2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340033

Primary color decoding using deep learning on source reconstructed EEG signal responses

Simen Fløtaker[1], Andres Soler[1] and Marta Molinas[1]

***Abstract*— The brain's response to visual stimuli of different colors might be used in a brain-computer interface (BCI) paradigm, for letting a user control their surroundings by looking at specific colors. Allowing the user to control certain elements in its environment, such as lighting and doors, by looking at corresponding signs of different colors could serve as an intuitive interface. This paper presents work on the development of an intra-subject classifier for red, green, and blue (RGB) visual evoked potentials (VEPs) in recordings performed with an electroencephalogram (EEG). Three deep neural networks (DNNs), proposed in earlier papers, were employed and tested for data in source- and electrode space. All the tests performed in electrode space yielded better results than those in source space. The best classifier yielded an accuracy of 77% averaged over all subjects, with the best subject having an accuracy of 96%.**

***Clinical relevance*— This paper demonstrates that deep learning can be used to classify between red, green and blue visual evoked potentials in EEG recordings with an average accuracy of 77%.**

I. INTRODUCTION

Brain-computer interfaces (BCIs) are systems built to let the user control devices with their brain activity. Such systems can be of great assistance for persons with physical disabilities, as an alternative to traditional systems, which often require physical interaction with the device. For a BCI to be implemented, it requires some form of measurement of brain activity. One common way of doing this is electroencephalography (EEG). EEG can be non-invasive and it can meet high real-time demands, making it a suitable component for a BCI [1].

After recording the brain activity, the BCI also has to interpret that data. This often involves classifying the data into a set of classes, each corresponding to a desired action of the BCI. Performing this classification is a crucial component of the BCI. Without a robust classification method, the actions performed by the BCI might be spurious, which is unacceptable for most control systems. Creating a robust classifier for any signal requires firstly identifying the features of the signal that are relevant to the task, and secondly performing a classification based on these features. Traditionally, feature extraction is done manually and the classification by machine learning (ML). Understanding what features are relevant for which tasks may require expertise in the field, and can be very difficult for novel tasks. Deep learning is an interesting alternative to traditional ML, as it learns both features and classification from data [2]. This facilitates finding novel features for any task, as well as lessening the need for expertise in the field.

For any BCI paradigm, one has to establish a set of brain activities the user should exert, each different enough to be distinguishable. Eliciting such signals may be nonintuitive and difficult for the user [1]. Moreover, Allison et al. [1] point out that the signals for different users might be very different, even if attempting to elicit the same brain activity. As humans, our response to colors is very well-trained, we can identify whether something is blue or red without thinking. Since color vision is such a primal part of human life and we easily distinguish between colors, it is natural to assume that color stimuli elicit distinguishable brain activity. A classifier able to separate what color a subject is looking at could be of use in a BCI. For instance, looking at signs of different colors could allow control of the user's environment, such as opening and closing doors, and turning on and off lights.

Several previous studies have explored classification of brain activity in subjects visually stimulated with red, green and blue (RGB) colors. One study achieved an accuracy of 58%, for a naive Bayes RGB classifier [3]. In [4], the same dataset used in this study was used to train and test machine learning classifiers. Their best results were obtained with a minimum distance to mean with geodesic filtering (Fg-MDM) Riemannian classifier, yielding an average accuracy of 74.48% per subject. In this study, in an attempt to better untangle the encoded colors, source reconstruction, a method for estimating the magnitude and location of neural activity in the brain from EEG signals, was used in combination with deep learning.

This paper is structured into four main sections: This introduction, material and methods, results, and discussion. The dataset used in this paper, the source reconstruction, and classification methods are all described in the section materials and methods. Finally, a conclusion is provided. A more exhaustive report on this work is available online [5].

II. MATERIAL AND METHODS

A. DATASET

The classifiers developed in this work were trained and tested on a dataset where the participants were exposed to primary colors (RGB). The colors were displayed on a screen in front of the participant for intervals of 1.3 seconds, in random order with 140 repetitions for each color. Between each repetition, a gray screen with a cross in the middle was displayed for a random interval of 1.3-1.6 seconds. This

[1]Department of Engineering Cybernetics, Norwegian University of Science and Technology, Norway. simenpf@stud.ntnu.no, andres.f.soler.guevara@ntnu.no, marta.molinas@ntnu.no

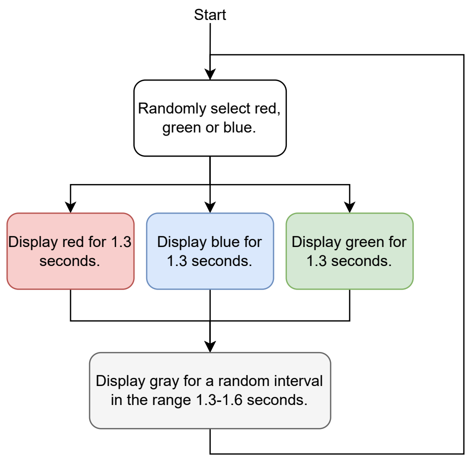

Fig. 1. The stimulus protocol during EEG recording.

