Event-Related Spectrogram Representation of EEG for CNN-Based P300 Speller

Ayana Mussabayeva, Zangar Ermaganbet, Prashant Kumar Jamwal, and Muhammad Tahir Akhtar
Department of Electrical and Computer Engineering,
Nazarbayev University, Kabanbay Batyr Ave. 53, Nur-Sultan, 010000, Kazakhstan
E-mails: {ayana.mussabayeva, zangar.ermaganbet, prashant.jamwal, muhammad.akhtar}@nu.edu.kz, akhtar@ieee.org

***Abstract*—P300 speller is one of the most popular applications for electroencephalography (EEG) features extraction and classification. It is used for enabling paralyzed people to communicate with the outer world by inputting text using EEG signals, evoked by their brain. Classical P300 speller uses EEG time-series classification, which does not cover all of the frequency bands of brain activity and is very subject-dependent. In order to make the system more stable and robust, frequency domain spectrograms such as intertrial coherence (ITC) and event-related spectral perturbation (ERSP) are generated from EEG data of amyotrophic lateral sclerosis (ALS) patients. The obtained spectrograms are further classified using a convolutional neural network (CNN). CNN-based ensemble model is proposed for classifying both ITC and ERSP in order to achieve more trusted results. Classification of EEG data in the form of images can make the whole process much more uniform, robust and independent from the data acquisition properties.**

***Index Terms*—Brain-Computer Interface, P300 Speller, Event-Related Spectral Perturbation, Intertrial Coherence, Convolutional Neural Network.**

I. Introduction

Brain-computer interface (BCI) systems process human brain activity signals to control some devices. The activity of a human brain can be obtained using different neuroimaging techniques, such as electroencephalography (EEG), functional magnetic resonance imaging (fMRI), near-infrared spectroscopy (NIRS), etc. However, the most popular data acquisition technique for BCI devices is EEG, as it is non-invasive and provides a fast response to the electrical activity of the neurons.

Generally, BCI devices collect brain activity data, extract the essential features from the processed data, and translate those features into device commands. BCI systems have become a popular application of brain signals processing in different spheres over the last decade. Recently entertainment and gaming industry has started using BCI devices for controlling video games [1]. Apart from that, BCI systems are widely used for designing brain-controlled assisting devices, such as prosthesis [2] or wheelchair [3]. Another significant type of BCI device is the speller system. BCI spellers enable people to communicate with the outer world by just processing their EEG signals. That is the only way to communicate for people suffering from severe motor neuron disorders, such as amyotrophic lateral sclerosis (ALS), peripheral neuropathy, cerebral palsy, etc.

EEG-based speller system is called P300 speller, introduced back in the 1980s [4]. P300 is a sharp positive voltage deflection detected at about 300 ms after a human recognizes some target visual stimulus. P300 is one of the event-related potential (ERP) components. When using a P300 speller, a user observes a flashing matrix of characters, intensified by rows and columns. The user concentrates on a particular character, which should be spelled. When the chosen character is intensified, a sharp P300 component can be detected in the EEG signal of the user. The classification of the EEG signals in the P300 speller is presented as a binary classification problem when the current EEG signal should be classified as target or non-target.

Classically, the EEG signal is presented as a time series, which makes its processing very dependent on the sampling frequency, chosen periods, and other data acquisition parameters. Frequency-based spectrogram representation can be used to overcome this problem. Representing EEG signals in the form of spectrograms enables classification of the data as images, by using standard image-processing models, such as convolutional neural network (CNN). In this work, intertrial coherence (ITC) and event-related spectral perturbation (ERSP) spectrograms [5] are generated from ALS patients' data and processed using a CNN-based ensemble classifier.

The remainder of this paper is organised as follows. Section II reviews the dataset used for the experiments, the generation of the spectrograms, and the proposed classification models, based on CNN. Section III represents the performance evaluation and the results obtained on validation and test. Finally, the conclusions are presented in Section IV.

II. Methodology

A. Dataset Description

The dataset used in the experiments is an open-access dataset provided by BCI Horizon 2020 [6]. The interface of the P300 speller used is the classical 6×6 matrix graphical user interface (GUI) for English-speaking users.

