

Technical Report

30 March 2019

Passive Head-Mounted Display Music-Listening EEG dataset

~

G. Cattan, P. L. C. Rodrigues, M. Congedo

GIPSA-lab, CNRS, University Grenoble-Alpes, Grenoble INP.
Address : GIPSA-lab, 11 rue des Mathématiques, Grenoble Campus BP46, F-38402, France

Keywords: *Electroencephalography (EEG), Virtual Reality (VR), Passive Head-Mounted Display (PHMD), experiment.*

Abstract - We describe the experimental procedures for a dataset that we have made publicly available at <https://doi.org/10.5281/zenodo.2617084> in mat (Mathworks, Natick, USA) and csv formats. This dataset contains electroencephalographic recordings of 12 subjects listening to music with and without a passive head-mounted display, that is, a head-mounted display which does not include any electronics at the exception of a smartphone. The electroencephalographic headset consisted of 16 electrodes. Data were recorded during a pilot experiment taking place in the GIPSA-lab, Grenoble, France, in 2017 (1). Python code for manipulating the data is available at <https://github.com/plcrodrigues/py.PHMDML.EEG.2017-GIPSA>. The ID of this dataset is *PHMDML.EEG.2017-GIPSA*.

Résumé - Dans ce document, nous décrivons une expérimentation dont les données ont été publiées sur <https://doi.org/10.5281/zenodo.2617084> aux formats mat (Mathworks, Natick, USA) et csv. Cette base de données contient les enregistrements électroencéphalographiques de 12 sujets écoutant de la musique avec et sans casque de réalité virtuelle passif, c'est à dire, un casque n'incluant aucune électronique à l'exception d'un smartphone. Le casque d'électroencéphalographie comportait 16 électrodes. Les données ont été enregistrées durant une étude pilote réalisée au GIPSA-lab, Grenoble, France en 2017 (1). Un exemple d'application en Python pour manipuler les données est disponible sous le lien <https://github.com/plcrodrigues/py.PHMDML.EEG.2017-GIPSA>. L'identifiant de cette base de données est *PHMDML.EEG.2017-GIPSA*.

Introduction

Virtual Reality (VR) is a promising field of application for Brain-Computer Interfaces (BCI) based on electroencephalography (EEG) as, in VR environments, BCI candidates to reduce the distance between a user and the virtual avatar. Passive head-Mounted Devices (PHMD) are VR devices consisting in a plastic or cardboard mask in which we insert a smartphone. They are cheap, affordable and thus may help to populate the VR+BCI technology for the general public, that is, they provide a potentially ubiquitous technology. However, to our knowledge, the impact of PHMD on the quality of EEG signal has never been assessed. We provide EEG recordings of subjects with and without a passive a head-mounted display while listening to classical music. A Fourier analysis (2) of these data has been published in (1). An example of application of this dataset is available online at <https://github.com/plcrodrigues/py.PHMDML.EEG.2017-GIPSA>.

Participants

12 subjects participated to the experiment (3 females), with mean (sd) age 26.25 (2.63). Subjects were volunteers recruited at the University of Grenoble-Alpes. Before the experiment, the subjects were informed that they will be exposed to electromagnetic contamination due to the proximity of a smartphone put in front of their eyes. All participants provided written informed consent confirming the notification of the experimental process, the data management procedures and the right to withdraw from the experiment at any moment.

Material

EEG signals were acquired using a research grade amplifier (g.USBamp, g.tec, Schiedlberg, Austria) and the EC20 cap equipped with 16 wet electrodes (EasyCap, Herrsching am Ammersee, Germany), placed according to the 10-10 international system (**Figure 1**). The locations of the electrodes were FP1, FP2, FC5, FC6, FZ, T7, CZ, T8, P7, P3, PZ, P4, P8, O1, Oz, and O2. The reference was placed on the right earlobe and the ground at the AFZ scalp location. The amplifier was linked by USB connection to the PC where the data were acquired by means of the software OpenVibe (3,4). We acquired the data with no digital filter and a sampling frequency of 512 samples per second.

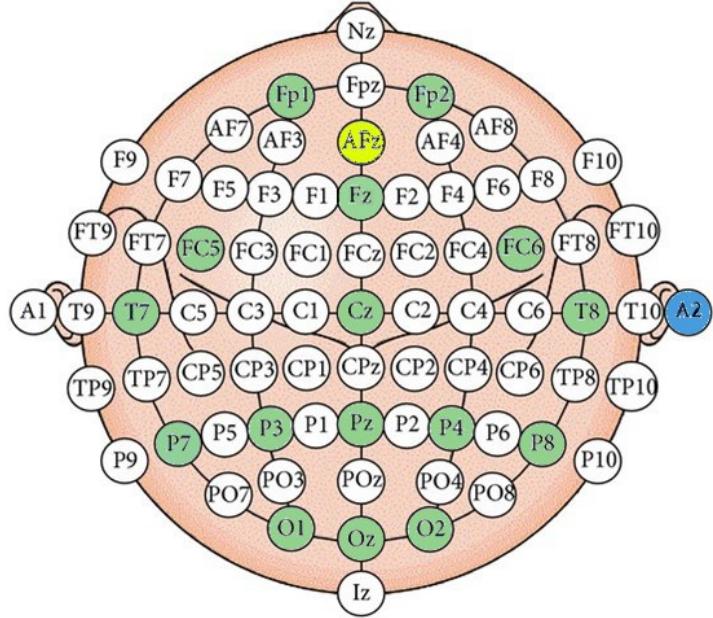


Figure 1. In green, the 16 electrodes placed according to the 10-10 international system. We used AFz (in yellow) as ground and A2 (in blue) as a reference.

We used two identical smartphones in order to quickly switch between the two experimental conditions (with and without PHMD). In both conditions the subject wore a SamsungGear (Samsung, Seoul, South Korea) device (**Figure 2**). In one condition (without PHMD) the smartphone was switched-off, and in the other (with PHMD) it was switched-on. Anything else in the two conditions was identical. The smartphone used as VR devices was a Samsung S6 running under Android OS Nougat. The Specific Absorption Rate (SAR) of the smartphone was 0.382 Watt/Kg (Head) and 0.499 Watt/Kg (Body).



Figure 2. The SamsungGear (Samsung, Seoul, South Korea) consists in a plastic mask in which we insert a smartphone.

