[45] in which every method treats variability and noise in a particular way.

All these ERP based BCI studies generally follow a common methodology: 1) conduct an ERP experiment to acquire EEG data by placing an electrode cap on the scalp of the subject; 2) pre-process the collected data to remove the artifacts and enhance the signal-to-noise ratio—for instance, restricting frequency contents to a range of (0.1 Hz–30 Hz) as suggested by [30], [46]; 3) extract raw temporal features with a predefined window size of 600 ms or 1000 ms followed by a customized spatiotemporal feature extraction method such as wavelets, spatial filters, beamformers, etc.; and 4) use extracted features in a classification rule [30], [31], [35], [47]–[50].

The majority of the previous studies use discriminative spatial, temporal, or spatiotemporal features of ERPs. Some studies attempted to characterize spectral information of P300 waves by identifying an optimal cut-off frequency range for data preprocessing [30], [46]. Nevertheless, even in these studies, after the bandpass filtering stage, temporal features have been essentially used for classification purposes. At the same time, although the wavelet-based methods can extract combined time and frequency features [39], [51], their computational load and difficulty in identifying an optimal wavelet basis function render their online BCIs applications limited.

In this investigation, we examine the discriminatory effect of spatiospectral features to identify the most relevant set of neural activities from ERPs that best represent users' mental intent. In other words, given ERPs from both target and non-target stimuli, we aim to extract discriminative spatiospectral patterns that are useful for improving the accuracy of a BCI speller system. In this regard, we first model EEG time-series using a mathematical form, which is the sum of sinusoids with unknown amplitudes, frequencies, and phases. A nonlinear least square technique is then used to estimate the unknown parameters of this mathematical form [52]. The effect of this signal processing step is to represent the high-dimensional ERP waveforms in a substantially lower-dimensional space. We then use these low-dimensional feature vectors in the state-of-the art machine learning techniques to construct accurate predictive models of users' mental intent. We further study the discriminatory effect of each channel and propose an efficient strategy to choose subject-specific subset of channels that generally leads to a lower number of channels, while keeping the classifier performance comparable. The focus of our study is warranted because developing accurate predictive models with a discriminative set of neural features and a minimal number of sensors empowers ERP-based BCI research and can potentially lead to accurate and robust neural decoding systems.

II. Material and Methods

A. Participants

Seven healthy young adults (age range 22 ± 3 years) with no history of neurological, physical, or psychiatric illness participated in this study. All the participants were naive BCI users who had not participated in any related experiments before. Informed consents were received from all participants. The study has been approved by Institutional Research Ethics Committee of Nazarbayev University.

B. Experimental Setup

1) Electroencephalography: Scalp EEG was recorded using a 16-channel, active Ag/AgCl electrodes (g.USBamp, g.LADYbird, Guger Technologies OG, Schiedlberg, Austria) with a sampling frequency = 256 Hz, resolution = 16-bit, dynamic range = ±3:2768 mV, and bandwidth = 0–1000 Hz. The EEG electrodes were positioned according to the International 10–20 system. The right earlobe of participants were used for a ground electrode, whereas the FCz location were used for a reference electrode.

2) Data Acquisition: We implemented the Farwell & Donchin style speller [17], using a 6×6 grid of alphanumeric characters, presented via an LCD monitor as shown in Fig. 1(A). Participants were seated in a comfortable chair facing an LCD monitor with the distance about 60 cm in between, in a quiet office. Each participant performed a single session during which their EEG signals were measured.

Before starting the calibration session, each participant was instructed that the target character on the visual stimuli matrix (see Fig. 1(A)) is intensified in green for two seconds after which columns and rows begin to flash. Once the session started, participants were asked to concentrate on the target character, attending to the number of times it flashed. Additionally, we asked each participant to count the number of times the target letter flashed to ensure their attentiveness.

The total number of target letters that a participant had to attend was equal to five. These targets were presented in five consequent sequences with an inter-sequence duration (ISD) of two seconds. In total, five repetitions of each target character were made where one repetition (or trial) consisted of a complete set of 12 random flashes of every row and column (six rows and six columns). Inter-stimulus interval (ISI) was set to 150 ms; likewise, the stimulus duration was 100 ms (i.e., the time length a row/col is highlighted). The minimum time between the same target letter highlights was set to 600 ms, also denoted as a

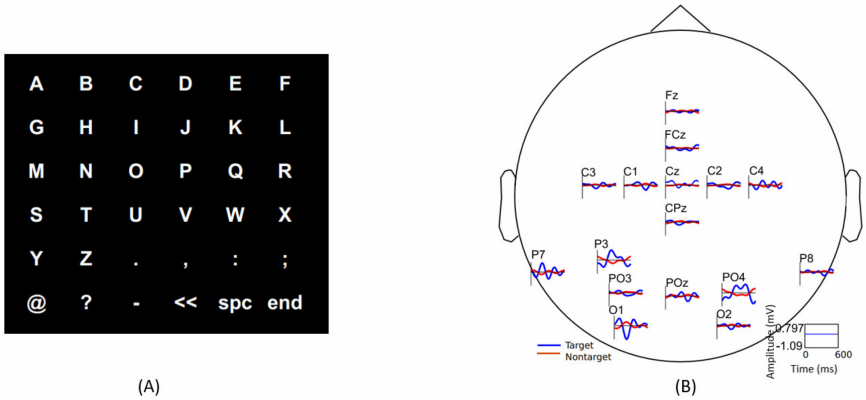


Fig. 1. (A) Visual stimuli character matrix used to evoke ERPs. (B) The electrode montage used in the current study shown along with target and nontarget ERPs. Target ERPs are defined as events and characters that a user wants to communicate, and any other events are defined as a non-target ERPs.

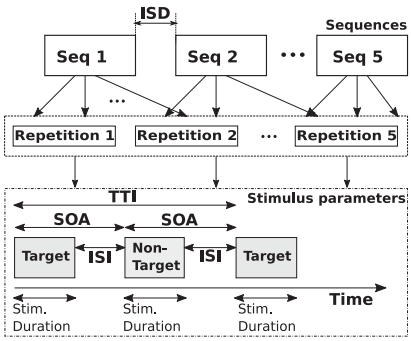


Fig. 2. The parameters of of the visual stimulus presentations during the ERP experiment. ISD: inter-sequence duration, ISI: inter-stimulus interval, TTI: target-to-target interval, SOA: stimulus onset asynchrony.

TABLE I
NUMBER OF TRAINING TRIALS FOR EACH OF THE SEVEN AVAILABLE SUBJECTS FOR TARGET AND NON-TARGET CLASSES. WE TREAT EACH TRIAL AS A SINGLE OBSERVATION IN D_{TRAIN}

Subject	I	II	III	IV	V	VI	VII
Target	221	231	220	221	223	238	435
Non-target	677	667	659	650	661	1460	2547

target-to-target (TTI) interval. The stimulus onset asynchrony (SOA) was set to 250 ms. The SOA is the time period between the start of one stimulus event, and the start of the next event.

For one of the subjects, we collected data from ten target letters as he reported his attention shift during the experimental session and agreed for more sequences to complete. This can be observed in Table I where the number of targets and non-target samples for subject VII is remarkably larger than others.

C. Pre-Processing

The following steps summarize the initial signal processing steps applied for a single trial detection of ERPs:

- *Segmentation:* Continuous EEG data were segmented into target and non-target trials with 600 ms duration after the stimulus event onset. Further, for a signal modeling (described in Section II-D), we only considered observations in the interval of 200 to 500 ms where P300 waveforms are prominent.
- *Detrending:* Arbitrary offsets in data were removed by subtracting the total mean from each channel. This step ensures the removal of noise in the form of slow drifts which might occur due to sweating and a poor sensor-to-head contact [53].
- *Bad trial removal:* EEG data trials were artifact edited using a statistical thresholding procedure to eliminate bad trials associated with egregious movement artifacts. The criteria to detect bad trials was based on calculating the mean absolute value per trial and excluding trials with values greater than 3-standard deviations over the median trial.
- *Bad channel removal:* All channels were analyzed across trials to identify electrodes which are corrupted with excessive noise arising from improper connection to the scalp of a participant. Channels with excessively high power were identified by computing: 1) the total power for each channel; 2) the mean channel power; and 3) the variance of channel power over all epochs. We removed any channels with power more than three standard deviation and replaced them with a common-averaged reference channel.
- *Spatial Filtering:* We employed a spatial whitening filter to minimize noise due to source mixing and volume conduction. The whitening filter uses linear re-weighting of the electrodes and maps raw electrode readings to a new space where the sensors are uncorrelated and have unit power.
- *Spectral Filtering:* EEG data was band-pass filtered between 0.1–15 Hz range. In this regard, the signal was first Fourier transformed and then a weighting was applied to suppress and remove unwanted frequency contents outside the frequency range of interest. Finally, the weighted signal wave was inverse Fourier transformed to obtain the filtered signal.