protocol is illustrated in Fig. 1. The dataset consists of 60-channel EEG recordings during color presentation and structural MRI from 31 participants (10 females) with an average age of 28.8 (sd 7.4) years old, the participants had normal or corrected-to-normal vision without color impairments. The study was carried out in accordance with the Declaration of Helsinki and all participants provided their informed consent prior to participation. The study was approved by the Data Protection Authority (NSD, reference number 968653). The dataset was recorded at the Aalto NeuroImaging facility of Aalto University.

B. PREPROCESSING

The raw EEG recordings were preprocessed before being used to train and test the neural network classifiers. A notch filter of 50Hz was applied, in order to reject the interference from the powerline. A bandpass filter with the frequency range 0.1-45Hz was applied, in order to filter out frequencies not of interest. After the filtering, the data were downsampled to 200Hz. The recordings were split into separate epochs, one for each stimulus event. The epoch interval was chosen to be from -0.2 seconds before the stimulus to 1.25 seconds after. Baseline correction was applied, by calculating the mean of the 0.2 seconds of data from all channels and then subtracting these means from their respective channels throughout the whole epoch. Blinking artifacts were detected by a peak-finding algorithm. Epochs were discarded if a blink artifact was found within a 200ms interval centered around the onset of the stimulus. Signal-space projection (SSP) was employed to reduce the remaining blink artifacts. A criterion for the maximal acceptable peak-to-peak amplitude of 150 µV within each epoch was set. Thus, any epoch where the difference between the maximal and minimal value for at least one EEG channel was larger than 150 µV, was discarded. All preprocessing was done using MNE-Python [6].

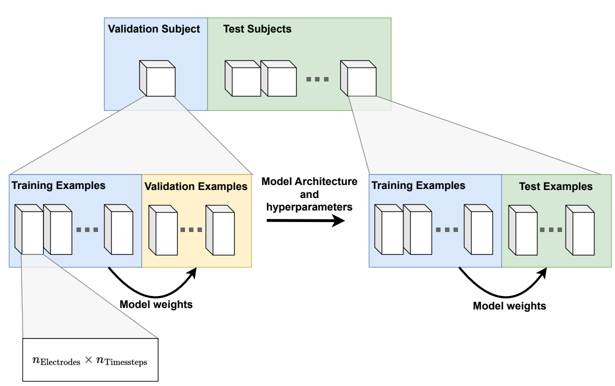

Fig. 2. The structure of the dataset. The validation subject is used to test different hyperparameters and architectures, and the best choice is selected for testing. The validation subject (sub-18), was randomly selected.

C. SOURCE RECONSTRUCTION

The forward model was created using individual magnetic resonance imaging (MRI) data for each subject. Coregistration was manually performed for each subject, such that digitized electrode positions were best transformed into the MRI frame for the forward modeling. A boundary element model (BEM) was used to define the conduction of the brain volume. The sources were distributed on the surface of the white matter. The inverse problem was then solved with the dSPM method [7]. The source space data were aggregated into regions of interest (ROIs) using the automatic parcellation of the brain volume proposed by [8], resulting in 150 dipoles (75 in each hemisphere). In order to aggregate the sources in a ROI into one single value, two steps were taken: First, find the sign of the value for each source and select the sign most represented as the dominant sign. Flip all source values that do not have the dominant sign. Use the mean of all resulting source values as the ROI value. For most applications, the direction of the dipoles is not relevant [9], rather the amplitude is the important information. By flipping the signs, an average value of the amplitudes in the ROI is obtained, avoiding the cancellation of opposing signs during the averaging. The source reconstruction and parcellation was done using MNE-Python [6].

D. CLASSIFICATION METHODS

The work was focused on exploring EEG classification using source space representation. Since the data is of a similar nature in both source- and electrode space ($n_{\text{channels}} \times n_{\text{times}}$, with n_{channels} being the number of dipoles or electrodes for source- and electrode space respectively), the same neural network architectures can be used for both types of data, by modifying the input layer of the network. Three neural networks were employed in this work, using Keras [10]:

- **Shallow EEG-GCNN** (Graph Convolutional Neural Network), [11]
- **EEGNet** (Convolutional Neural Network), [12]
- **Deep ConvNet** (Convolutional Neural Network), [13]

All three networks were implemented with the same hyperparameters as in the papers they were originally proposed (except for some minor modifications necessary for integration). The graph convolutional neural network (GCNN) uses an adjacency matrix to represent the data structure as a graph. Each dipole is treated as a node and the time series of that dipole is treated as its feature vector. The edges between node i and j is represented by the value of element a_{ij} in the adjacency matrix. In this work, if the ROIs represented by two nodes i and j share a border, a_{ij} was set to 1, otherwise it would be set to 0.