Procedures

Subjects were asked to sit on a desk in front of screen at a distance of about 50 cm. In order to mimic real-world usage we did not employ any instrumental noise-reduction device such as a Faraday cage. The EEG cap and the Samsung Gear were then placed on the subject. We continuously swapped the smartphones into the Samsung Gear (**Figure 3**). In both condition the screen of the smartphone was black and a purple marker was stuck on the left part of the screen in both devices. Having one marker might seems unnatural as one eye is looking at something that the other cannot. However, in a pilot study it was established that it was difficult for the subjects to reproduce stereoscopic vision with two markers because small differences in shape and position between the two markers were unavoidable. Additionally, there was a tiny white line on the middle of the switched-on smartphone to mark separation between left and right part of the screen on the running smartphone. This line was hidden by the Samsung Gear when the smartphone was put into it. The luminosity of the screen was comparable in the two conditions. Subjects were asked to focus on the marker and to listen to the music that was diffused during the experiment (Bach Invention from one to ten on harpsichord). The music was presented via the speakers of a personal computer. The marker and the music were introduced to homogenize the mental activity of the subjects during EEG recording. In addition, fixating the purple marker aimed at minimizing eye movement artifacts. The experiment comprised ten blocks. There were five blocks in the condition switched-on and five blocks in the condition switched-off (**Figure 3**). Each block consisted of one minute of EEG data recording with the eyes opened. Hence, a total of ten minutes were recorded for each subject. The sequence of the ten blocks were randomized prior to the experiment for each subject using a random number generator featuring no autocorrelation. This experimental design allows the use of exact randomization test for testing hypotheses (5). For ensuing analysis (1), we tagged the beginning of each block on the EEG using a trigger channel. Note that when the experimental condition was the same between two blocks, we removed and placed the same smartphone within the headset as if it were swapped.

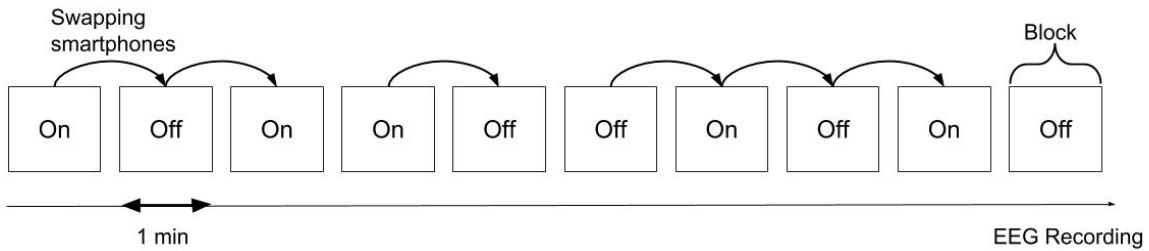


Figure 3. Example of a sequence comporting ten randomized blocs.

Organization of the dataset

For each subject we provide a single *mat* (and *csv*) file (Mathworks, Natick, USA) containing the complete recording of the session. The file is a 2D-matrix where the rows contain the observations at each time sample. Columns 2 to 17 contain the recordings on each of the 16 EEG electrodes. The first column of the matrix represents the timestamp of each observation and column 18 contains the triggers for the experimental condition. The rows in column 18 are filled with zeros, except at the timestamp corresponding to the beginning of a block, when the row gets a value of one (smartphone switched-off) or two (smartphone switched-on). The attribute names of the matrix are provided in the *Header.mat* (or *Header.csv*) file.

We supply an online and open-source example working with Python (6) and using the framework MNE (7,8). This example shows how to download the data using Python. The code shows also how to extract the blocks of given subjects from the dataset and plot the spectral characteristics of the signal.

References

1. Cattan G, Andreev A, Mendoza C, Congedo M. The Impact of Passive Head-Mounted Virtual Reality Devices on the Quality of EEG Signals. In Delft: The Eurographics Association; 2018. Available from: <http://dx.doi.org/10.2312/vrphys.20181064>
2. Congedo M. EEG Source Analysis [Internet] [Habilitation à diriger des recherches]. Université de Grenoble; 2013. Available from: <https://tel.archives-ouvertes.fr/tel-00880483>
3. Renard Y, Lotte F, Gibert G, Congedo M, Maby E, Delannoy V, et al. OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments. Presence Teleoperators Virtual Environ. 2010 Feb 1;19(1):35–53.

4. Arrouët C, Congedo M, Marvie J-E, Lamarche F, Lécuyer A, Arnaldi B. Open-ViBE: A Three Dimensional Platform for Real-Time Neuroscience. *J Neurother*. 2005 Jul 8;9(1):3–25.
5. Edgington ES. Randomization tests. New York: Marcel Dekker; 1980.
6. Rodrigues P. Codes for working with the “Passive Head-Mounted Display Music-Listening EEG” developed at the GIPSA-lab [Internet]. 2019. Available from: <https://github.com/plcrodrigues/py.PHMDML.EEG.2017-GIPSA>
7. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MNE software for processing MEG and EEG data. *NeuroImage*. 2014 Feb 1;86:446–60.
8. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-Python. *Front Neurosci* [Internet]. 2013;7. Available from: <https://www.frontiersin.org/articles/10.3389/fnins.2013.00267/full>

EEGdenoiseNet: A benchmark dataset for end-to-end deep learning solutions of EEG denoising

Haoming Zhang^{1,†}, Mingqi Zhao^{1,2,†}, Chen Wei¹,
Dante Mantini^{2,3}, Zherui Li¹, Quanying Liu^{1,*}

July 29, 2021

¹ Shenzhen Key Laboratory of Smart Healthcare Engineering, Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen 518055, P.R. China

² Movement Control and Neuroplasticity Research Group, KU Leuven, Leuven 3001, Belgium

³ Brain Imaging and Neural Dynamics Research Group, IRCCS San Camillo Hospital, Venice 30126, Italy

† Equal contribution

* Corresponding author liuqy@sustech.edu.cn to Q.L.

Abstract

Deep learning networks are increasingly attracting attention in various fields, including electroencephalography (EEG) signal processing. These models provided comparable performance with that of traditional techniques. At present, however, lacks of well-structured and standardized datasets with specific benchmark limit the development of deep learning solutions for EEG denoising. Here, we present EEGdenoiseNet, a benchmark EEG dataset that is suited for training and testing deep learning-based denoising models, as well as for performance comparisons across models. EEGdenoiseNet contains 4514 clean EEG segments, 3400 ocular artifact segments and 5598 muscular artifact segments, allowing users to synthesize contaminated EEG segments with the ground-truth clean EEG. We used EEGdenoiseNet to evaluate denoising performance of four classical networks (a fully-connected network, a simple and a complex convolution network, and a recurrent neural network). Our analysis suggested that deep learning methods have great potential for EEG denoising even under high noise contamination. Through EEGdenoiseNet, we hope to accelerate the development of the emerging field of deep learning-based EEG denoising. The dataset and code are available at <https://github.com/ncclabsustech/EEGdenoiseNet>.