D. Signal Modeling

Let the set of pairs $D_{\text{train}} = \{(\boldsymbol{z}_1, y_1), \ldots, (\boldsymbol{z}_n, y_n)\}$ denote n trials of EEG recordings for each subject where y_i is the scaler class variable (target or non target) and $\boldsymbol{z}_i \in \mathbb{R}^{(pc)\times 1}$ is

an instance of EEG observation over c channels and p time points defined as

$$\boldsymbol{z}_i = [\boldsymbol{x}_{i1}^T, \boldsymbol{x}_{i2}^T, \ldots, \boldsymbol{x}_{ic}^T]^T, i = 1, \ldots, n, \tag{1}$$

with $\boldsymbol{x}_{ij} \in \mathbb{R}^{p\times 1}, j = 1, \ldots, c$, and T denoting the transpose operator. Table I shows the number of training trials for each of the seven subjects who participated in this study. Furthermore, in our case, $p = \lfloor 256\text{ Hz} \times (500 - 200)\text{ms} \rfloor = 77$ time points where $\lfloor . \rfloor$ denotes the floor function. The purpose of signal modeling is to represent $\boldsymbol{x}_{ij}$, which is the ith observation vector from the jth channel, by an estimated signal vector $\boldsymbol{s}_{ij}$ that lives in a k-dimensional subspace such that $k \ll p$. In other words, this estimation problem can be seen as a dimensionality reduction or, equivalently, a feature extraction operation to train our predictive model. As the signal model, we choose the sum of sinusoids with unknown frequencies, phases, and amplitudes [54]. That is to say,

$$\begin{aligned} s_{ij}[l] &= \sum_{k=1}^{M} A_k \cos\left(2\pi f_k l + \phi_k\right) \\ &= \sum_{k=1}^{M} \xi_{1k} \cos\left(2\pi f_k l\right) + \xi_{2k} \sin\left(2\pi f_k l\right), l = 0, \ldots, p-1 \end{aligned} \tag{2}$$

where $s_{ij}[l]$ denotes the l^{th} element of $\boldsymbol{s}_{ij}$, M is the model order to be determined, $\xi_{1k} = A_k \cos(\phi_k)$, $\xi_{2k} = -A_k \sin(\phi_k)$, and $0 < f_k < f_{k+1} < 0.5, k = 1, \ldots, M$. Note that all parameters used in model (2) are functions of i and j because they can vary for each trial in each channel; however, to ease the notation, we have already omitted indices i and j from $A_{i,j,k}$, $\phi_{i,j,k}$, $f_{i,j,k}$, $\xi_{1,i,j,k}$, and $\xi_{2,i,j,k}$ to write (2). The sinusoidal model (2) is linear in terms of $\xi_{hk}, h = 1, 2$, and nonlinear in terms of f_k. Without making any distributional assumption on the observations, we follow the nonlinear least squares approach to estimate the $3M$ unknown parameters on each observation segment $\boldsymbol{x}_{ij}$ [52]. However, using separability property of parameters [52], we reduce the computational complexity by solving a M-dimensional nonlinear least squares problem and a $2M$-dimensional linear one to estimate f_k's and ξ_{hk}'s, respectively.

In this regard, we first estimate f_k's by conducting a grid search over $\{0.005, 0.01, 0.015, \ldots, f_{\max}\}^M$ where $f_{\max} = 0.05 \approx \frac{\text{upper cutoff freq. of band pass filter}}{\text{sampling frequency}} = \frac{12}{256}$ to minimize the least squares quadratic risk or, equivalently, by (cf. Section 3.5.4 in [52])

$$\hat{\boldsymbol{f}}_{ij} \triangleq [\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_M] = \underset{f_1,\ldots,f_M}{\operatorname{argmax}} \ \boldsymbol{x}_{ij}^T \boldsymbol{H} \left(\boldsymbol{H}^T \boldsymbol{H}\right)^{-1} \boldsymbol{H}^T \boldsymbol{x}_{ij}, \tag{3}$$

where $\boldsymbol{H}$ is the observation matrix with elements determined as

$$\boldsymbol{H}[r, s] = \begin{cases} \cos\left(2\pi f_{\lceil s/2 \rceil}(r-1)\right), \text{for } s \text{ odd}, \\ \sin\left(2\pi f_{\lceil s/2 \rceil}(r-1)\right), \text{for } s \text{ even}, \end{cases} \tag{4}$$

for $r = 1, \ldots, p$, $s = 1, \ldots, 2M$ and $\lceil a \rceil$ denoting the least integer greater or equal to a. Once $\hat{f}_i$'s are obtained, they are replaced in (4) to obtain $\hat{\boldsymbol{H}}$. We can then estimate ξ_{hk} as

$$\hat{\boldsymbol{\xi}}_{ij} \triangleq [\hat{\xi}_{11}, \hat{\xi}_{21}, \hat{\xi}_{12}, \ldots, \hat{\xi}_{2M}]^T = \left(\hat{\boldsymbol{H}}^T \hat{\boldsymbol{H}}\right)^{-1} \hat{\boldsymbol{H}}^T \boldsymbol{x}_{ij}. \tag{5}$$

As a result, the set of pairs $D_{\text{train}} = \{(\boldsymbol{z}_1, y_1), \ldots, (\boldsymbol{z}_n, y_n)\}$ with $\boldsymbol{z}_i \in \mathbb{R}^{(pc)\times 1}$ is then converted into the set of pairs

$$S_{\text{train}} = \{(\hat{\boldsymbol{\theta}}_1, y_1), \ldots, (\hat{\boldsymbol{\theta}}_n, y_n)\}, \tag{6}$$

where $\hat{\boldsymbol{\theta}}_i \in \mathbb{R}^{(3Mc)\times 1}$, $3M \ll p$, and

$$\hat{\boldsymbol{\theta}}_i = [\hat{\boldsymbol{\theta}}_{i1}^T, \hat{\boldsymbol{\theta}}_{i2}^T, \ldots, \hat{\boldsymbol{\theta}}_{ic}^T]^T, i = 1, \ldots, n, \tag{7}$$

with $\hat{\boldsymbol{\theta}}_{ij}$ being the estimated signal parameter vector defined as

$$\hat{\boldsymbol{\theta}}_{ij} = [\hat{\boldsymbol{\xi}}_{ij}^T, \hat{\boldsymbol{f}}_{ij}^T]^T, i = 1, \ldots, n, j = 1, \ldots, c, \tag{8}$$

and $\hat{\boldsymbol{f}}_{ij}$ and $\hat{\boldsymbol{\xi}}_{ij}$ are obtained from (3) and (5), respectively.

E. Model Order and Classification Rule Selection

The purpose of model selection in our analysis is two folds: 1) selecting the model order M used in (2); and 2) choosing the classification rule. In this regard, we use the set of signal parameter vectors S_{train} given in (6) and choose the model order M that along with a classification rule leads to the best accuracy measured by the Area Under receiver operating Curve (AUC) [55]. Previous research have shown and discussed a number of desirable properties of using area under the receiver operating characteristic curve (AUC) as a measure of classification performance compared to the overall classification accuracy [56], [57]. In particular, AUC is a measure that is insensitive to [56]–[58]: (1) decision threshold; (2) *a priori* class distribution; and (3) cost of different types of misclassification. In order to evaluate the AUC on the set of training observations, we use a 10-fold cross-validation. In K-fold cross-validation, one randomly divides the set of training data to K subsets, successively holds out these subsets from training data, construct the classifier on the reduced set of training data, and determine the error rate of the constructed (surrogate) classifiers on the held-out subset [59].

As the classification rule, we use five models: (1) Logistic Regression with L_2-Ridge penalty (LRR) [60] and regularization parameter $\lambda \in \{10^{-8}, 10^{-4}, 1\}$; (2) Naive Bayes with Laplace estimator $\mu \in \{0.1, 1, 10\}$ of conditional probabilities (NB; see [61, p. 99]); (3) Tree-Augmented Naive Bayes with Laplace estimator $\mu \in \{0.1, 1, 10\}$ (TAN; [62]); (4) Support Vector Machines with Gaussian kernel (also known as Radial Basis Function) with kernel variance $\sigma^2 \in \{0.01, 0.1, 1, 10, 100\}$ and a complexity parameter $C = \{1, 10, 50, 100, 200, 500, 1000\}$ (SVM-RBF; see [63, p. 190–191]); and (5) logistic regression with L_1-Least Absolute Shrinkage and Selection Operator penalty (LRlasso; [64]) and regularizaion parameter (λ) path determined from glmnet package in R [65]. We have chosen these classifiers as they have shown remarkable performance in a wide variety of applications. For example consider the comparison between Naive Bayes, decision trees, and SVM, which was conducted in [66], [67]. In particular and interestingly, the result of [66] shows that although SVM and Naive Bayes are comparable to decision trees in terms of predictive accuracy,

TABLE II

AUC OF CONSTRUCTED CLASSIFIERS FOR DIFFERENT MODEL ORDER M AND THE ESTIMATED OPTIMUM VALUE OF THE FREE PARAMETER EVALUATED USING 10-FOLD CROSS-VALIDATION. THE CLASSIFIER USES $3Mc$ FEATURES EXTRACTED BY APPLYING SIGNAL MODELING (2)–(8) ON OBSERVATIONS COLLECTED FROM ALL CHANNELS. THE OPTIMUM CONSTRUCTED MODEL FOR EACH SUBJECT IS SHOWN IN BOLD