The data is collected from 8 subjects suffering from ALS. According to the statistics, ALS is one of the most common neuron diseases. ALS is also known as Lou Gehrig's disease. Each year more than six thousand people are diagnosed with ALS all over the world [7]. The dataset presents the data collected from five male and three female subjects. The details about the subjects are presented in Table I. There are two

TABLE I
ALS PATIENTS INFORMATION FROM BCI HORIZON 2020 DATASET

Subject	Age	Sex	Type of ALS
A01	56	Male	Spinal
A02	59	Male	Spinal
A03	43	Male	Spinal
A04	75	Female	Bulbar
A05	60	Female	Bulbar
A06	40	Male	Spinal
A07	61	Male	Bulbar
A08	72	Female	Bulbar

different types of ALS: spinal and bulbar. Spinal onset ALS usually starts from paralysis of the limbs of the patients. In bulbar onset, ALS breathing and speaking muscles are the first to be affected and paralyzed. People older than 65 years usually suffer from bulbar onset ALS, while younger patients are commonly diagnosed with spinal onset ALS. This tendency can also be noted from the table above.

The dataset was recorded using g.tec g.MOBILAB equipment using g.Ladybird active electrodes. The provided EEG signal was acquired using eight active electrodes: Fz, Cz, P3, Pz, P4, P07, P08, and Oz.

B. Spectrograms Generation

To generate spectrogram representation of 8-channel EEG data, independent component analysis (ICA) has been applied to the data. ICA is a statistical tool used in blind source separation (BSS). BSS is a problem where the mixture of initial data is known while the mixing system and initial data are unidentified. The main principles of ICA are the independence and nongaussianity of the components.

There are different approaches to ICA: it can be computed by maximizing nongaussianity, likelihood, or minimizing mutual information. This paper used the Infomax ICA algorithm using the EEGLAB toolbox [8]. The algorithm is based on the idea of maximizing the entropy of the data [9]. The independent component features map is further used for spectrogram generation of two different types: ERSP and ITC. Time-frequency transformation of the data to obtain ERSP and ITC has been completed using EEGLAB, a toolbox that provides a user-friendly interface for physiological data analysis. Besides time-frequency transformation [8]. The toolbox also allows applying ICA and filtering techniques.

Spectrograms have been generated by applying the divisive baseline method in ERSP and FFT in both spectrograms using the EEGLAB toolbox. The spectrograms are separated into two classes, the first including target stimuli recording and the second containing only non-target stimuli. The time frame of 825 ms (125 ms stimuli with 100 ms before and 600 ms after stimuli) has been transformed to create a single spectrogram.

ERSP spectrogram is presented by Fig. 1. Baseline method applied spectrogram that describes power (dB) at a given time-frequency using the color. ERSP can be computed using either Wavelet transform or fast Fourier transform (FFT)

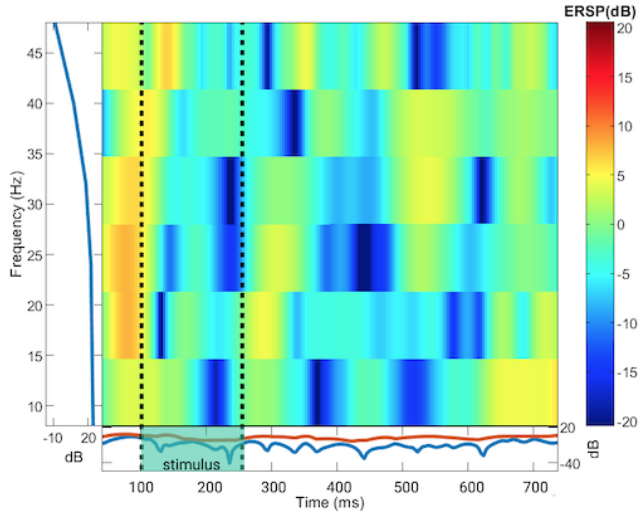

Fig. 1. ERSP spectrogram of the EEG signal: the bottom panel represents the ERSP envelope, where the red line indicates the most positive and the blue line shows the most negative value at each time point; the left panel shows the baseline log spectrum