Keywords: Deep learning, Neural network, EEG dataset, Benchmark dataset, EEG artifact removal, EEG denoising

1 Introduction

Electroencephalography (EEG) solutions permit recording of changes in electrical potential on the scalp, which are generated by neurons in the gray matter. EEG is one of the most important direct and noninvasive approaches for studying brain activity under task and resting conditions. It has been widely used in psychology, neurology and psychiatry research, as well as for brain-computer interface [1, 2, 3, 4, 5, 6].

EEG signals contain not only brain activity, but also a variety of noise and artifacts, including ocular [7], myogenic artifacts [8, 9], and, in rare cases, cardiac artifacts. Therefore, a basic step in using EEG data to study neural activity is denoising or artifact attenuation [10]. Ocular and myogenic artifacts contaminate EEG signals in different ways. The former is often visible as relatively large pulses in the frontal region [11], while the latter frequently appears in the temporal and occipital regions, and has a wide frequency spectrum [9, 12].

Various traditional denoising techniques have been developed to remove artifacts from EEG data, such as regression-based methods, adaptive filter-based methods and blind source separation (BSS)-based methods. Among them, the regression-based method first obtains the noise signal through the noise template, and then subtracts the estimated noise signal from the EEG data to eliminate the artifacts [12, 13, 14, 15]. On the contrary, methods based on adaptive filters rely on dynamically estimating filter coefficients based on the input EEG signal itself, thereby filtering out noise [16, 17]. BSS-based methods decompose the EEG signal into multiple components [18, 19, 20], assign them to neural and artifactual sources, and reconstruct a clean signal by recombining the neural components [9, 12, 21]. However, BSS-based methods can only be used when a large number of electrodes are available, which are not suitable for single-channel denoising.

Deep learning (DL) have been increasingly attracting attention in the past few years [22, 23, 24, 25]. Due to the increase in computing resources, the boosting data size, and the availability of new network architectures and learning algorithms, the performance of DL neural networks has made great breakthroughs, and deep learning has been successfully applied to solve various technical problems, such as image processing [22, 23, 26, 27] and natural language processing [24, 25, 28]. DL methods have begun to be introduced into the field of EEG signal analysis [29], such as EEG-based classification [30, 31, 32], EEG reconstruction [33, 34] and EEG signal generation [35, 36]. Recently, deep learning has also been applied to EEG denoising, providing performance comparable to the traditional denoising method [37, 38, 39, 40].

Deep neural networks can learn the hidden state of neural oscillations in EEG, thereby eliminating fluctuations that are not from real neural activity but from biological artifacts. The performance of deep neural networks fundamentally depends on the size of the training and test datasets; or in other words, it requires big data [41, 42, 43]. A big dataset with the gold standard clean EEG is essential for evaluating newly developed supervised deep learning models. Some EEG datasets have been collected while participants are at rest [44, 45], during cognitive tasks [46, 47, 48], or motor-related tasks [49, 50, 51, 52]. However, none of them are specifically developed for training end-to-end deep learning models for EEG artifact removal. To the best of our knowledge, there is no open EEG dataset suitable for training deep learning

models for EEG denoising. The lack of ground-truth clean EEG data and benchmarks have largely limited the development of DL methods for EEG denoising.

In this study, we present a publicly available structured dataset, named EEGdenoiseNet, which is particularly suitable for deep network-based EEG artifacts attenuation (**Sec 2**). Specifically, the dataset contains 4514 clean EEG segments as ground truth, and 3400 pure EOG segments and 5598 pure EMG segments as ocular artifacts and myogenic artifacts respectively. In addition, we also implement four deep neural networks as benchmarks (**Sec 3**), including a fully-connected neural network (FCNN), a simple convolution neural network (CNN), a complex CNN, and a recurrent neural network (RNN). We train the deep learning models in a supervised end-to-end fashion, and the denosing performance are presented as benchmarks (**Sec 4**).

2 EEGdenoiseNet Dataset

2.1 Data acquisition and preprocessing

Our main goal is to construct a dataset suitable for EEG denoising research based on deep learning networks. In this regard, we downloaded EEG, EOG and EMG data from several publicly available data repositories which were published in previous studies (see Table 1) [53, 54, 55, 56, 57, 58, 59, 60]. These studies have been ethically approved by their respective local ethical committees, and followed the Helsinki Declaration of 1975, revised in 2000.

To generate clean EEG, pure EOG and pure EMG, we firstly preprocessed the data. Then segmented them into 2-second segments. Afterwards, we re-scaled the segments to the same variance. Finally, each segment was visually checked by an expert to ensure they are clean and usable. We set the length of segments to 2 seconds according to the previous knowledge of EEG signals. On the one hand, a 2s segment is long enough to recover the temporal and spectral characteristics of EEG, as well as EOG and EMG. On the other hand, it is difficult to obtain artifact-free EEG segments longer than 2s due to the random eye blinks or movements. The segments in each category have been uploaded to a publicly available repository (<https://github.com/ncclabsustech/EEGdenoiseNet>).

Specifically, for the EEG segments (Figure 1a)[53], the dataset included 52 participants who performed both real and imaginary left and right hand movement task, with 64 channel EEG recorded simultaneously at 512 Hz sampling frequency. For both real and imagined movement task, a participant repeated 2 second baseline and 3 seconds movement with 4.1 to 4.8 second random interval for 20 minutes. The data was band-pass filtered between 1 to 80 Hz, notched at powerline frequency, and then re-sampled to 256 Hz. To obtain the clean EEG as ground truth, the 64-channel EEG signals were processed by ICLLabel, a toolbox to remove EEG artifacts with independent component composition (ICA) [9]. Then the pure EEG signals were segmented into one-dimensional segments of 2 seconds. It is worth noting that, in order to ensure the universality of this data set, we did not construct clean EEG signals with a specific number of channels due to the diversity of EEG caps, but constructed a dataset with single-channel EEG signal.

For the ocular artifact segments (Figure 1*b*), multiple open-access EEG datasets with additional EOG channels are used [54, 55, 56, 57, 58, 59]. The horizontal and vertical raw electroculography (EOG) signals of the datasets are band-pass filtered between 0.3 and 10 Hz, and then re-sampled to 256 Hz. Finally, the EOG signals are segmented into one-dimensional segments of 2 seconds.