Subj.	M	Classifier	AUC	Subj.	M	Classifier	AUC	Subj.	M	Classifier	AUC
I	1	LRR ($\lambda = 10^{-8}$)	67.8%	IV	1	LRR ($\lambda = 1$)	73.2%	VII	1	LRR ($\lambda = 10^{-8}$)	44.3%
I	**2**	**LRR ($\boldsymbol{\lambda = 10^{-8}}$)**	**81.6%**	IV	**2**	**LRR ($\boldsymbol{\lambda = 10^{-8}}$)**	**79.5%**	VII	2	LRR ($\lambda = 10^{-8}$)	74.8%
I	3	LRR ($\lambda = 10^{-8}$)	80.8%	IV	3	LRR ($\lambda = 1$)	78.6%	VII	**3**	**LRR ($\boldsymbol{\lambda = 1}$)**	**79.8%**
I	1	NB ($\mu = 1$)	49.6%	IV	1	NB ($\mu = 1$)	55%	VII	1	NB ($\mu = 1$)	60.2%
I	2	NB ($\mu = 1$)	49.6%	IV	2	NB ($\mu = 10$)	60.9%	VII	2	NB ($\mu = 1$)	67.2%
I	3	NB ($\mu = 1$)	47.9%	IV	3	NB ($\mu = 1$)	59.7%	VII	3	NB ($\mu = 0.1$)	70.5%
I	1	TAN ($\mu = 1$)	49.6%	IV	1	TAN ($\mu = 0.1$)	55.3%	VII	1	TAN ($\mu = 0.1$)	69.2%
I	2	TAN ($\mu = 1$)	49.6%	IV	2	TAN ($\mu = 10$)	61.4%	VII	2	TAN ($\mu = 10$)	69%
I	3	TAN ($\mu = 1$)	47.9%	IV	3	TAN ($\mu = 10$)	59.2%	VII	3	TAN ($\mu = 0.1$)	70.4%
I	1	SVM-RBF ($\sigma^2 = 10, C = 100$)	51%	IV	1	SVM-RBF ($\sigma^2 = 10, C = 500$)	62.6%	VII	1	SVM-RBF ($\sigma^2 = 1, C = 1000$)	61.5%
I	2	SVM-RBF ($\sigma^2 = 10, C = 1000$)	60.6%	IV	2	SVM-RBF ($\sigma^2 = 10, C = 50$)	66.8%	VII	2	SVM-RBF ($\sigma^2 = 10, C = 50$)	61.1%
I	3	SVM-RBF ($\sigma^2 = 10, C = 1000$)	55.6%	IV	3	SVM-RBF ($\sigma^2 = 10, C = 100$)	65.1%	VII	3	SVM-RBF ($\sigma^2 = 10, C = 200$)	61.7%
I	1	LRlasso ($\lambda = 7 \times 10^{-6}$)	50.5%	IV	1	LRlasso ($\lambda = 7.3 \times 10^{-3}$)	58.7%	VII	1	LRlasso ($\lambda = 3 \times 10^{-4}$)	60.6%
I	2	LRlasso ($\lambda = 9 \times 10^{-6}$)	66.8%	IV	2	LRlasso ($\lambda = 1.5 \times 10^{-3}$)	65.1%	VII	2	LRlasso ($\lambda = 2 \times 10^{-4}$)	55.8%
I	3	LRlasso ($\lambda = 10^{-5}$)	63.2%	IV	3	LRlasso ($\lambda = 3.9 \times 10^{-3}$)	68.4%	VII	3	LRlasso ($\lambda = 1.1 \times 10^{-3}$)	59.7%
II	1	LRR ($\lambda = 10^{-8}$)	76.1%	V	1	LRR ($\lambda = 1$)	82.3%				
II	2	LRR ($\lambda = 10^{-4}$)	87.5%	V	**2**	**LRR ($\boldsymbol{\lambda = 1}$)**	**94%**				
II	**3**	**LRR ($\boldsymbol{\lambda = 10^{-4}}$)**	**90.3%**	V	3	LRR ($\lambda = 1$)	67.8%				
II	1	NB ($\mu = 1$)	49.6%	V	1	NB ($\mu = 10$)	69.3%				
II	2	NB ($\mu = 1$)	49.6%	V	2	NB ($\mu = 1$)	85.6%				
II	3	NB ($\mu = 0.1$)	51.5%	V	3	NB ($\mu = 10$)	87.9%				
II	1	TAN ($\mu = 1$)	49.6%	V	1	TAN ($\mu = 10$)	70.1%				
II	2	TAN ($\mu = 1$)	49.6%	V	2	TAN ($\mu = 1$)	86%				
II	3	TAN ($\mu = 0.1$)	52.1%	V	3	TAN ($\mu = 10$)	88.2%				
II	1	SVM-RBF ($\sigma^2 = 1, C = 1000$)	61.2%	V	1	SVM-RBF ($\sigma^2 = 1, C = 20$)	70.4%				
II	2	SVM-RBF ($\sigma^2 = 10, C = 1000$)	70.3%	V	2	SVM-RBF ($\sigma^2 = 10, C = 10$)	81.4%				
II	3	SVM-RBF ($\sigma^2 = 10, C = 500$)	65%	V	3	SVM-RBF ($\sigma^2 = 10, C = 10$)	81%				
II	1	LRlasso ($\lambda = 6 \times 10^{-6}$)	63.6%	V	1	LRlasso ($\lambda = 8.9 \times 10^{-3}$)	67.5%				
II	2	LRlasso ($\lambda = 10^{-5}$)	81.5%	V	2	LRlasso ($\lambda = 3.6 \times 10^{-3}$)	80.4%				
II	3	LRlasso ($\lambda = 2 \times 10^{-5}$)	81.6%	V	3	LRlasso ($\lambda = 9.4 \times 10^{-3}$)	80.9%				
III	1	LRR ($\lambda = 1$)	89.6%	VI	**1**	**LRR ($\boldsymbol{\lambda = 1}$)**	**85.9%**				
III	2	LRR ($\lambda = 1$)	91.2%	VI	2	LRR ($\lambda = 1$)	81.7%				
III	**3**	**LRR ($\boldsymbol{\lambda = 1}$)**	**91.6%**	VI	3	LRR ($\lambda = 1$)	84.3%				
III	1	NB ($\mu = 10$)	72.3%	VI	1	NB ($\mu = 0.1$)	71.7%				
III	2	NB ($\mu = 10$)	80.1%	VI	2	NB ($\mu = 10$)	72%				
III	3	NB ($\mu = 10$)	83.5%	VI	3	NB ($\mu = 1$)	73.8%				
III	1	TAN ($\mu = 1$)	74%	VI	1	TAN ($\mu = 1$)	75.2%				
III	2	TAN ($\mu = 10$)	80.4%	VI	2	TAN ($\mu = 10$)	72.3%				
III	3	TAN ($\mu = 10$)	83.7%	VI	3	TAN ($\mu = 1$)	75.3%				
III	1	SVM-RBF ($\sigma^2 = 10, C = 100$)	77.3%	VI	1	SVM-RBF ($\sigma^2 = 10, C = 1000$)	66.8%				
III	2	SVM-RBF ($\sigma^2 = 10, C = 100$)	79.6%	VI	2	SVM-RBF ($\sigma^2 = 10, C = 50$)	64.6%				
III	3	SVM-RBF ($\sigma^2 = 100, C = 200$)	83.4%	VI	3	SVM-RBF ($\sigma^2 = 10, C = 50$)	71.2%				
III	1	LRlasso ($\lambda = 9 \times 10^{-4}$)	77.8%	VI	1	LRlasso ($\lambda = 4 \times 10^{-4}$)	68.9%				
III	2	LRlasso ($\lambda = 6.2 \times 10^{-3}$)	84.5%	VI	2	LRlasso ($\lambda = 6.1 \times 10^{-3}$)	55.6%				
III	3	LRlasso ($\lambda = 8.4 \times 10^{-3}$)	86.1%	VI	3	LRlasso ($\lambda = 4.8 \times 10^{-3}$)	60%				

they are significantly better than decision trees in terms of AUC. The authors of [68] concluded that in general the SVM-RBF is superior to linear SVM. Experiments conducted in [62], [69] show that TAN generally outperforms NB. The results of [70] show that SVM-RBF is generally more effective in classification but the experiments in [71] suggest that LRlasso is more accurate than SVM (with linear and polynomial kernel).

Table II shows the AUC of these constructed classifiers for model orders $M = 1, 2, 3$ with the estimated optimum value of their free parameters (in terms of maximizing AUC) evaluated using 10-fold cross-validation for each subject. Interestingly, LRR shows a remarkably good performance in all cases and outperforms other classifiers in terms of achieving a higher AUC. Note that in a highly imbalanced design, using the accuracy to judge the predictive capacity of a classifier could be misleading [72]. For example, as seen in Table II, SVM-RBF, although generally is a good classifier, in a number of cases in our application has led to a low AUC because it was highly biased towards the majority class (and as a result a sensitivity of 0% but an accuracy >70%).

F. Channel Ranking

We conducted a channel-wise analysis to examine and rank the predictive capacity of each channel in discriminating target and non-target trials across all subjects. This analysis is warranted because: 1) it can shed light on the presence of channels that carry the most discriminative contents; and 2) it can potentially lead to a lower number of channels, while keeping the classifier performance comparable. Using the BCI system with few channels would also minimize the setup time for the EEG cap and make it more portable, comfortable, and inexpensive for the end user [73], [74].

In this regard, we constructed a LRR classifier using $3M$ features extracted using (2)–(8) from observations $\boldsymbol{x}_{ic}$ defined in (1). Table III shows the AUC of the constructed channel-specific classifiers, their average AUC across subjects, and the channel rank based on the average AUC. We also report the sensitivity of each channel in detecting positive responses. As discussed before, a sensitivity of 0% here implies the learning rule is completely biased towards the non-target class (i.e., the majority class here). The analysis suggests that channels Cz and C2 (ranked 1st and 2nd) and channels POz and PO4 (ranked 15th and 16th) have the highest and lowest discriminatory effects, respectively.

III. RESULTS

A. Classifier Validation

We have examined the performance of our constructed LRR models for each subject in predicting the class variable in a set of independent test dataset D_{test} collected from the same subjects. Table IV shows the number of test observations for each subject.