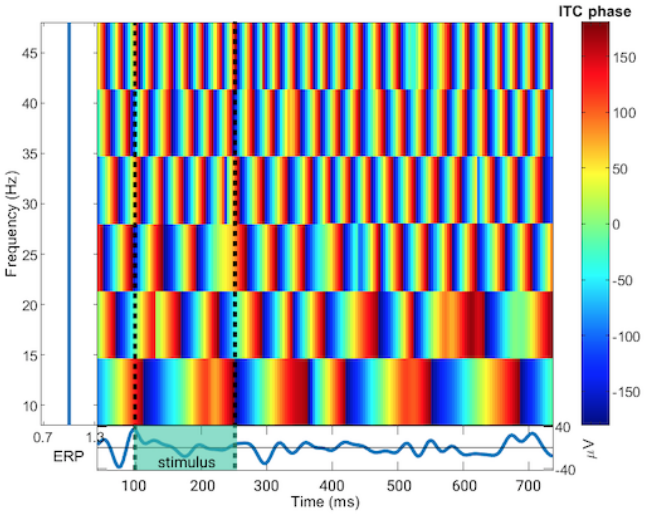

Fig. 2. ITC spectrogram of the EEG signal: the bottom panel represents the time-domain average of the input data; the left panel shows the frequency means ITC

to calculate time-frequency representation. In addition, the baseline method is required to create ERSP due to the absence of visual patterns in ERS spectrograms that are computed without baseline subtraction. When generating an ERSP plot, the ERSP envelope is also usually considered. ERSP envelope is the lowest and the highest dB values at each time i.

The generated ITC spectrogram is shown in Fig. 2. ITC is a frequency representation that describes the consistency of phase across trials. ITC is created by applying the time-frequency transformation to compute phase information of the data. For that purpose, wavelet, short-time Fourier transform

(STFT), or Slepian multi-paper decomposition can be used in the used EEGLAB toolbox. In addition, ITC represents good patterns for P300 speller since it captures the power of ERP components. The training sets consist of 1920 spectrograms for each type. The test sets contain 656 ITC spectrograms and 656 ERSP spectrograms. The data is perfectly balanced, meaning that 50% of data represents the target class and 50% is non-target.

C. Convolutional Neural Network

The classification of the generated spectrograms is performed using a CNN-based model. Over the last years, CNN has started being applied for brain signals classification. For instance, a CNN-based classifier can be used for successful seizure prediction when processing multi-channel EEG signal [10].

Identification of target P300 components can also be performed using a CNN-based model. For instance, a P300 speller with applied CNN along with batch normalization, achieved 96.77% of accuracy for one subject and 93.3% for another [11]. This result, however, shows that the classification was subject-dependent, and both training and test datasets should be collected from the same user. In this work, to achieve the subject-independency of the proposed classifier, it is trained on the merged dataset from all of the eight subjects.

The proposed classifier is based on a classical 2D CNN architecture. The proposed structure has been inherited from the classical image-processing CNN model AlexNet [12]. The spectrograms are compressed to the size of 51×46 RGB images, which are fed to the CNN. The CNN is presented in Fig. **??**. The first layer of the network is a convolution layer, which is followed by the maximum pooling layer. Each hidden convolutional layer has uses rectified linear unit (ReLU) as an activation function. The ReLU is calculated as

$$\mathrm{ReLU}(\tilde{X}_i) = \max(0, \tilde{X}_i), \tag{1}$$

where $\tilde{X}_i$ is the feature map resulted from the previous layer.

Each layer extracts features simultaneously on time and frequency. The first convolution layer uses 32 kernels of size 26×26 and a padding $P = 1$ for calculating the height of the output feature map. As a result, the output of the first convolution layer is a feature map of size $26 \times 23 \times 32$. The pooling layers use 2×2 kernel to reduce the dimension of the feature map.

There are two combinations of convolution and pooling layers, which are connected sequentially for efficient dimensionality reduction. The second convolution layer uses 7×7 kernel. The kernel size is decided to be of equal width and height for minimizing the loss of information in both time and frequency dimensions.