For the myogenic artifact segments (Figure 1*c*), a facial electromyography (EMG) dataset is used [60]. We choose facial EMG because they are the main sources of myogenic artifacts. The raw EMG signal is band-pass filtered between 1 to 120 Hz and notched at the powerline frequency, and then resampled to 512 Hz. We resample the EMG to 512 Hz instead of 256 Hz, because the EMG signal is concentrated in the high frequency range, so a higher sampling rate is required (according to the Nyquist sampling theorem). In the end, we extract one-dimensional 2-second EMG segments.

For all three categories, the segments are standardized by subtracting their mean and dividing by their standard deviation, and then visually inspected by an expert. We obtain a total of 4514 EEG segments, 3400 ocular artifact segments, and 5598 myogenic artifact segments. The segments of each category are respectively saved as Matlab matrix files and Python numpy matrix files in the public data repository. Figure 2 shows an example of the clean EEG, horizontal EOG, vertical EOG and EMG.

2.2 Data Usage

The contaminated signals can be generated by linearly mixing the pure EEG segments with EOG or EMG artifact segments, according to Eq. (1) (see Figure 1*c*):

$$y = x + \lambda \cdot n \quad (1)$$

where y denotes the mixed one-dimensional signal of EEG and artifacts; x denotes the clean EEG signal as the ground truth; n denotes (ocular or myogenic) artifacts; λ is a hyperparameter to control the signal-to-noise ratio (SNR) in the contaminated EEG signal y . Specifically, the SNR of the contaminated segment can be adjusted by changing the parameter λ according to Eq. (2):

$$SNR = 10 \log \frac{RMS(x)}{RMS(\lambda \cdot n)} \quad (2)$$

in which the Root Mean Squared (RMS) value is defined as Eq. 3:

$$RMS(g) = \sqrt{\frac{1}{N} \sum_{i=1}^N g_i^2} \quad (3)$$

where N denotes the number of temporal samples in the segment g , and g_i denotes the i^{th} sample of a segment g . Notably, lower λ represents higher SNR, as less EOG or EMG artifacts are added in the EEG signal. In return, lower SNR means higher noise level. According to previous studies, the SNR of EEG contaminated by ocular artifacts is usually ranging from

-7dB to 2dB [61], while the SNR of EEG contaminated by myogenic artifacts are between -7dB and 4dB [62, 63].

In this way, we obtain a pair of EEG data (x, y) . To train the end-to-end deep learning methods for EEG denoising, the clean EEG x can be regarded as the ground truth, and the contaminated EEG y can be used as the inputs.

3 Benchmarking deep learning algorithms

The second goal of this study is to provide a set of benchmark algorithms. We train four standard deep-learning neural networks, then validate the networks. The evaluation metrics can be used as benchmarks for the EEG denoising algorithms.

3.1 Generating semi-synthetic data

The semi-synthetic ocular artifact contaminated signals are from 3400 EEG segments and 3400 ocular artifact segments, with 80% for generating the training set, 10% for generating the validation set, and 10% for generating the test set [64]. Each set were generated by randomly linearly mixing EEG segments and ocular artifact segments according to section 2.2, with SNR raging from ten different SNR levels (-7dB, -6dB, -5dB, -4dB, -3dB, -2dB, -1dB, 0dB, 1dB, 2dB). This procedure expanded the data size of each set to ten times. The EEG segments are treated as ground truth, and the corresponding mixed segments are treated as contaminated EEG.

The myogenic artifacts contaminated signals come from 4514 EEG segments and 5598 myogenic artifact segments. To match the sampling frequency of EEG segments with myogenic artifact segments, we upsample the EEG segments to 512Hz. To match the number of EEG segments with myogenic artifact segments, we randomly reuse some EEG segments. We mix the EEG segments and the myogenic artifact segments as Eq.(1) to generate the training data, test data, and validation data. Likewise, the EEG segments are treated as ground truth, and the corresponding mixed segments are treated as contaminated EEG.

3.2 Network architectures

3.2.1 Fully-connected Neural Network

A fully-connected network with four hidden layers using ReLu as activation function is provided as a benchmarking algorithm (Figure 4a). The number of neurons in each layer is equal to the number of temporal samples of the input signal (*i.e.*, 512 for ocular artifact reduction, and 1024 for myogenic artifact reduction). Dropout regularization [65] is used to reduce overfitting. The contaminated EEG is fed in from the first layer of the neural network, and then the denoised EEG is output from the last layer.

3.2.2 Simple Convolution Neural Network

A simple convolution network is implemented (Figure 4b). The simple CNN consists of four 1D-convolution layers with small 1×3 kernels, 1 stride, and 64 feature maps (k3n64s1). Each

1D-convolution layer is followed by a batch-normalization (BN) layer [66] and a ReLu activation function. To reconstruct the signal, the last convolutional layer is followed by a flatten layer and a dense layer with 512 or 1024 neurons as outputs (the same as the input).

3.2.3 Complex Convolution Neural Network

An one-dimensional residual convolutional neural network (1D-ResCNN), adapted from [38], is implemented (Figure 4c). Compared with simple CNN, the 1D-ResCNN has a more complex structure, so it is called complex CNN. The main difference between them is that a ResNet with skip-layer connections is added into the complex CNN to avoid gradient explosion so that we can train a deeper network to obtain better feature extraction capabilities [23]. To extract multi-scale features, we repeatedly stack residual blocks, using 1×3 , 1×5 , 1×7 multi-scale convolutional kernels twice and arranging three sets of residual blocks branches in parallel [27, 67].

3.2.4 Recurrent Neural network

A Long Short-Term Memory (LSTM) network (Figure 4d), adapted from [68], is regard as the benchmark of recurrent neural networks (RNNs). LSTM can learn long-term dependencies, which may help distinguish the long-term features in noise and EEG signals. Each EEG sample is sequentially input to LSTM cells, and the output is obtained from the state of each cell through a fully-connected network. This RNN model is initialised to have 1 hidden state, and the output network is a three-layer fully-connected network with ReLu activation function, dropout regularization, and 512 or 1024 neurons per layer.

3.3 Learning process

In order to facilitate the learning procedure, we normalized the input contaminated EEG segment and the ground-truth EEG segment by dividing the standard deviation of contaminated EEG segment according to Eq. (4):

$$\hat{x} = \frac{x}{\sigma_y}, \quad \hat{y} = \frac{y}{\sigma_y} \quad (4)$$

where σ_y is the standard deviation of the contaminated EEG signal segment y . The standard deviation of each noise segment is saved, so that the magnitude of the denoised EEG segment can be restored by multiplying the network output by the corresponding standard deviation value.