TABLE III
AUC AND SENSITIVITY (IDENTIFIED AS "SENS.") OF LRR CLASSIFIERS CONSTRUCTED FOR EACH CHANNEL EVALUATED USING 10-FOLD CROSS-VALIDATION. THE CLASSIFIER USES $3M$ FEATURES EXTRACTED BY APPLYING SIGNAL MODELING (2)–(8) ON OBSERVATIONS COLLECTED FROM EACH CHANNEL INDIVIDUALLY. CHANNELS WITH A SENSITIVITY $>1\%$ ARE SHOWN IN BOLD

Channel →	**FCz**, AUC (Sens.)	**C3**, AUC (Sens.)	**C1**, AUC (Sens.)	**C2**, AUC (Sens.)	**Cz**, AUC (Sens.)	**C4**, AUC (Sens.)	**CPz**, AUC (Sens.)	**P7**, AUC (Sens.)
Subject I	**61.7% (4%)**	54.3% (0%)	54.7% (0%)	56.6% (0%)	**58.4% (1.4%)**	57.3% (0%)	53.9% (0%)	55% (0%)
Subject II	57.7% (0%)	50.0 (0%)	**58.2% (2.2%)**	**62.8% (6.9%)**	59% (0%)	**62.2% (2.6%)**	51.5 (0%)	50.8% (0%)
Subject III	**69.6% (15.4%)**	**62.9% (4.1%)**	**64.3% (5.9%)**	**83.9% (48.1%)**	**82.3% (46.8%)**	**62.6% (1.3%)**	**67% (10%)**	**67.6% (11.3%)**
Subject IV	55.4% (0%)	58.1% (0%)	58.2% (0%)	**68.6% (14.9%)**	57.5% (0%)	55.1% (0%)	**58.6% (1.4%)**	**63.9% (1.8%)**
Subject V	58.1% (0%)	**65.9% (8.9%)**	**82.6% (45.3%)**	**82.5% (47.0%)**	**81.6% (43.0%)**	**69.9% (12.1%)**	**66.1% (6.3%)**	**71.1% (14.8%)**
Subject VI	53.6% (0%)	**69.8% (2.1%)**	62.3% (0%)	65.8% (0%)	**72.3% (6.7%)**	67.8% (0%)	**72% (4.2%)**	**69.6% (1.6%)**
Subject VII	63.4% (0%)	58.4% (0%)	57.1% (0%)	**68.6% (2.0%)**	**68.8% (2.2%)**	65.2% ($<1\%$)	64% (0%)	68.1% ($<1\%$)
Average AUC Across Subjects	59.9%	59.9%	62.5%	69.8%	68.5%	62.8%	61.8%	63.7%
Rank Based on Average AUC	13	14	7	1	2	6	10	5

Channel →	**P3**, AUC (Sens.)	**P4**, AUC (Sens.)	**P8**, AUC (Sens.)	**PO3**, AUC (Sens.)	**POz**, AUC (Sens.)	PO4, AUC (Sens.)	O1, AUC (Sens.)	O2, AUC (Sens.)
Subject I	55.9% (0%)	55.4% (0%)	57.5% (0%)	55.5% (0%)	54% (0%)	**60.6% (3.1%)**	53.5% (0%)	**60.6% (1.4%)**
Subject II	50.4% (0%)	53.1% (0%)	52% (0%)	**53% (1.2%)**	54% (0%)	54.2% (0%)	**62% (3.4%)**	59.6% (0%)
Subject III	**65.6% (7.7%)**	**60.8% (3.1%)**	**67.1% (8.6%)**	**68% (9.5%)**	**65.9% (6.8%)**	41.6% (0%)	60.4% ($<1\%$)	**71.9% (18.1%)**
Subject IV	**63.4% (1.4%)**	**63.6% (3.6%)**	60.8% ($<1\%$)	61.4% ($<1\%$)	**58.8% (1.4%)**	56.6% (0%)	**66.3% (10%)**	56.5% (0%)
Subject V	**68.2% (13%)**	**68.6% (11.2%)**	**67.8% (11.2%)**	**68.4% (13.4%)**	**66.7% (5.8%)**	**62.7% (2.6%)**	**79% (34.9%)**	**63.5% (5.8%)**
Subject VI	65.1% (0%)	63.1% (0%)	60.6% ($<1\%$)	61.2% (0%)	56.5% (0%)	51% (0%)	67.4% (0%)	**73.9% (7.1%)**
Subject VII	63.7% (0%)	63.5% (0%)	**68.1% (1.6%)**	66.5% ($<1\%$)	53.8% (0%)	59% (0%)	**67% (1.8%)**	**67.7% (2.7%)**
Average Performance	61.7%	61.1%	62.3%	62%	58.5%	55.1%	65.1%	64.8%
Rank Based on Average AUC	11	12	8	9	15	16	3	4

TABLE IV
NUMBER OF TEST TRIALS FOR EACH OF THE SEVEN AVAILABLE SUBJECTS FOR TARGET AND NON-TARGET CLASSES. WE TREAT EACH TRIAL AS A SINGLE OBSERVATION IN D_{TEST}

Subject	I	II	III	IV	V	VI	VII
Target	79	69	70	69	72	76	136
Non-target	220	230	222	221	222	490	857

In order to apply our constructed models in classifying test data, we first use the signal modeling procedure described in (2)–(8) to transform D_{test} into $S_{\text{test}} = \{(\hat{\boldsymbol{\theta}}_1, \boldsymbol{y}_1), \ldots, (\hat{\boldsymbol{\theta}}_n, \boldsymbol{y}_{n_t})\}$ where n_t is the total number of test observations and each $\hat{\boldsymbol{\theta}}_i$ represents a feature vector of $3Mc$ variables estimated on each observation. Table V presents the AUC, the accuracy, and the confusion matrix for the constructed LRR models. The results of this table confirm that the constructed LRR models are remarkably accurate in predicting class variables.

B. Discriminatory Power Spectrum

The signal model (2) allows us to decompose observations into a sum of sinosuids. As we have seen in the previous section, treating the components of these sinusoids as features used in our classification rules led us to construct remarkably accurate classifiers to discriminate positive and negative class for each subject. At the same time, the signal model (2) allows to analyze the frequency spectrum of EEG waveforms. In this regard, power spectrum for each trial is obtained by estimating the frequency and the corresponding power from (3) and (5), respectively, where the power of $\hat{f}_i$ component for $i = 1, ..., M$ is obtained by $\sqrt{\hat{\xi}_{1i}^2 + \hat{\xi}_{2i}^2}$. Fig. 3 shows the power spectral difference between positive and negative class averaged across all channels and trials for each subject and across all subjects. This figure shows that: 1) there is a clear difference between the spectral peaks across positive and negative classes; 2) the difference in spectral peaks is subject dependent; 3) the difference between spectral peaks occurs generally at the first four harmonics of the fundamental frequency of $0.005 \times 256 = 1.28$ Hz; and 4) the difference between spectral peaks is larger at first, second, fourth, and third harmonics, respectively (see the plot on the left right corner of Fig. 3).

C. Channel Rank and Channel-Wise Classifier Validation

We sought to examine the ranking of channels found previously in D_{test}. The results are presented in Table VI. This table shows that channels C2 and Cz and channels POz and PO4 have still the highest and lowest discriminatory effects, respectively. We then examined the predictive capacity of the group of channels with a sensitivity $>1\%$ found in training data (bold cases in Table III)—this choice of sensitivity threshold has been made to guard the classifier against those channels that have been completely biased towards the majority class. In this regard, we evaluated the performance measures of a LRR classifier constructed using only these selected channels via 10-fold CV as well as independent observations D_{test} (see Table VII). Comparing Table VII and Table V suggests the possibility of achieving a classification performance using a subset of channels which is comparable to the full set of channels used. Nevertheless, these selected subset of channels proved to be subject dependent. For example, the classifier constructed using observations collected from channel FCz seems to result in a relatively high classification performance in subject III but not in other subjects.

IV. COMPARING SIGNAL MODELING WITH OTHER FEATURE EXTRACTION METHODS

Here we examine the efficacy of the signal modeling strategy described in Section II-D in comparison with other feature extraction techniques including Principal Component Analysis (referred to as PCA) [75], Morlet Wavelets [76], [77] (referred to as WAV), and the use of full power spectrum estimated by the

TABLE V

PERFORMANCE MEASURES OF THE CONSTRUCTED CLASSIFIERS (LRR WITH λ AND M DETERMINED IN TABLE II) ON INDEPENDENT TEST DATA. P AND N DENOTE POSITIVE (TARGET) AND NEGATIVE (NON-TARGET) CLASSES, RESPECTIVELY

Subject	I	II	III	IV	V	VI	VII
Classifier	LRR	LRR	LRR	LRR	LRR	LRR	LRR
Free Parameters	$M=2$, $\lambda=10^{-8}$	$M=3$, $\lambda=10^{-4}$	$M=3$, $\lambda=1$	$M=2$, $\lambda=10^{-8}$	$M=2$, $\lambda=1$	$M=1$, $\lambda=1$	$M=3$, $\lambda=1$
AUC	79.3%	92.6%	92.9%	79.1%	91.3%	88.8%	78.7%
Accuracy	75.9%	86.6%	89.0%	77.6%	86.4%	89.8%	86.7%
Confusion Matrix	predicted class / true class: P 38 41, N 31 189	predicted class / true class: P 51 18, N 22 208	predicted class / true class: P 52 18, N 14 208	predicted class / true class: P 27 42, N 23 198	predicted class / true class: P 46 26, N 14 208	predicted class / true class: P 26 50, N 8 482	predicted class / true class: P 33 103, N 29 828