The second pooling layer is followed by the linear fully-connected layers, which transform the feature map as a single vector. The last layer's activation function is a softmax function, calculated as

$$\mathrm{softmax}(\tilde{X}_i) = \frac{e^{x_{ij}}}{\sum_{k=1}^{K} e^{x_{ik}}}, \tag{2}$$

where exponential of each data point x_{ij} is normalized by the sum of exponentials of all K data points of the feature map $\tilde{X}_i$. The softmax normalizes the output of the last layer, translating the output values into a probability distribution. As a result, the network outputs a vector of two elements, where each element represents the probability of the input EEG specrtogram either containing the target P300 component or not. When the spectrogram is defined to be a target response to the stimulus, the following condition is met:

$$P(X|y = 1) > P(X|y = -1), \tag{3}$$

where X denotes the current spectrogram and y is the label, showing whether it contains target P300 component ($y = 1$) or not ($y = -1$).

The same CNN architecture is used for two models: CNN-1 and CNN-2. CNN-1 is trained on the training set of ERSP spectrograms, while CNN-2 is trained on ITC spectrograms.

The training sets consist of data collected from six subjects, which are A01, A02, A03, A04, A05, A07. Thus, 50% of the training data was collected from bulbar onset ALS patients, and the same amount was provided from spinal onset ALS patients (see Table I). The rest two subjects (A06, A08) are used for the test.

D. Ensemble Voting

Over the last years, ensemble learning has become a popular technique for features classification in BCI research. The designed CNN models are combined in one ensemble voter. This is done to achieve more trusted results. Ensemble learning can combine different types of classifiers in one model. For instance, support vector machine (SVM), stepwise linear discriminant analysis (SWLDA), and CNN can be combined for features classification in P300 speller [13]. Apart from that, a fusion of the same type of classifiers with different hyperparameters also provides more trusted results. Combining several CNN models in one ensemble voter can provide high accuracy for P300 component recognition [14]. In this work, two CNN models with the same architecture are trained on different types of spectrograms. Both types (ERSP and ITC) are obtained from the same EEG time-series data. The models are combined in an ensemble voting classifier CNN-ENS, as presented in Fig 3.

The classification result of the ensemble-averaged voting model is the average value of the inner CNN classifiers, which is computed as

$$P(X|y = 1) = \frac{P_{\text{CNN-1}}(X|y = 1) + P_{\text{CNN-2}}(X|y = 1)}{2}, \tag{4}$$

where $P_{\text{CNN-1}}(X|y = 1)$ is the prediction of CNN-1 and $P_{\text{CNN-2}}(X|y = 1)$ is the prediction of CNN-2. It is assumed that the proposed CNN-ENS model can provide more stable

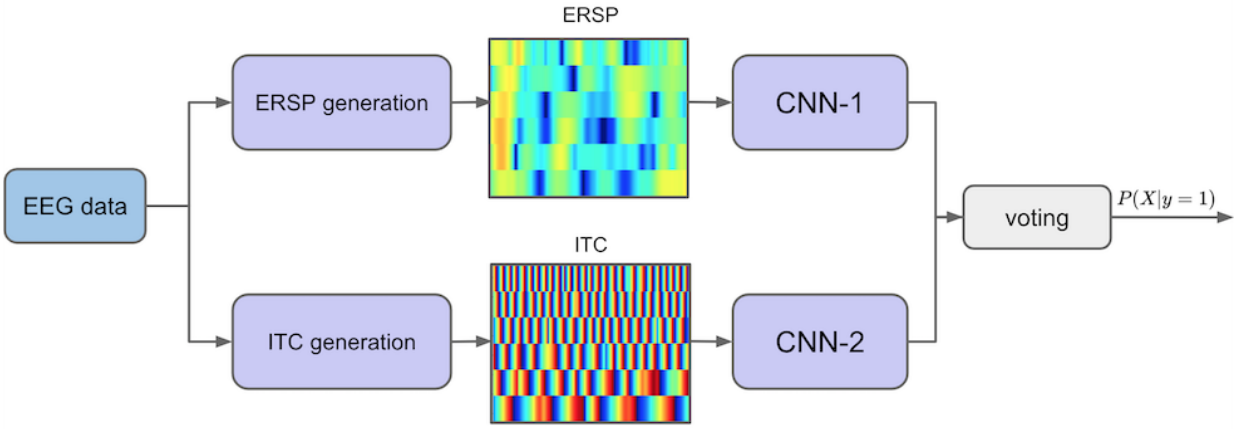

Fig. 3. CNN-ENS ensemble voting model's architecture

and trusted results compared to the standalone classifiers (CNN-1 and CNN-2).