The networks are trained in an end-to-end manner, which means that we input the normalized contaminated EEG segment into the neural networks and then directly output the denoised EEG segment. To this end, the goal of a denoising network is to learn a nonlinear function f that maps the contaminated EEG \hat{y} to the denoised EEG \tilde{x} :

$$\tilde{x} = f(\hat{y}, \boldsymbol{\theta}) \quad (5)$$

where $\hat{y} \in \mathbb{R}^{1 \times T}$ denotes the contaminated EEG segment, $\tilde{x} \in \mathbb{R}^{1 \times T}$ as the output of neural network (the denoised EEG segment), and the vector $\boldsymbol{\theta}$ contains all parameters to be learned.

We use the mean squared error (MSE) as loss function $L_{MSE}(f)$ (see Eq. (6)). The learning process is implemented with gradient descent to minimize the error between the denoised segment and the ground-truth clean segment.

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N \left\| \tilde{x}_i - \hat{x}_i \right\|_2^2 \quad (6)$$

where N denotes the number of temporal samples of segment; \tilde{x}_i denotes i^{th} sample of the output of the neural network; \hat{x}_i denotes the i^{th} sample of the ground truth x .

For ocular artifact removal, we train the FCNN with 60 epochs, RNN with 100 epochs, while the simple CNN and complex CNN models are trained over 40 epochs. For myogenic artifact removal, we train the FCNN and RNN with 60 epochs, while the simple CNN and complex CNN models are trained over 10 epochs. The Adam algorithm is applied to optimize the loss function, and the parameter were set to $\alpha = 5e^{-5}$, $\beta_1 = 0.5$, $\beta_2 = 0.9$. To increase the statistical power, the four networks are trained, validated and tested independently for 10 times with randomly generated datasets via EEGdenoiseNET.

All the four networks are implemented in Python 3.7 with Tensorflow 2.2 library, running on a computer with two NVIDIA Tesla V100 GPUs. The codes for the benchmarking algorithms are publicly available online at Github [69].

3.4 Performance Evaluation as Benchmarks

There are several metrics are used to qualitatively evaluate the performance of networks, including the network convergence, the relative root mean squared error, and the correlation coefficient.

The network convergence is the first index to evaluate the performance of networks, which can provide rich information about the learning procedure and generalization ability. The convergence curve of both training and validating processes are obtained by calculating the averaged loss (in Eq. (6)) with respect to the number of epochs.

We then quantitatively examine the performance of the networks by applying three objective measures to the denoised data [62], including Relative Root Mean Squared Error (RRMSE) in the temporal domain ($RRMSE_{temporal}$, see Eq. (7)), RRMSE in the spectral domain ($RRMSE_{spectral}$, see Eq. (8)) and the correlation coefficient (CC see Eq. (9)).

$$RRMSE_{temporal} = \frac{RMS(f(y) - x)}{RMS(x)} \quad (7)$$

$$RRMSE_{spectral} = \frac{RMS(PSD(f(y)) - PSD(x))}{RMS(PSD(x))} \quad (8)$$

where the function $PSD()$ denotes to the power spectral density of an input segment. The frequency range of PSD is 0-120Hz. The fft-length equal to the length of the input segment.

$$CC = \frac{Cov(f(y), x)}{\sqrt{Var(f(y))Var(x)}} \quad (9)$$

To compare the deep learning methods with the traditional methods, we implement two traditional EEG denoising methods: i) empirical mode decomposition (EMD) and ii) filtering. In the EMD method, the artifactual intrinsic mode functions (IMFs) are defined by the distance metric used in clustering [70]. In the filtering method, the ocular and myogenic artifacts are removed using a high-pass filter (12 Hz) and a band-pass filter (12-40 Hz), respectively. These two traditional methods are tested 10 times with randomly generated datasets. The corresponding codes are available online at <https://github.com/ncclabsustech/Single-Channel-EEG-Denoise>.

4 Results

To give a qualitative overview of the denoising results, we display some sample fragments in the test in the time domain and frequency domain for ocular artifact removal (see Figure 5) and for myogenic artifact removal (see Figure 6). For each network and artifact type, we show two examples: one of the best results (left column) and one of the worst result (right column). Generally, both in ocular and myogenic artifact removal, the artefacts are greatly attenuated, and the noise-free EEG samples are well-reconstructed. The frequency domain diagram shows that the artifacts in the low frequency bands are well detected and attenuated, but the high frequency is affected by residual noise.

The quantitatively results are examined. We first present the convergence of the four networks. The training and validation loss of the networks can show a quantitative overview of the training and validating process. For all the 4 networks and 2 artifact types, the training loss is consistently lower than the validation loss as expected. For the ocular artifact removal (see Figure 7a), the training and validation loss decrease with the increase of epochs. Specifically, the loss of simple CNN and complex CNN drop notably fast and eventually diminish after 20 epochs. The FCNN loss and the RNN loss, however, starting from a relatively high level, remain at a significant level after 20 epochs. For the myogenic artifact removal (see Figure 7b), the training loss of four networks decreased with respect to the number of epochs, similar to the ocular artifact removal. The loss of simple CNN and complex CNN decrease faster during training, but increased during validation. This phenomenon indicates that the two convolutional networks seem not learn the true characteristics of the EMG signal, which means that CNNs suffer from an overfitting problem when removing myogenic artifacts.

We present the quantitative benchmarks ($RRMSE_{temporal}$, $RRMSE_{spectral}$ and CC) from the four networks and the two traditional methods at multiple SNR levels (see Figure 8). Generally, for both ocular and myogenic artifact removal, the performance decreases with the decrease of SNR level. The traditional methods showed higher $RRMSE$ and lower CC compared with the four deep learning networks. The difference of performance is larger at the large noise level (low SNR), while the difference reduces at low noise level (large SNR, e.g. $SNR > 0$). Among the deep learning methods, RNN has the lowest $RRMSE$ and the highest CC for ocular artifact removal (see Figure 8a), and the complex CNN has the lowest $RRMSE$ and the highest CC in myogenic artifact removal.