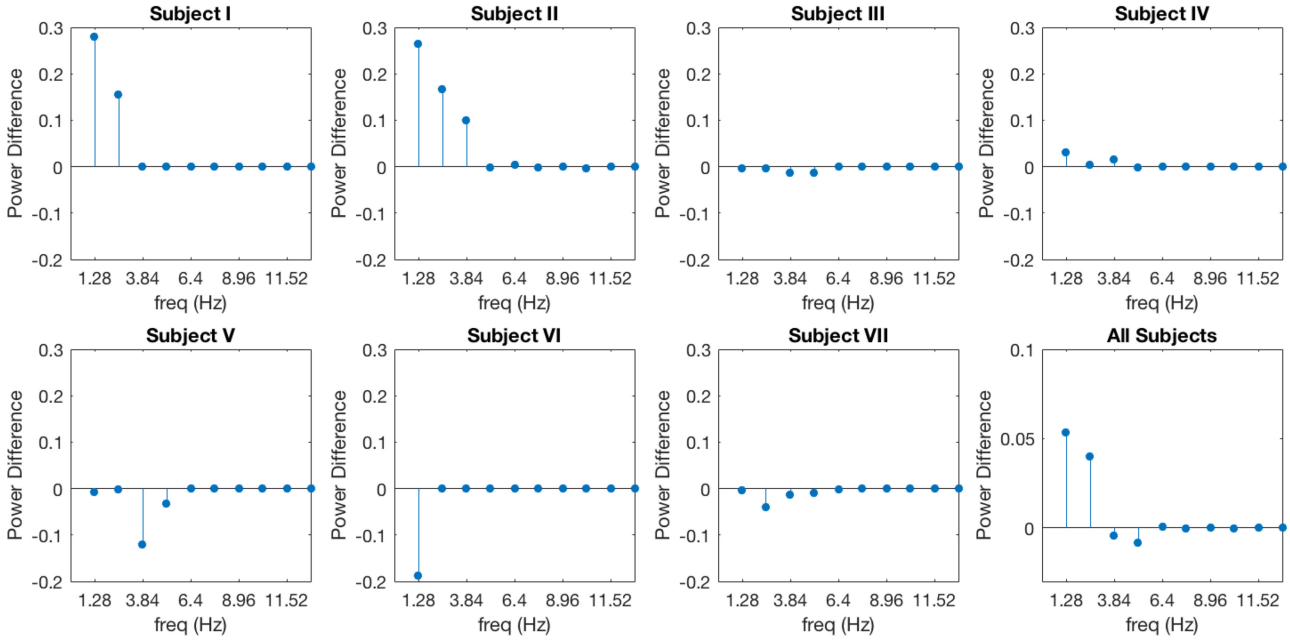


Fig. 3. Power spectral difference between positive and negative class (averaged over all channels and trials) for each subject and across all subjects (the figure on the right lower corner). The frequencies and the power are estimated from the sinusoidal model (2) where M is determined from Table II.

TABLE VI

AUC AND SENSITIVITY (IDENTIFIED AS "SENS.") OF LRR CLASSIFIERS CONSTRUCTED FOR EACH CHANNEL EVALUATED USING TEST DATA

Channel →	**FCz**, AUC (Sens.)	**C3**, AUC (Sens.)	**C1**, AUC (Sens.)	**C2**, AUC (Sens.)	**Cz**, AUC (Sens.)	**C4**, AUC (Sens.)	**CPz**, AUC (Sens.)	**P7**, AUC (Sens.)
Subject I	64.5% (1.2%)	54.3% (0%)	57.2% (0%)	60.2% (0%)	58.3% (0%)	56.5% (0%)	55.1% (0%)	53.8% (0%)
Subject II	62.9% (1.4%)	55.1 (0%)	57.4% (1.4%)	65.2% (5.8%)	64.1% (0%)	63.8% (0%)	52.4 (0%)	56.1% (0%)
Subject III	74.8% (15.7%)	66.6% (8.5%)	68.4% (10%)	87.3% (52.8%)	86.6% (50%)	54.9% (4.2%)	70.2% (14.2%)	64.1% (17.1%)
Subject IV	64.9% (0%)	58.1% (0%)	56% (0%)	71.7% (15.9%)	66.5% (0%)	57.3% (0%)	67.1% (1.4%)	69.7% (2.9%)
Subject V	66.1% (1.4%)	73.4% (12.5%)	82.1% (40.8%)	84.2% (41.6%)	82.9% (41.6%)	71.6% (13.8%)	68.4% (8.3%)	72.2% (12.5%)
Subject VI	52.4% (0%)	66.0% (1.3%)	63.8% (0%)	72.5% (1.3%)	74.6% (2.6%)	66.5% (0%)	69.6% (1.3%)	69.7% (1.3%)
Subject VII	60.0% (0%)	58.4% (0%)	57.6% (0%)	67.9% (1.4%)	68.7% (4.4%)	64.3% (0%)	61.8% (0%)	61.4% (1.4%)
Average AUC Across Subjects	63.65%	61.7%	63.2%	72.7%	71.6%	62.1%	63.5%	63.85%
Rank Based on Average AUC	9	13	11	1	2	12	10	7
Channel →	**P3**, AUC (Sens.)	**P4**, AUC (Sens.)	**P8**, AUC (Sens.)	**PO3**, AUC (Sens.)	**POz**, AUC (Sens.)	**PO4**, AUC (Sens.)	**O1**, AUC (Sens.)	**O2**, AUC (Sens.)
Subject I	54.5% (0%)	52% (0%)	55.4% (0%)	54.3% (0%)	54.9% (0%)	65.2% (5.0%)	60.5% (0%)	66.6% (1.2%)
Subject II	55.3% (0%)	55.1% (1.4%)	54.6% (0%)	55.3% (1.4%)	57.2% (0%)	58.4% (0%)	64.6% (1.4%)	64.2% (0%)
Subject III	65.7% (8.5%)	66% (4.2%)	69.3% (14.2%)	71% (20%)	61.6% (5.7%)	49.8% (0%)	60.2% (0%)	75.3% (27.1%)
Subject IV	71.4% (7.2%)	62.6% (8.7%)	66.4% (2.8%)	66.8% (1.4%)	49.8 (0%)	59.5% (1.4%)	69.2% (5.7%)	66.2% (0%)
Subject V	66.5% (12.5%)	67.4% (11.1%)	71.6% (15.2%)	73.5% (13.8%)	66.8% (5.5%)	52.1% (4.1%)	76.6% (23.6%)	58.3% (2.7%)
Subject VI	68.2% (0%)	65.2% (1.3%)	67.2% (1.3%)	63% (1.3%)	56.1% (0%)	52.6% (0%)	67.1% (1.3%)	69.9% (3.9%)
Subject VII	65.3% (0%)	61% (0%)	69.7% (2.9%)	68.1% (0%)	53.6% (0%)	55.6% (0%)	64.7% (2.9%)	66.3% (0%)
Average Performance	63.8%	61.3%	64.8%	64.5%	57.1%	56.1%	66.1%	66.6%
Rank Based on Average AUC	8	14	5	6	15	16	4	3

TABLE VII
PERFORMANCE MEASURES OF THE LRR CLASSIFIERS CONSTRUCTED USING $3M\times$ (NUMBER OF CHANNELS IN BOLD IN TABLE III) EXTRACTED FEATURES OBTAINED FROM (2)–(8) EVALUATED USING 10-FOLD CROSS-VALIDATION AS WELL AS INDEPENDENT TEST DATA. P AND N DENOTE POSITIVE (TARGET) AND NEGATIVE (NON-TARGET) CLASSES, RESPECTIVELY

Subject	I	II	III	IV	V	VI	VII
Channels	FCz, C2, PO4, O2	C1, Cz, C4, PO3, O1	FCz, C3, C1, Cz, C2, C4 CPz, P7, P3, P4, POz, O2	Cz,CPz,P7,P3,P4 P8,PO3,POz,O1	C3, C1, Cz, C2, C4, CPz P7, P3, P4, P8, PO3, POz PO4, O1, O2	C3, C2, CPz, P7, O2	Cz, C2, P8, O1, O2
AUC (10-CV)	74.4%	90.0%	86.6%	75.7%	92.1%	79.3%	76.7%
Accuracy (10-CV)	77.8%	84.9%	84.9%	77.3%	86.4%	86.6%	86.4%
AUC (Test Data)	79.9%	91.3%	93.6%	71.9%	91.1%	76.5%	75.6%
Accuracy (Test Data)	76.6%	86.3	89.04%	77.6%	86.3%	86.2%	87.8%
Confusion Matrix on Test Data	predicted class / true class: P 20 59, N 11 209	predicted class / true class: P 48 21, N 20 210	predicted class / true class: P 54 16, N 16 206	predicted class / true class: P 26 43, N 22 199	predicted class / true class: P 46 26, N 14 208	predicted class / true class: P 6 70, N 8 482	predicted class / true class: P 23 113, N 8 849

multitaper technique (referred to as PSD) [78] estimated using the MNE python toolbox [79].

In our comparative analysis, we not only use our collected data on Healthy Individuals that was described in Section II (hereafter, referred to as HI data), but we also use a publicly available dataset collected on eight subjects with amyotrophic lateral sclerosis (ALS) [80] (hereafter, referred to as ALS data). As a result, we need to apply the proposed signal modeling scheme on the ALS data as well. The ALS data consists of a pair of train and test datasets collected for each subject. The training data for each subject contains a set of 300 and 1500 observations for target and non-target stimuli, respectively. The test data for each subject contains 400 and 2000 observations for the matching non-target and target stimuli. A summary of experimental protocol used in [80] to collect the ALS data is provided in Supplementary Materials Section I.

Similar to the analysis performed on our collected HI data, we first apply the procedure detailed in Section II-E to determine the best combination of sinusoidal model order and classifier on ALS data. The results of this analysis is tabulated in Table S1 in Supplementary Materials Section II. This table shows that the LRR classifier outperforms all other classifiers in terms of achieving a higher AUC determined by 10-fold cross-validation on training data. Interestingly, and in contrast with HI dataset, a model order of $M = 1$ (i.e., the first harmonic at 1.28 Hz) consistently leads to the highest AUC in the ALS data. As in Sections II-F, III-A, and III-C, we tabulated the results of channel ranking on ALS training data, classifier validation (on ALS test data), and channel-wise classifier validation (on ALS test data) in Tables S2, S3, and S4, respectively (see in Supplementary Materials Section III). The results confirm that the constructed LRR models are remarkably accurate in predicting class variables.