III. Results and Discussions

A. Performance Evaluation

The proposed models are compared by using several metrics. The first one is the accuracy, which is calculated as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

where TP and TN are the numbers of true positives and true negatively predicted spectrograms respectively. The predictions which have been done wrong are false positive (FP) and false negative (FN) results.

In order to evaluate models independently from the distribution of the dataset, F-score is calculated as

$$\text{F-score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}. \tag{6}$$

F-score is also called F-measure or F1-score. The meaning of this metric can be understood if it is represented in terms of recall and precision values as their harmonic mean:

$$\text{F-score} = \frac{2(\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}. \tag{7}$$

Precision value indicates that the spectrogram labelled as target response is target response indeed. It is computed as

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{8}$$

The recall value shows whether the number of FN is low, and it is calculated as

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{9}$$

The proposed models are compared with the baseline method. Classical linear discriminant analysis (LDA) [15] model has been chosen as a baseline method. LDA is one of the most popular classifiers in BCI research, as it is computationally efficient and provides robust results. It can be applied to either supervised or unsupervised P300 speller [16].

It is assumed that the covariance matrices of each class are identical and full rank matrices in LDA, which results in a linear structure when using Bayes' rule. Different solving methods can be applied for LDA implementation, such as singular value decomposition (SVD), eigenvalue decomposition (ED), or least-squares solution (LSS). SVD is more preferrable for a large number of features, thus in addition to LDA, SVD was applied for our data.

Another baseline method chosen is stepwise linear discriminant analysis (SWLDA). SWLDA is more efficient for feature dimension reduction than LDA when applied to single-trial data in P300 speller [17]. SWLDA uses ordinary least squares regression to weight input features in order to predict labels of two classes.

B. Validation and Test Results

There are three proposed models compared. The first one is CNN-1, which is trained on the set of ERSP spectrograms. CNN-2 is trained and tested on ITC plots. The third model is the ensemble voter CNN-ENS, which combines both of the above-mentioned models.

There are 30 training epochs used for training and validation of the proposed models. The models are validated using K-fold validation. K is chosen to be equal to 10.

The loss function chosen for the given binary classification problem is the log loss. Before calculating the loss, the output labels are re-scaled from $-1, 1$ to $0, 1$, where 0 indicates the non-target spectrogram, while 1 still indicates target class. As a result, the new labels are obtained as

$$\hat{y} \in \{0, 1\}. \tag{10}$$

The log loss for each spectrogram X_i is calculated as

TABLE II
VALIDATION AND TEST RESULTS

Model	Data	Validation		Test	
		Accuracy (%)	F-score (%)	Accuracy (%)	F-score (%)
LDA	time-series EEG	83.36	82.64	77.45	76.99
SWLDA	time-series EEG	84.59	84.95	78.03	77.54
CNN-1	ERSP	87.34	88.02	81.98	81.69
CNN-2	ITC	85.61	84.97	78.84	78.36
CNN-ENS	ERSP + ITC	88.05	88.11	81.98	81.55

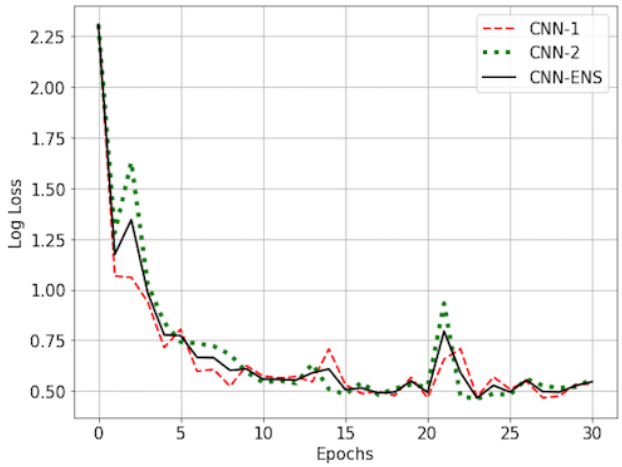

Fig. 4. Training loss provided by CNN-1, CNN-2, CNN-ENS

$$\text{Loss}_i = \hat{y}_i \log(P(X_i|\hat{y}_i) + (1-\hat{y}_i)\log(1-P(X_i|\hat{y}_i), \quad (11)$$

where $\hat{y}_i$ is the re-scaled true label of the spectrogram X_i.