To further comprehensively compare benchmarks, we separately plot the benchmarks at multiple SNR levels in boxplot (see Figure 9), and conduct ANOVA analyses. For the ocular

artifact removal (see Figure 9a), DL-based methods have significantly better denoising performance compared to two traditional methods, in terms of $RRMSE_{temporal}$, $RRMSE_{spectral}$ and CC ($p < 0.001$ for each of three metrics). Similarly, the DL-based methods outperform traditional methods for myogenic artifact removal ($p < 0.001$ for each of three metrics) (see Figure 9b). In the time domain, $RRMSE_{temporal}$ of RNN is significantly higher than FCNN, simple CNN, and complex CNN ($p = 0.007$, $p < 0.001$, and $p < 0.001$, respectively); FCNN has significantly higher $RRMSE_{temporal}$ than the complex CNN ($p = 0.020$). At the frequency domain, RNN has significantly higher $RRMSE_{spectral}$ than the FCNN and the complex CNN ($p = 0.020$ and $p = 0.006$, respectively). CC of the complex CNN is significantly higher than RNN and FCNN ($p < 0.001$, and $p = 0.011$, respectively). The same effect is shown on simple CNN and RNN ($p = 0.004$).

We finally evaluate the performance of the different methods for different frequency bands, by calculating the average power ratio of each frequency band (delta [1-4 Hz], theta [4-8 Hz], alpha [8-13 Hz], beta [13-30 Hz], and gamma [30-80 Hz] bands) to whole band (1-80 Hz) for ocular artifact removal and myogenic artifact removal (see Table 2 and 3). For the ocular artifact removal (see Table 2), the mix of ocular artifacts increased the power ratio of delta and theta bands, while reduced the ratio of the other bands. The simple CNN showed the closest delta, theta and beta power ratio compared to those of the ground truth, the same effect observed on the complex CNN for theta band, on the EMD for alpha band, on the RNN for gamma band, and on the FCNN for theta and gamma band. In myogenic artifact removal (see Table 3), the add of myogenic artifact increased gamma power ratio and decreased other power ratios. The FCNN showed the closest ratio in beta bands compared to those of the ground truth, the same effect on the EMD for alpha band, on the Complex CNN for theta and gamma bands, on the RNN for delta and alpha bands.

5 Discussion

In this study, we have provided an EEG benchmark dataset, EEGdenoiseNet, for training and testing end-to-end deep learning models. To obtain the ground-truth clean EEG data, the raw EEG data is denoised by ICLLabel [9] and then manually inspected for a double check. Although there are other publicly available EEG datasets, they are not specifically developed for EEG denoising. Instead, they are mainly focused on the resting state study [44, 45], psychological study [46, 47, 48], or motor imaginary or motor tasks [49, 50, 51, 52]. A previous study has offered a semi-simulated dataset for EOG artifact removal, but EMG signals are not included [71]. Effective use of these datasets for DL-based denoising requires extensive EEG background knowledge, including properties of EEG and artifacts, data format conversion, and signal processing. In contrast, the segments in our dataset have been pre-processed, so users can immediately generate a large set of semi-synthetic noisy EEG segments with ground truth for their DL-based denoising models without being distracted by detailed electrophysiological knowledge. With this advantage, our well-structured dataset would greatly promote the development of DL-based EEG denoising.

Another major challenge to compare the performance of different denoising algorithms is the lack of specific benchmarks. The use of standard benchmarks greatly simplify the comparisons of performance across multiple DL models. To fill this gap, we provided a set

of benchmark algorithms along with a standardized EEG dataset. We chose four well-known and relatively basic networks, i.e. a FCNN, a simple CNN, and a RNN for benchmarking. Performance of these DL models in providing artifact-corrected EEG data has been measured using several standard metrics, such as RRMSE, PSD and CC. Furthermore, we define the network convergence, expressed by loss as a function of epoch number, as a qualitative part of the benchmarks. We expect our work to contribute to the DL-based EEG denoising field, in particular for we standardizes evaluation metrics of performance.

Our benchmarks of four deep learning networks and two traditional methods have demonstrated the feasibility of using DL-based methods to attenuate artifacts from EEG signals. Our comparisons of the four networks (*i.e.*, FCNN, simple CNN, complex CNN, RNN) with two traditional methods (*i.e.*, EMD and filter) suggest that DL-based methods outperform the traditional method, and the supervised end-to-end deep learning has great potential to remove artifacts in EEG signals. Specifically, for the ocular artifacts, the range of *CC* values in our four networks are at equivalent level of the *CC* values reported in a previous study, which used a regression-based method and an offline ICA-based method [72]. A consistent result has been also reported by a DL-based ocular artifact removal study [37]. However, these studies have not offered benchmarks for comparing with other methods [73, 74, 75, 76]. For the myogenic artifacts, comparable *RRMSE_{temporal}* values have been reported in previous literature, such as an ICA-based method [63] and a canonical correlation analysis-based method [77].

The performance of the neural networks depends on the data quality and frequency characteristics of artifacts. The neural networks provide better results for high SNR signals than low SNR signals (Figure 8). Moreover, the high-frequency artifacts, such as EMG artifacts, are more difficult for neural networks to deal with (Figure 8-9). This phenomenon may be explained by the *F-Principle* of neural network [78]. The F-Principle proves that deep learning networks often learn low-frequency information in the early stages of training, and then learn high-frequency information as training iterations increase.

One advantage of deep learning for EEG artifact removal is its flexibility and generalizability. Although the DL-based denoising methods require a large amount of ground-truth EEG data in the training stage, once the model is trained, it can be easily applied to new data, such as multi-channel EEG data or task-related EEG data, regardless of the corresponding reference channels for artifact removal. Another advantage lies in the handling of complex (*e.g.*, nonlinear and non-stationary) artifact mixtures. Due to the hierarchical structure of deep neural networks, DL models can directly learn the true nature of neural activities from training data in the hidden space, and then generate the cleaned EEG data according to the new contaminated EEG input, whereas traditional methods usually linearly attenuate artifacts. Therefore, methods based on deep learning are expected to provide better performance than traditional methods in noise removal.

Several limitations should also be mentioned. First of all, an important potential problem is the size of the dataset and the type of data. Although we provided thousands of segments of EEG, ocular and myogenic artifacts in EEGdenoiseNet, it is possible that more complex neural networks might require larger amounts of data for training and testing. Another drawback is the diversity of the EEG type and artifact type. EEG data may be collected in resting

state or in different task conditions; furthermore, artifacts in EEG recordings are not only limited to ocular and myogenic. For example, removing motion artifacts is important for EEG mobile applications. Criteria for reviewing and approving additional EEG data submissions to EEGdenoiseNet would be helpful. Such an evolving dataset will contribute to improve the generalability of the DL-based EEG denoising networks to diverse brain states. Third, we only focused on the denosing of 2s-long EEG segments in this study. It is worth noting that some EEG tasks might be longer than 2 seconds, not to mention the case of resting EEG. In the future, EEGdenoiseNet needs to be extended to adapt to the denoising of continuous EEG. The continuous artifact removal problem can be solved by defining pseudo-segments in continuous EEG recordings, and extracting the hidden relationships between consecutive segments, such that the previous segment can be used in the training stage as input to constrain the denoising process of the current segment. Forth, here we only focus on single-channel EEG denoising, and the deep learning model learns the temporal information of EEG signals and EOG/EMG artifacts. To use supervised models to learn spatial features, a benchmark data set with multi-channel EEG data must be provided in future. Finally, we did not explore unsupervised deep learning models in this study. When there is no gold standard for clean EEG signals and artifacts, unsupervised deep learning may be of great importance.