As discussed earlier, we now compare the results of the sinusoidal signal modeling with PCA, WAV, and PSD on both HI and ALS datasets. Fig. 5(A, B, C) and 5(D, E, F) show the performance of LRR and the signal modeling mechanism (bold cases in Tables II and S1 for HI and ALS data, respectively) in comparison with a combination of other classifiers and other feature extraction techniques. The results confirm that our constructed LRR models with features extracted from the sinusoidal models outperform other methods in terms of achieving the highest AUC across both datasets. Fig. 5(A, E) shows the proposed signal modeling approach outperforms WAV, PSD, PCA methods across both HI and ALS data. Furthermore, analysis of the range and distributional characteristics of AUC scores across subjects as well as the level of the scores shows that our approach is robust as shown in Fig. 5(C, F).

V. DISCUSSION

Constructing accurate predictive models is at the heart of BCI technologies because these models can ultimately translate brain activities into communication and control commands [28]. In this study, we hypothesize that a set of spatiospectral features extracted from ERP waveforms enable constructing accurate predictive models that represent the mental intent of users of a BCI speller. In order to capture the dominant discriminatory spectral contents of our recorded EEG waveforms, we model observations by a mathematical form, which is the sum of sinusoids with unknown amplitudes, frequencies, and phases. A nonlinear least square technique is then used to estimate the unknown parameters of this signal model [52]. We benefit from this signal modeling stage in two ways. First, the set of model parameters include dominant components of the signal spectrum. The set of spectral features appeared in the accurate predictive models developed herein proves that the major discriminatory spectral contents of our recorded ERPs align with the low-frequency brain rhythms of delta and low theta band activity (frequencies <6.4 Hz). Secondly, the dimensionality reduction performed by the signal modeling is warranted because generally the larger is the ratio of the sample size (i.e., the number of trials) to the dimensionality of observations, the more likely is to train an accurate classifier. In our analysis, the constructed subject-specific classifiers exhibit an AUC ranging from 78.7% to 92.6% on test data. As discussed by Pines and Everett [81], the accuracy of a predictive model measured by AUC can be seen as "poor", "fair", "good", and "excellent" when the AUC lies within 60%–70%, 70%–80%, 80%–90%, and 90%–100%, respectively. The type of spectral analysis that we have conducted here depends on the order of the sinusoidal model that leads to the highest classification performance (see Section II-E). At the same time, in order to have an efficient estimation-classification process, we have restricted M to $1, 2,$ or 3 at most. In other words, the number of estimated dominant and discriminatory frequency contents for each individual

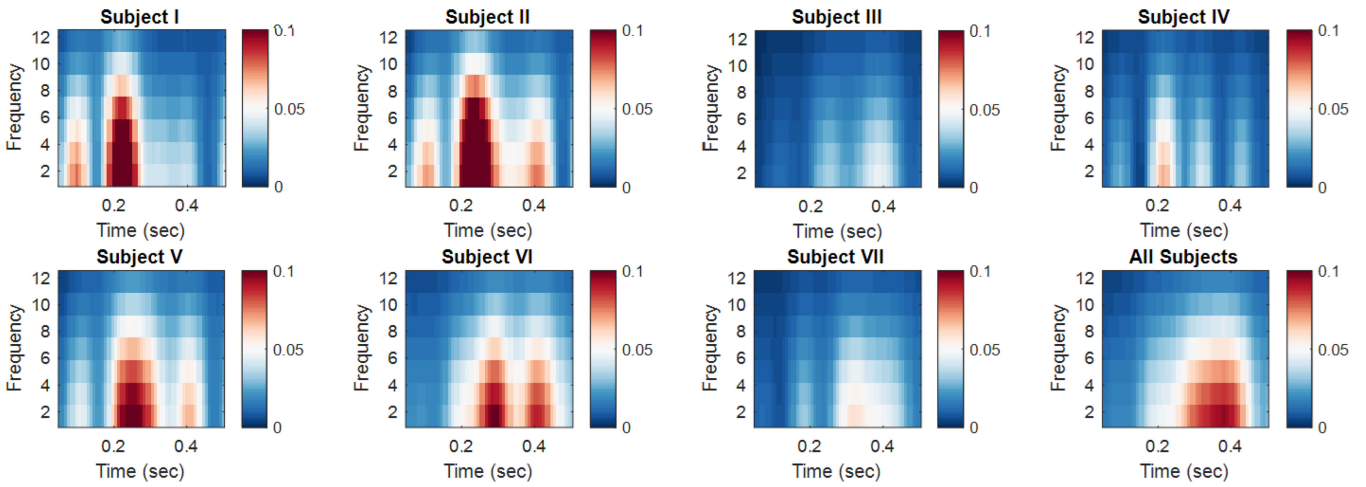
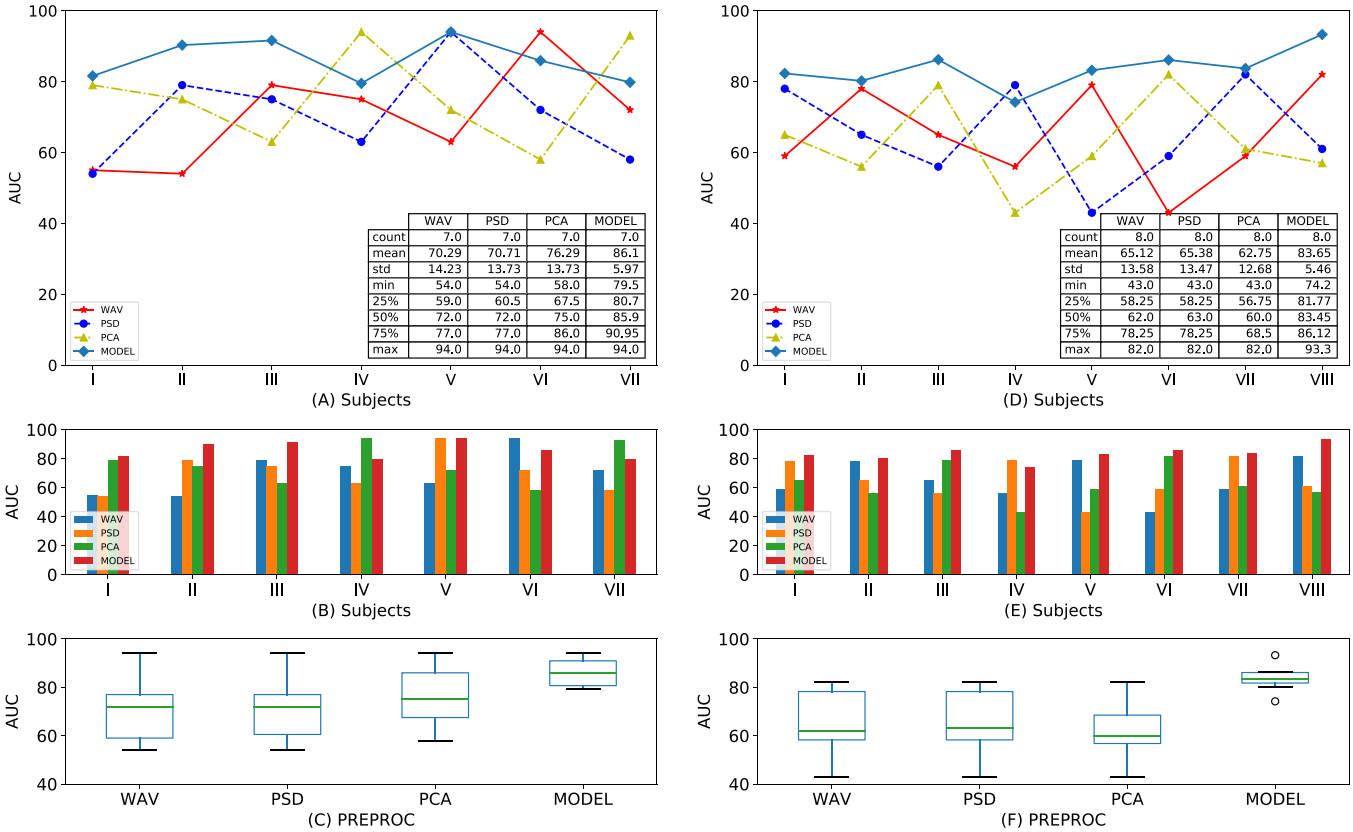


Fig. 4. The absolute difference of power spectrum between target and non-target ERPs trials for all subjects in time-frequency representation. This representation was obtained using a Hanning taper approach with an adaptive time window of six cycles per frequency in 20 ms steps for frequencies from 0.1 to 12 Hz.



	WAV	PSD	PCA	MODEL
count	7.0	7.0	7.0	7.0
mean	70.29	70.71	76.29	86.1
std	14.23	13.73	13.73	5.97
min	54.0	54.0	58.0	79.5
25%	59.0	60.5	67.5	80.7
50%	72.0	72.0	75.0	85.9
75%	77.0	77.0	86.0	90.95
max	94.0	94.0	94.0	94.0

	WAV	PSD	PCA	MODEL
count	8.0	8.0	8.0	8.0
mean	65.12	65.38	62.75	83.65
std	13.58	13.47	12.68	5.46
min	43.0	43.0	43.0	74.2
25%	58.25	58.25	56.75	81.77
50%	62.0	63.0	60.0	83.45
75%	78.25	78.25	68.5	86.12
max	82.0	82.0	82.0	93.3

Fig. 5. Performance measures of the logistic regression (LRR) model using Morlet wavelets (WAV), power-spectral density (PSD), principal-component analysis (PCA) and signal modeling (MODEL) features on healthy individuals' data (A, B, C), and ALS subjects data (D, E, F). *Top row*: Descriptive statistics of LRR. *Middle row*: Subject-specific performance of LRR using four different preprocessing methods. *Bottom row*: Box and Whisker plot analysis of the LRR model on four different preprocessing methods across all subjects.

is subject to the classification performance as well as the efficiency of the estimation process. This can be seen from Fig. 3 where the power spectrum differences for each subject have non-zero magnitudes at most at three distinct frequencies.