Only intra-subject classifiers have been explored in this work. Hence, each classifier had to be trained and tested on data from only one subject. To explore different configurations and hyperparameters before testing, the following routine was developed: The subjects are randomly divided into two groups: validation subject and test subjects. The dataset from the validation subject are each randomly segmented into its own training set and validation set. The datasets from the test subjects are each randomly segmented into a training set and a test set. Fig.2 depicts this structure. In practice, the segmentation within each subject was done with cross-validation. The purpose of separating into validation subject and test subjects is to explore different DNN architectures and hyperparameter configurations. Different configurations can be trained on the validation subject, and then evaluated on its validation set. Several configurations can be found by iterating this process. However, the accuracy found for the validation subject must be considered overly optimistic due to possible overfitting. So the final configuration is tested on the test subject, to see how well it generalizes.

Using only a subset of ROIs/electrodes for classification may be beneficial. Some regions of the brain may be more descriptive than others for the task at hand, thus using only those could reduce the overall signal-to-noise ratio in the data, and improve the classification. For instance, since the aim is to discriminate visual evoked potentials (VEPs), the occipital lobe might be such a region, as it is the brain area that interprets visual stimuli [14]. No optimal selection was performed to select a certain set of ROIs/electrodes in this work, but two different configurations were tested: using all ROIs/electrodes and using a selected set of ROIs/electrodes. In source space, the selected set was all 24 ROIs of the occipital lobe. In electrode space, the selected set was eight electrodes placed in the vicinity of the occipital lobe.

III. RESULTS

EEGNet and deep ConvNet (DCN) classifiers were trained and tested in both source- and electrode space. All these classifiers were developed for both channel configurations (all channels and selected channels). The GCNN was only built for source space using all ROIs. The results presented are the accuracies and standard deviations of intra-subject classifiers tested using a 5-fold cross-validation, those accuracies are presented in Table II. Two trends can be observed from the results: DCN performs better than EEGNet, and electrode space classifiers tend to perform better than source space classifiers. For all classifiers, especially those performing well, there is a noticeable difference between the best and the worst-performing subjects. This can be partially explained by some subjects not having correct behavior during the EEG recording. The subjects were observed during the recording, and notes were taken of some subjects being sleepy or moving excessively. The results show that these subjects tend to perform worse than the average. In addition, for one subject two of the EEG channels located in the vicinity of the occipital lobe did not function correctly and delivered no signal. This subject was among the worst-performing subjects in all tests. In a previous study using the same dataset [4], a choice was made to leave out a set of subjects. The motivation for leaving out these subjects was a set of requirements for a session to be allowed in their study, such as correct behavior of the subjects and no flat channels on the visual cortex. For the same subset of subjects used in this previous study, the best results in this paper (using deepConvnet with all electrodes) yield an average accuracy of 84%. Table I compares the results of this study to those reported in [4].

TABLE I
CLASSIFICATION RESULTS USING THE SAME SUBJECT SUBSET AS IN [4]

	This study		Previous study [4]
	Electrode space	**Source Space**	
Best acc.	0.96	0.87	0.93
Average acc.	0.84	0.58	0.75
Average std.	0.04	0.05	0.08
Worst acc.	0.66	0.38	0.54

IV. DISCUSSION

All average performances of EEGNet and DCN were better in electrode space than source space, regardless of electrode and ROI selection. This was not the expected result, seeing as the source space representation theoretically has a higher spatial resolution [9], and thus different conditions should be easier to discriminate. One possible explanation for this unexpected result is that both networks were developed for electrode space classification. Although both representations have a similar nature, it is not necessarily the case that a given DNN architecture is equally suitable for source- and electrode space. When averaging the sources much of the spatial resolution might be lost. It is effectively a spatial downsampling of the source space. Thus, some of the advantages gained in spatial resolution, by using source reconstruction, may be lost. As mentioned, the use of only some specific ROIs may serve as a method for improving classification. This method also has the advantage of reducing data dimensionality. Instead of reducing the data size by averaging into different regions, one could rather select a subset of sources from the entire set of sources (ca. 8000). By selecting certain regions expected or demonstrated to be relevant for decoding color stimuli, one could keep a high spatial resolution in those areas, while also reducing the data size.