6 Conclusion

In this study, we provided a dataset containing thousands of clean EEG, ocular artifacts and muscular artifact segments, which is suited for benchmarking DL-based EEG denoising methods. The dataset is well-structured and publicly available online in different formats. In addition, we included a set of benchmark tools to facilitate the evaluation of newly developed DL-based EEG denoising models. Our benchmarking results suggested that DL methods have great potential to remove both ocular and myogenic artifacts from EEG data, even at high noise levels. Our study may accelerate the development of DL-based EEG denoising field.

Acknowledgement

This work was funded in part by the National Natural Science Foundation of China (62001205), Guangdong Natural Science Foundation Joint Fund (2019A1515111038), Shenzhen Science and Technology Innovation Committee (20200925155957004, KCXFZ2020122117340001, SGDX2020110309280100), Shenzhen Key Laboratory of Smart Healthcare Engineering (ZDSYS20200811144003009).

Competing interests

The authors declare no competing interests.

References

- [1] Hanshu Cai, Zhidiao Qu, Zhe Li, Yi Zhang, Xiping Hu, and Bin Hu. Feature-level fusion approaches based on multimodal eeg data for depression recognition. *Information Fusion*,

59:127–138, 2020.

- [2] Lindsay M Oberman, Edward M Hubbard, Joseph P McCleery, Eric Lewin Altschuler, Vilayanur S Ramachandran, and Jaime A Pineda. Eeg evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, 24(2):190–198, 2005.
- [3] Jonathan R Wolpaw, Dennis J Mcfarland, Gregory W Neat, and Catherine Forneris. An eeg-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3):252–259, 1991.
- [4] Christoph Herrmann and Tamer Demiralp. Human eeg gamma oscillations in neuropsychiatric disorders. *Clinical Neurophysiology*, 116(12):2719–2733, 2005.
- [5] Quanying Liu, Seyedehrezvan Farahibozorg, Camillo Porcaro, Nicole Wenderoth, and Dante Mantini. Detecting large-scale networks in the human brain using high-density electroencephalography. *Human Brain Mapping*, 38(9):4631–4643, 2017.
- [6] Mingqi Zhao, Marco Marino, Jessica Samogin, Stephan P Swinnen, and Dante Mantini. Hand, foot and lip representations in primary sensorimotor cortex: a high-density electroencephalography study. *Scientific Reports*, 9(1):1–12, 2019.
- [7] Rodney J Croft and Robert J Barry. Removal of ocular artifact from the eeg : a review. *Neurophysiologie Clinique-clinical Neurophysiology*, 30(1):5–19, 2000.
- [8] Suresh D Muthukumaraswamy. High-frequency brain activity and muscle artifacts in meg/eeg: a review and recommendations. *Frontiers in Human Neuroscience*, 7:138–138, 2013.
- [9] Luca Piontonachini, Kenneth Kreutzdelgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 2019.
- [10] Mike X Cohen. *Analyzing neural time series data: theory and practice*. MIT press, 2014.
- [11] Maximilien Chaumon, Dorothy V M Bishop, and Niko A Busch. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, 250:47–63, 2015.
- [12] Brenton W Mcmenamin, Alexander J Shackman, Jeffrey S Maxwell, David R W Bachhuber, Adam M Koppenhaver, Lawrence L Greischar, and Richard J Davidson. Validation of ica-based myogenic artifact correction for scalp and source-localized eeg. *NeuroImage*, 49(3):2416–2432, 2010.
- [13] Gabriele Gratton, Michael GH Coles, and Emanuel Donchin. A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, 55(4):468–484, 1983.
- [14] Rodney J Croft and Robert J Barry. Eog correction : Which regression should we use? *Psychophysiology*, 37(1):123–125, 2000.

- [15] Breton W McMenamin, Alexander J Shackman, Jeffrey S Maxwell, Lawrence L Greischar, and Richard J Davidson. Validation of regression-based myogenic correction techniques for scalp and source-localized eeg. *Psychophysiology*, 46(3):578–592, 2009.
- [16] Ping He, G Wilson, and C Russell. Removal of ocular artifacts from electro-encephalogram by adaptive filtering. *Medical and biological engineering and computing*, 42(3):407–412, 2004.
- [17] Ping He, Glenn Wilson, Christopher Russell, and Maria Gerschutz. Removal of ocular artifacts from the eeg: a comparison between time-domain regression method and adaptive filtering method using simulated data. *Medical and biological engineering and computing*, 45(5):495–503, 2007.
- [18] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
- [19] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.
- [20] Xun Chen, Xueyuan Xu, Aiping Liu, Soojin Lee, Xiang Chen, Xu Zhang, Martin J McKeown, and Z Jane Wang. Removal of muscle artifacts from the eeg: a review and recommendations. *IEEE Sensors Journal*, 19(14):5353–5368, 2019.
- [21] Tzyy-Ping Jung, Scott Makeig, Marissa Westerfield, Jeanne Townsend, Eric Courchesne, and Terrence J Sejnowski. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111(10):1745–1758, 2000.
- [22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [24] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
- [26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [27] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