This then raises the question as to whether any important discriminatory frequency content has been missing by such classification-restricted spectrum analysis. In this regard, we calculated the time-frequency representation (TFR) of subject-specific EEG recordings across all subjects as shown in Fig. 4. We computed TFR using using a Fourier approach: 1) applied a sliding Hanning taper with an adaptive time window of six cycles for each frequency in 20 ms steps for frequencies from 0.1

to 12 Hz for each class of ERP trials separately; 2) performed grand averaging in time-frequency domain; 3) took the absolute difference of target and non-target [82]. One drawback of this analysis is that it does not take into account the efficacy or the efficiency of classification rules. On the positive side, time-frequency transforms and distributions simultaneously retrieve the temporal and frequency domain structure of observations. This result confirms that delta and lower theta bands contain the major part of discriminatory frequency contents for each subject. This is interesting because many current ERP-based BCIs are driven by a wider range of spectral bandwidth that also include faster rhythmic brain activities in the alpha or beta band [83, p. 95], [84]. Furthermore, there is no consensus regarding the appropriate cut-off frequencies of a band-pass filter used to extract frequency content for P300 feature extraction in the BCI literature. For instance, Farquhar *et al.* [30] suggested that the most discriminative frequency range lies within (0.1 Hz–12 Hz). While other studies use different frequency bands to extract spectral features, such as (0.1 Hz–21.33 Hz) used in [22], (0.1 Hz–10.66 Hz) in [85], (1 Hz–17 Hz) in [86], (0.1 Hz–30 Hz) in [51], [87] (see [46] for an extensive review) and recently it was suggested that the best frequency range lies between (0.1 Hz–21.33 Hz) reported in [22].[1] In contrast, we found that spectral range of $<$6.4 Hz contains most discriminative frequencies for enhanced decoding of attention modulated P300 features and can be used in the design of a BCI speller. Our finding is corroborated with the previous research in neurophysiology of attention modulated ERPs that show the contribution of delta (lower theta) oscillations better represent cognitive processes such as attention, perception, and decision-making [88]–[90]. We contend that our signal modeling approach captured these intrinsic features of ERPs, and as a result, led to accurate mental state decoding models.

Additionally, we ranked the efficacy of each channel in predicting users' mental intent based on the extracted spectral features. The analysis shows that central channels, over the sensorimotor cortex, Cz and C2 (ranked 1st and 2nd) and the parieto-occipital channels POz and PO4 (ranked 15th and 16th) have the highest and lowest discriminatory effects, respectively. This is an interesting finding that aligns with other channel localization studies where the best discriminative channels are spread over the sensorimotor cortex despite a slight subject-specific variation [91]–[93].

Based on this channel rank analysis, we also proposed a channel selection procedure based on the extracted novel spectral features with the aim to reduce the number of channels while generally keeping a comparable classification performance. On five subjects we observed a drop in the AUC, which was as low as 0.2% for subject V to 12.6% for subject VI. However, for two subjects (I and III) the reduced set of channels slightly improved the classification performance in terms of AUC. Nevertheless, this analysis confirms the possibility of achieving an "excellent" classification performance in a BCI application with very few channels—for instance, we achieved an AUC of 91.3% for subject II with as few as five channels.

Furthermore, the performance of our proposed method can be seen in Fig. 5 cross-validated across seven healthy and eight ALS subjects. These results manifest that the signal modeling based approach is more robust against variability and highly accurate when compared to a set of widely used methods (i.e., PCA, Wavelets, PSDs). This finding is important because most BCIs suffer from the presence of non-task relevant features, enormous variability, and high-dimensionality [29] that lead to the failure of a previously functioning BCI system.

VI. Conclusion

The signal modeling strategy used herein along with the logistic regression with L_2-ridge penalty have enabled us to extract intrinsic spectral features of ERPs and to construct remarkably accurate classifiers for predicting users' mental intent in a BCI speller system. The proposed framework is simple and efficient to implement and it is believed to be used in the future in other ERP-based BCI applications. The set of dominant and discriminatory frequency contents of ERPs used in our constructed predictive models lies in a range less than 6.4 Hz, which aligns with the low-frequency rhythmic brain activity of delta and low theta bands. This is interesting because many current BCIs are focused on and are driven by a wider frequency bandwidth (0.1 Hz–30 Hz) that also include faster rhythmic brain activities in the alpha or beta bands. As also discussed in [84], introducing slower oscillations such as delta or theta band activity in BCIs could potentially extent applications of BCI. Our analysis also confirms the possibility of using very few channels to construct remarkably accurate classifiers in BCI systems. We believe analyses of this type is important in BCIs to minimize the number of electrodes and the setup time, which in turn leads to more comfortable and robust BCI systems. The next natural step in this line of work is to use the spectral contents of ERPs to construct subject-independent predictive models of users' mental intent.

[1]Note that these studies use a band-pass filter as a spectral filter to restrict the bandwidth to discriminatory frequency content; however, after spectral filtering time-domain features are still being used.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002.

[2] A. Vallabhaneni, T. Wang, and B. He, "Brain–computer interface," in *Neural Engineering*. New York, NY, USA: Springer, 2005, pp. 85–121.

[3] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces in the completely locked-in state and chronic stroke," *Prog. Brain Res.*, vol. 228, pp. 131–161, 2016.

[4] M. C. Kiernan *et al.*, "Amyotrophic lateral sclerosis," *Lancet*, vol. 377, no. 9769, pp. 942–955, 2011.

[5] J. Morris, "Amyotrophic lateral sclerosis (ALS) and related motor neuron diseases: An overview," *Neurodiagnostic J.*, vol. 55, no. 3, pp. 180–194, 2015.

[6] U. Chaudhary, B. Xia, S. Silvoni, L. G. Cohen, and N. Birbaumer, "Brain–computer interface–based communication in the completely locked-in state," *PLoS Biol.*, vol. 15, no. 1, 2017, Art. no. e1002593.

[7] L. M. McCane *et al.*, "P300-based brain-computer interface (BCI) event-related potentials (ERPs): People with amyotrophic lateral sclerosis (ALS) vs. age-matched controls," *Clin. Neurophysiol.*, vol. 126, no. 11, pp. 2124–2131, 2015.

[8] E. W. Sellers and E. Donchin, "A P300-based brain–computer interface: Initial tests by ALS patients," *Clin. Neurophysiol.*, vol. 117, no. 3, pp. 538–548, 2006.

[9] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A P300 event-related potential brain–computer interface (BCI): The effects of matrix size and inter stimulus interval on performance," *Biol. Psychol.*, vol. 73, no. 3, pp. 242–252, 2006.

[10] B. He, S. Gao, H. Yuan, and J. R. Wolpaw, "Brain–computer interfaces," in *Neural Engineering*. New York, NY, USA: Springer, 2013, pp. 87–151.

[11] J. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. New York, NY, USA: Oxford Univ. Press, 2012.

[12] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *J. Neural Eng.*, vol. 4, no. 2, pp. R32–R57, Mar. 2007.

[13] S. Gao, Y. Wang, X. Gao, and B. Hong, "Visual and auditory brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1436–1447, May 2014.

[14] A. Zhumadilova, D. Tokmurzina, A. Kuderbekov, and B. Abibullaev, "Design and evaluation of a P300 visual brain-computer interface speller in cyrillic characters," in *Proc. 26th IEEE Int. Symp. Robot–Human Interactive Commun.*, Aug. 2017, pp. 1006–1011.

[15] E. Baykara *et al.*, "Effects of training and motivation on auditory P300 brain–computer interface performance," *Clin. Neurophysiol.*, vol. 127, no. 1, pp. 379–387, 2016.

[16] N. A. Bhagat *et al.*, "Design and optimization of an EEG-based brain machine interface (BMI) to an upper-limb exoskeleton for stroke survivors," *Frontiers Neurosci.*, vol. 10, pp. 1–17, 2016, Art. no. 122.

[17] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.

[18] R. Fazel-Rezai and W. Ahmad, *P300-Based Brain–Computer Interface Paradigm Design*. Rijeka, Croatia: InTech, 2011.

[19] J. Jin, H. Zhang, I. Daly, X. Wang, and A. Cichocki, "An improved P300 pattern in BCI to catch users attention," *J. Neural Eng.*, vol. 14, no. 3, 2017, Art. no. 036001.

[20] H. Serby, E. Yom-Tov, and G. F. Inbar, "An improved P300-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 1, pp. 89–98, Mar. 2005.

[21] G. Townsend and V. Platsko, "Pushing the P300-based brain–computer interface beyond 100 BPM: Extending performance guided constraints into the temporal domain," *J. Neural Eng.*, vol. 13, no. 2, 2016, Art. no. 026024.

[22] H. Cecotti and A. J. Ries, "Best practice for single-trial detection of event-related potentials: Application to brain-computer interfaces," *Int. J. Psychophysiol.*, vol. 111, pp. 156–169, 2017.

[23] K.-R. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz, "Machine learning techniques for brain-computer interfaces," *Biomed. Technol.*, vol. 49, no. 1, pp. 11–22, Dec. 2004.

[24] B. Blankertz *et al.*, "The Berlin brain-computer interface: Progress beyond communication and control," *Frontiers Neurosci.*, vol. 10, pp. 1–24, 2016, Art. no. 530.

[25] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih, and A. Kübler, "P300 brain computer interface: Current challenges and emerging trends," *Frontiers Neuroeng.*, vol. 5, pp. 1–14, 2012, Art. no. 14.

[26] J. E. Huggins *et al.*, "Workshops of the fifth international brain-computer interface meeting: Defining the future," *Brain–Comput. Interfaces*, vol. 1, no. 1, pp. 27–49, 2014.

[27] A. Kübler, "Quo vadis P300 BCI," in *Proc. 5th Int. Winter Conf. Brain–Comput. Interface*, 2017, pp. 36–39.