Due to the fact that the network processes spectrograms in batches, the final loss function is calculated as

$$\text{LogLoss} = \frac{1}{N}\sum_{i=1}^{N}\text{Loss}_i, \quad (12)$$

where N is the number of spectrograms in each batch. In this paper, $N = 64$, meaning that there are 64 spectrograms used at each epoch for training and validation.

The log loss obtained during the training process is presented in Fig. 4. It is observable that the loss decreases with the number of epochs, meaning that all of the models are learning properly. It is seen that the loss provided by CNN-ENS is something between the result of CNN-1 and CNN-2, as it considers both types of spectrograms.

The results obtained during the validation and training are presented in Table II. It is seen that the proposed models generally achieve about 80% of accuracy on test data. The proposed model provides high performance if compared with other EEG spectrogram classification methods. For instance, the CNN model for STFT-based spectrograms for P300 Speller achieved only 75.86% of accuracy, requiring averaging 500 reference EEG signals [18]. Our methodology does not require signal averaging, which makes it more applicable for online P300 spellers in the future.

In addition to that, the proposed methodology is compared to the time-series data classification using LDA and SWLDA as seen from Table II. Spectrogram representation turns out to be more efficient on both validation and test. SWLDA provides slightly better results than LDA, but still, it reaches only 77.54% of F-score on a test, while the classification of ITC and ERSP spectrograms using standalone CNN achieves 78.36% and 81.69% respectively.

According to the experimental results, ERSP seems to be more representative than ITC, as CNN-1 reaches almost 82% on training data, while CNN-2 with the same architecture achieves less than 79% of accuracy when tested on ITC spectrograms.

Slightly better results on validation were provided by an ensemble of two models. Usage of both spectrograms enables to achieve 88.05% of accuracy. The validation for ERSP reaches 87.34%, while ITC provides only 85.61%. Thus, it can be assumed that ensemble learning may be more efficient for some data distributions. Nevertheless, it is observable that the test accuracy provided by the CNN-ENS does not outperform CNN-1, meaning that there have been no test cases, on which ITC recognition has positively influenced the final result. Nevertheless, the results show that the ensemble voting model can be more efficient for P300 component identification.

Even though ensemble learning provides only a slight improvement, it evokes the idea of using dual-input CNN architecture with a fusion of features in the future. Such methodology has been applied for different spectrograms classification in various areas, such as sound signals classification [19] etc.

IV. CONCLUSIONS

In this paper, two different spectrogram types have been generated from 8-channel EEG data of ALS patients. ERSP and ITC spectrograms have been fed to two identical CNN models (CNN-1 and CNN-2) and the ensemble model (CNN-ENS). By comparing the results provided by CNN-1 and CNN-2, it turned out that ERSP provides better features representation than ITC. The ensemble classifier achieved slightly better results on validation, meaning that the usage of both ERSP and ITC may provide better classification results.

Nevertheless, to summarize, results show that spectrogram representation of EEG can be efficiently used for subject-independent P300 speller for ALS patients. In addition to that, both bulbar onset and spinal onset type ALS patients' EEG signals can be efficiently represented as ERSP and ITC spectrograms for the same classification model.

One of the limitations of the given study is that the CNN-ENS model requires more computational resources to generate and process two different spectrograms. Parallel computing may be effectively applied for generating both ERSP and ITC at the same time to overcome this problem. Apart from the possible usage of parallel computing, the further direction of the proposed topic considers dataset extension. It is planned to use data from patients with cerebral palsy or peripheral neuropathy to verify the proposed methods for other types of patients. The improvement of the CNN architecture is also planned to be completed by adding residual blocks [20] and features fusion. It is assumed that using both feature maps from ERSP and ITC in a single dual-input architecture may provide better results than the ensemble model.