The architectures evaluated in this study have been used with their original hyperparameters, the results also serve as an example that certain DNNs can be applicable across different tasks. One of the concerns regarding deep learning,

TABLE II
CLASSIFICATION RESULTS

Classifier	Accuracy best	Accuracy average	SD average	Accuracy worst
DCN, source space (selected, 24)	0.87	0.54	0.05	0.38
DCN, source space (all, 150)	0.79	0.53	0.05	0.43
EEGNet, source space (selected, 24)	0.71	0.48	0.05	0.37
EEGNet, source space (all, 150)	0.68	0.53	0.05	0.42
GCNN, source space (all, 150)	0.47	0.36	0.05	0.29
DCN, electrode space (selected, 8)	**0.96**	**0.77**	**0.04**	**0.46**
DCN, electrode space (all, 60)	0.95	0.72	0.04	0.37
EEGNet, electrode space (selected, 8)	0.92	0.67	0.05	0.41
EEGNet, electrode space (all, 60)	0.88	0.59	0.06	0.34

raised in [2], is the often unjustified selection of parameters in DNNs, making it difficult to rule out that some tuning based on the test set has occurred. That EEGNet and DCN, with their original hyperparameters, classify with higher accuracy than previous studies on RGB stimuli, is further evidence that these architectures are suitable for EEG decoding, and that their parameters are not overfitted to the test sets of the original papers

There are several reasons to believe that further work can achieve better performance than that reported in this project. Both EEGNet and DCN performed reasonably well, however, no modifications have been done to the design or hyperparameters of these architectures. It is reasonable to assume that these architectures could be tailored more specifically to the task of classifying RGB stimuli, and thus achieve better results. Such tailoring could for instance involve testing different numbers and widths of layers in the model.

The results show that choosing a subset of channels in the vicinity of the occipital lobe does yield an increase in performance. In light of this, it seems natural that there is an optimal subset of channels to use. This should be explored by employing a structured search for an optimal channel subset. Not only could this help develop a better-performing classifier, but choosing a subset of channels would also result in less data and fewer electrodes needed. This would make the classifier more applicable to a BCI, where few electrodes and low computational cost are important factors.

Variability of light conditions and color tones were not considered in this study. Towards a BCI implementation the variability of these parameters should taken into account, further studies should clarify the color classification in more naturalistic scenarios. The color vision impairment can be a limitation of the usability of a BCI based on colors, future evaluations on color blind participants could help to clarify the boundaries of this approach.

V. CONCLUSIONS

In conclusion, the results reported in this study suggest that deep learning can be a suitable approach for classifying RGB stimuli. All architectures employed in this project have been implemented with minimal adaption to the task at hand. Thus, the level of accuracy of both EEGNet and DCN suggests that some DNNs can be suitable across different tasks. Moreover, it is reasonable to believe that modifying these architectures for the classification of RGB responses would yield better results.

REFERENCES

[1] B. Z. Allison, C. Neuper, D. S. Tan, and A. Nijholt, "Could anyone use a BCI?" *Human-Computer Interaction Series*, pp. 35–54, 2010.

[2] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, p. 031005, 4 2018.

[3] S. H. Åsly, L. A. Moctezuma, M. Molinas, and M. Gilde, "Towards eeg-based signals classification of RGB color-based stimuli." in *GBCIC*, 2019.

[4] S. L. Ludvigsen, E. H. Buøen, A. Soler, and M. Molinas, "Searching for unique neural descriptors of primary colours in EEG signals: A classification study," *Lecture Notes in Computer Science*, vol. 12960 LNAI, pp. 277–286, 2021.

[5] S. P. Fløtaker, M. Molinas, and A. Soler, "Deep learning for classification of primary color responses in EEG signals using source reconstruction," 2022. [Online]. Available: https://www.researchgate.net/publication/367346031_Deep_learning_for_classification_of_primary_color_responses_in_EEG_signals_using_source_reconstruction

[6] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, no. 267, pp. 1–13, 2013.

[7] A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren, "Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity," *Neuron*, vol. 26, pp. 55–67, 4 2000.

[8] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *NeuroImage*, vol. 53, pp. 1–15, 10 2010.

[9] C. M. Michel and D. Brunet, "EEG source imaging: A practical review of the analysis steps," *Frontiers in Neurology*, vol. 10, p. 325, 4 2019.

[10] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[11] N. Wagh and Y. Varatharajah, "EEG-GCNN: Augmenting electroencephalogram-based neurological disease diagnosis using a domain-guided graph convolutional neural network." *Proceedings of the Machine Learning for Health NeurIPS Workshop*, vol. 136, p. 367–378, 2020.

[12] L. Xu, M. Xu, Z. Ma, al, B. Zang, Y. Lin, Z. Liu, V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, p. 056013, 7 2018.

[13] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23730

[14] M. Bear, B. Connors, and M. Paradiso, *Neuroscience: Exploring the Brain, Enhanced Edition*. Jones & Bartlett Learning, 2020.