[28] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, pp. R1–R13 Jun. 2007.

[29] D. J. Krusienski *et al.*, "Critical issues in state-of-the-art brain–computer interface signal processing," *J. Neural Eng.*, vol. 8, no. 2, 2011, Art. no. 025002.

[30] J. Farquhar and N. J. Hill, "Interactions between pre-processing and classification methods for event-related-potential classification," *Neuroinformatics*, vol. 11, no. 2, pp. 175–192, 2013.

[31] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components—A tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, 2011.

[32] M. Siepmann and W. Kirch, "Effects of caffeine on topographic quantitative EEG," *Neuropsychobiology*, vol. 45, no. 3, pp. 161–166, 2002.

[33] L. Roijendijk, "Variability and nonstationarity in brain computer interfaces," Master's thesis, Dept. Cogn. Artif. Intell., Radboud Univ. Nijmegen, Nijmegen, The Netherlands, 2009.

[34] A. Rakotomamonjy and V. Guigue, "BCI competition III: Dataset II—Ensemble of SVMs for BCI P300 speller," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1147–1154, Mar. 2008.

[35] K. Yoon and K. Kim, "Multiple kernel learning based on three discriminant features for a P300 speller BCI," *Neurocomputing*, vol. 237, pp. 133–144, 2017.

[36] R. M. Mowla, E. J. Huggins, and E. D. Thompson, "Enhancing P300-BCI performance using latency estimation," *Brain–Comput. Interfaces*, vol. 4, no. 3, pp. 137–145, Jul. 2017.

[37] N. J. Mak *et al.*, "Optimizing the P300-based brain-computer interface: Current status, limitations and future directions," *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025003.

[38] V. Lafuente, M. J. Gorriz, J. Ramirez, and E. Gonzalez, "P300 brainwave extraction from EEG signals: An unsupervised approach," *Expert Syst. Appl.*, vol. 74, pp. 1–10, May 2017.

[39] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: Feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1057–1061, Jun. 2004.

[40] D. J. Krusienski, E. W. Sellers, and T. M. Vaughan, "Common spatio-temporal patterns for the P300 speller," in *Proc. 3rd Int. IEEE/EMBS Conf. Neural Eng.*, 2007, pp. 421–424.

[41] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: Application to brain–computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2035–2043, Aug. 2009.

[42] A. Turnip, K.-S. Hong, and M.-Y. Jeong, "Real-time feature extraction of P300 component using adaptive nonlinear principal component analysis," *Biomed. Eng. Online*, vol. 10, pp. 1–20, Sep. 2011.

[43] M. M van Vliet *et al.*, "Single-trial ERP component analysis using a spatio-temporal LCMV," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 55–66, Jan. 2016.

[44] B. Wittevrongel and M. M. Van Hulle, "Faster P300 classifier training using spatiotemporal beamforming," *Int. J. Neural Syst.*, vol. 26, no. 3, May 2016, Art. no. 1650014.

[45] B. Wittevrongel and M. M. Van Hulle, "Spatiotemporal beamforming: A transparent and unified decoding approach to synchronous visual brain-computer interfacing," *Frontiers Neurosci.*, vol. 11, p. 630, Nov. 2017.

[46] L. Bougrain, C. Saavedra, and R. Ranta, "Finally, what is the best filter for P300 detection?" in *Proc. Tools Brain–Comput. Interact.*, 2012, pp. 1–2. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00756669/

[47] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018, Art. no. 031005.

[48] P. Wang, J. Lu, B. Zhang, and Z. Tang, "A review on transfer learning for brain-computer interface classification," in *Proc. 5th Int. Conf. Inf. Sci. Technol.*, 2015, pp. 315–322.

[49] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[50] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage*, vol. 56, no. 2, pp. 387–399, 2011.

[51] R. M. Mowla, E. J. Huggins, and E. D. Thompson, "Enhancing P300-BCI performance using latency estimation," *Brain–Comput. Interfaces*, vol. 4, no. 3, pp. 137–145, Jul. 2017.

[52] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume III: Practical Algorithm Development*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2013.

[53] D. Talsma and M. G. Woldorff, "Methods for the estimation and removal of artifacts and overlap in ERP data," in *Event-Related Potentials: A Methods Handbook*. Cambridge, MA, USA: MIT Press, 2005, p. 115.

[54] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, 3rd ed. New York, NY, USA: Springer, 2011.

[55] M. Kubat, R. Holte, and S. Matwin, "Learning when negative examples abound," *Pattern Recognit. Lett.*, vol. 27, pp. 861–874, 2006.

[56] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.

[57] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, pp. 861–874, 2006.

[58] D. J. Hand and R. J. Till, "A simple generalization of the area under the ROC curve for multiple class classification problems," *Mach. Learn.*, vol. 45, pp. 171–186, 2001.

[59] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ, USA: Wiley, 2001.

[60] S. le Cessie and J. C. van Houwelingen, "Ridge estimators in logistic regression," *Appl. Statist.*, vol. 41, no. 1, pp. 191–201, 1992.

[61] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining Practical Machine Learning Tools and Techniques*, 4th ed. New York, NY, USA: Elsevier, 2017.

[62] J. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, vol. 29, pp. 131–163, 1997.

[63] A. R. Webb, *Statistical Pattern Recognition*, 1st ed. Hoboken, NJ, USA: Wiley, 2002.

[64] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[65] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Statist. Softw.*, vol. 33, pp. 1–22, 2010.

[66] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, Mar. 2005.

[67] P. Langley, W. Iba, and K. Thompson, "An analysis of Bayesian classifiers," in *Proc. 10th Nat. Conf. Artif. Intell.*, 1992, pp. 223–228.

[68] S. Song, Z. Zhan, Z. Long, J. Zhang, and L. Yao, "Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data," *PLoS One*, vol. 6, 2011, Art. no. e17191.

[69] J. Cheng and R. Greiner, "Comparing Bayesian network classifiers," in *Proc. 15th Conf. Uncertainty Artif. Intell.*, 1999, pp. 101–108.

[70] J. Fruitet, D. J. McFarland, and J. R. Wolpaw, "A comparison of regression techniques for a two-dimensional sensorimotor rhythm-based brain computer interface," *J. Neural Eng.*, vol. 7, 2010, Art. no. 016003.

[71] N. Ueda, Y. Tanaka, and A. Fujino, "Robust naive Bayes combination of multiple classifications," in *The Impact of Applications on Mathematics* (Mathematics for Industry), M. Wakayama *et al.*, Ed. Tokyo, Japan: Springer, 2014.

[72] M. Kubat, R. Holte, and S. Matwin, "Learning when negative examples abound," in *Proc. Eur. Conf. Mach. Learn.*, 1997, pp. 146–153.

[73] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, Jun. 2011.

[74] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 66, pp. 1–21, 2015.

[75] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.

[76] C. S. Herrmann, M. Grigutsch, and N. A. Busch, "EEG oscillations and wavelet analysis," in *Event-Related Potentials: A Methods Handbook*. Cambridge, MA, USA: MIT Press, 2005, p. 229.

[77] P. Flandrin, *Time-Frequency/Time-Scale Analysis* (Wavelet Analysis and Its Applications), vol. 10. New York, NY, USA: Academic, 1999.

[78] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques.* New York, NY, USA: Cambridge Univ. Press, 1993..

[79] A. Gramfort *et al.*, "MEG and EEG data analysis with MNE-Python," *Frontiers Neurosci.*, vol. 7, pp. 1–13, 2013, Art. no. 267.

[80] A. Ricco *et al.*, "Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis," *Frontiers Human Neurosci.*, vol. 7, pp. 1–9, 2013, Art. no. 732.

[81] J. M. Pines and W. W. Everett, *Evidence-Based Emergency Care Diagnostic Testing and Clinical Decision Rules*. Oxford, U.K.: Blackwell, 2008.

[82] D. D. Cox, "Spectral analysis for physical applications: Multitaper and conventional univariate techniques," *Technometrics*, vol. 38, no. 3, 1996.

[83] B. He, *Neural Engineering*. New York, NY, USA: Springer, 2013.

[84] J. A. Barios *et al.*, "Delta-theta intertrial phase coherence increases during task switching in a BCI paradigm," in *Biomedical Applications Based on Natural and Artificial Computing* (Notes in Computer Science). New York, NY, USA: Springer, 2017, pp. 96–108.

[85] J. M. Leiva and S. M. M. Martens, "MLSP competition, 2010: Description of first place method," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Aug. 2010, pp. 112–113.

[86] B. Labb, X. Tian, and A. Rakotomamonjy, "MLSP competition, 2010: Description of third place method." in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Aug. 2010, pp. 116–117.

[87] D. E. Thompson, S. Warschausky, and J. E. Huggins, "Classifier-based latency estimation: A novel way to estimate and predict BCI accuracy," *J. Neural Eng.*, vol. 10, no. 1, 2012, Art. no. 016006.

[88] C. Başar-Eroglu, E. Başar, T. Demiralp, and M. Schürmann, "P300-response: Possible psychophysiological correlates in delta and theta frequency channels: A review," *Int. J. Psychophysiol.*, vol. 13, no. 2, pp. 161–179, 1992.

[89] B. Güntekin and E. Başar, "Review of evoked and event-related delta responses in the human brain," *Int. J. Psychophysiol.*, vol. 103, pp. 43–52, 2016.

[90] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, 2007.

[91] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15–21, 2008.

[92] K. Colwell, D. Ryan, C. Throckmorton, E. Sellers, and L. Collins, "Channel selection methods for the P300 speller," *J. Neurosci. Methods*, vol. 232, pp. 6–15, 2014.

[93] M. Xu *et al.*, "Channel selection based on phase measurement in P300-based brain-computer interface," *PloS One*, vol. 8, no. 4, 2013, Art. no. e60608.