V. Acknowledgements

This work was supported by the Faculty Development Competitive Research Grants Program of Nazarbayev University under grant number 110119FD4525 (SOE2019007) and under grant number 021220FD0251 (SEDS2021002).

References

[1] A. K. Singh, Y. K. Wang, J. T. King, and C. T. Lin, "Extended interaction with a BCI video game changes resting-state brain activity," *IEEE Trans. Cogn. Dev. Syst.*, vol. 12, no. 4, pp. 809–823, 2020.

[2] H. I. Aly, S. Youssef, and C. Fathy, "Hybrid brain computer interface for movement control of upper limb prostheses," in *Proc. 2018 Int. Conf. Biomed. Eng. Appl. (ICBEA)*, 2018, pp. 1–6.

[3] N. Hasan, M. M. Hasan, and M. A. Alim, "Design of EEG based wheel chair by using color stimuli and rhythm analysis," in *Proc. 2019 1st Int. Conf. Adv. Sci., Eng. and Robot. Tech. (ICASERT)*, 2019, pp. 1–4.

[4] L. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroenceph. Clin. Neurophysiol.*, vol. 70, pp. 510–523, 1998.

[5] S. Makeig, "Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones," *Electroencephalogr. Clin. Neurophysiol.*, vol. 86, pp. 283–293, 1993.

[6] *P300 speller with Patients with ALS Database*. BCI Horizon 2020, 2020. [Online]. Available: http://bnci-horizon-2020.eu/database/data-sets.

[7] L. Xu, T. Liu, L. Liu, X. Yao, L. Chen, D. Fan, S. Zhan, and S. Wang, "Global variation in prevalence and incidence of amyotrophic lateral sclerosis: A systematic review and meta-analysis," *J. Neurol.*, vol. 267, pp. 944–953, 2020.

[8] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[9] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. Wiley, 2001, ISBN: 0-471-40540-X. DOI: 10.1002/0471221317.

[10] A. R. Ozcan and S. Erturk, "Seizure prediction in scalp EEG using 3D convolutional neural networks with an image-based approach," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 27, no. 11, pp. 2284–2293, 2019.

[11] Z. Lu, Q. Li, N. Gao, T. Wang, J. Yang, and O. Bai, "A convolutional neural network based on batch normalization and residual block for P300 signal detection of P300-speller system," in *Proc. 2019 IEEE Int. Conf. Mechatron. Autom. (ICMA)*, 2019, pp. 2303–2308.

[12] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *J. Neural Inf. Proces. Syst.*, vol. 25, pp. 1106–1114, Jan. 2012.

[13] T. Takeichi, T. Yoshikawa, and T. Furuhashi, "Detecting P300 potentials using weighted ensemble learning," in *Proc. 2018 Jt. 10th Int. Conf. Soft Comput. Intell. Syst. 19th Int. Symp. Adv. Intell. Syst. (SCIS-ISIS)*, Dec. 2018, pp. 950–954.

[14] S. Kundu and S. Ari, "Fusion of convolutional neural networks for P300 based character recognition," in *Proc. 2019 Int. Conf. Inf. Technol. (ICIT)*, Dec. 2019, pp. 155–159.

[15] R. D. Fisher and P. Langley, "Methods of conceptual clustering and their relation to numerical taxonomy," *Artif. Intell. Stat.*, vol. 18, pp. 77–116, 1986.

[16] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K. R. Müller, "Toward unsupervised adaptation of LDA for brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 587–597, 2011.

[17] X. Xiao, M. Xu, Y. Wang, T.-P. Jung, and D. Ming, "A comparison of classification methods for recognizing single-trial P300 in brain-computer interfaces," in *2019 41st Ann. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2019, pp. 3032–3035.

[18] H. Meng, H. Wei, T. Yan, and W. Zhou, "P300 detection with adaptive filtering and EEG spectrogram graph," in *Proc. 2019 IEEE Int. Conf. Mechatron. Autom. (ICMA)*, 2019, pp. 1570–1575.

[19] B. Rocha, D. Pessoa, A. Marques, P. de Carvalho, and R. P. Paiva, "Automatic classification of adventitious respiratory sounds: A (un)solved problem?" *Sensors*, vol. 21, Dec. 2020.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.