- [28] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.
- [29] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5):051001, aug 2019.
- [30] Siavash Sakhavi, Cuntai Guan, and Shuicheng Yan. Parallel convolutional-linear neural network for motor imagery classification. In *Signal Processing Conference*, 2015.
- [31] Yousef, Rezaei, Tabar, Ugur, and Halici. A novel deep learning approach for classification of eeg motor imagery signals. *Journal of Neural Engineering*, 2017.
- [32] Zhichuan Tang, Chao Li, and Shouqian Sun. Single-trial eeg classification of motor imagery using deep convolutional neural networks. *Optik - International Journal for Light and Electron Optics*, 2017.
- [33] Tian-jian Luo, Yachao Fan, Lifei Chen, Gongde Guo, and Changle Zhou. Eeg signal reconstruction using a generative adversarial network with wasserstein distance and temporal-spatial-frequency loss. *Frontiers in Neuroinformatics*, 14:15, 2020.
- [34] Isaac A. Corley and Yufei Huang. Deep eeg super-resolution: Upsampling eeg spatial resolution with generative adversarial networks. In *2018 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2018.
- [35] Kay Hartmann, Robin Schirrmeyer, and Tonio Ball. Eeg-gan: Generative adversarial networks for electroencephalographic (eeg) brain signals, 06 2018.
- [36] Fatemeh Fahimi, Strahinja Dosen, Kai Keng Ang, Natalie Mrachacz-Kersting, and Cuntai Guan. Generative adversarial networks-based data augmentation for brain-computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
- [37] Banghua Yang, Kaiwen Duan, Chengcheng Fan, Chenxiao Hu, and Jinlong Wang. Automatic ocular artifacts removal in eeg using deep learning. *Biomedical Signal Processing and Control*, 43:148–158, 2018.
- [38] Weitong Sun, Yuping Su, Xia Wu, and Xiaojun Wu. A novel end-to-end 1d-rescnn model to remove artifact from eeg signals. *Neurocomputing*, 2020.
- [39] Conor Hanrahan. Noise reduction in eeg signals using convolutional autoencoding techniques. 2019.
- [40] Haoming Zhang, Chen Wei, Mingqi Zhao, Haiyan Wu, and Quanying Liu. A novel convolutional neural network model to remove muscle artifacts from eeg. *arXiv preprint arXiv:2010.11709*, 2020.
- [41] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

- [42] Michaël Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. Fma: A dataset for music analysis. *arXiv preprint arXiv:1612.01840*, 2016.
- [43] Alexander Panchenko, Eugen Ruppert, Stefano Faralli, Simone Paolo Ponzetto, and Chris Biemann. Building a web-scale dependency-parsed corpus from commoncrawl. *arXiv preprint arXiv:1710.01779*, 2017.
- [44] Logan T Trujillo, Candice T Stanfield, and Ruben D Vela. The effect of electroencephalogram (eeg) reference choice on information-theoretic measures of the complexity and integration of eeg signals. *Frontiers in Neuroscience*, 11:425, 2017.
- [45] Mastaneh Torkamani-Azar, Sumeyra Demir Kanik, Serap Aydin, and Mujdat Cetin. Prediction of reaction time and vigilance variability from spatio-spectral features of resting-state eeg in a long sustained attention task. *IEEE journal of biomedical and health informatics*, 2020.
- [46] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.
- [47] Gijsbrecht Van Veen, Alexandre Barachant, Anton Andreev, Grégoire Cattan, Pedro Coelho Rodrigues, and Marco Congedo. Building brain invaders: Eeg data of an experimental validation. *arXiv preprint arXiv:1905.05182*, 2019.
- [48] Louis Korczowski, Martine Cederhout, Anton Andreev, Grégoire Cattan, Pedro C Rodrigues, Violette Gautheret, and Marco Congedo. Brain invaders calibration-less p300-based bci with modulation of flash duration dataset (bi2015a). *Archives Ouvertes*, 2019.
- [49] Matthew D Luciw, Ewa Jarocka, and Benoni B Edin. Multi-channel eeg recordings during 3,936 grasp and lift trials with varying weight and friction. *Scientific data*, 1(1):1–11, 2014.
- [50] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, aug 2017.
- [51] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.
- [52] Murat Kaya, Mustafa Kemal Binli, Erkan Ozbay, Hilmi Yanar, and Yuriy Mishchenko. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Scientific data*, 5:180211, 2018.
- [53] Cho Hohyun, Ahn Minkyu, Ahn Sangtae, Kwon Moonyoung, and Jun Sung Chan. Eeg datasets for motor imagery brain computer interface. *Gigaence*, 6(7), 2017.
- [54] Suguru Kanoga, Masaki Nakanishi, and Yasue Mitsukura. Assessing the effects of voluntary and involuntary eyeblinks in independent components of electroencephalogram. *Neurocomputing*, 193(Jun.12):20–32, 2016.

- [55] Mehrdad Fatourechi, Ali Bashashati, Rabab K. Ward, and Gary E. Birch. Emg and eog artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 118(3):480–494, 2007.
- [56] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller. Separability of four-class motor imagery data using independent components analysis. *Journal of Neural Engineering*, 3(3):208–216, 2006.
- [57] Schlogl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, and Pfurtscheller and. A fully automated correction method of eog artifacts in eeg recordings. *Clinical Neurophysiology Shannon*, 2007.
- [58] Clemens Brunner, Muhammad Naeem, Robert Leeb, Bernhard Graimann, and Gert Pfurtscheller. Spatial filtering and selection of optimized components in four class motor imagery eeg data using independent components analysis. *Pattern Recognition Letters*, 28(8):957–964, 2007.
- [59] Alois Schlögl, Julien Kronegg, Jane Huggins, and Steve Mason. Evaluation criteria for bci research. *Toward Brain-Computer Interfacing*, 01 2007.
- [60] Ville Rantanen, Mirja Ilves, Antti Vehkaoja, Anton Kontunen, Jani Llylykangas, Eeva Mäkelä, Markus Rautiainen, Veikko Surakka, and Jukka Lekkala. A survey on the feasibility of surface emg in facial pacing. volume 2016, pages 1688–1691, 08 2016.
- [61] Gang Wang, Chaolin Teng, Kuo Li, Zhonglin Zhang, and Xiangguo Yan. The removal of eog artifacts from eeg signals using independent component analysis and multivariate empirical mode decomposition. *IEEE Journal of Biomedical and Health Informatics*, 20(5):1301–1308, 2016.
- [62] Xun Chen, Hu Peng, Fengqiong Yu, and Kai Wang. Independent vector analysis applied to remove muscle artifacts in eeg data. *IEEE Transactions on Instrumentation and Measurement*, pages 1–10, 2017.
- [63] Wim De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Biomedical Engineering*, 53(12):2583–2587, 2006.
- [64] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International joint conference on Artificial intelligence*, 1995.
- [65] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- [66] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [67] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

- [68] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [69] Haoming Zhang, Mingqi Zhao, Chen Wei, Dante Mantini, Zherui Li, and Quanying Liu. Eegdenoisenet. <https://github.com/ncclabsustech/EEGdenoiseNet>.
- [70] Ahmet Mert and Aydin Akan. Hilbert-huang transform based hierarchical clustering for eeg denoising. 2014.
- [71] Manousos A Klados and Panagiotis D Bamidis. A semi-simulated eeg/eog dataset for the comparison of eog artifact rejection techniques. *Data in brief*, 8:1004–1006, 2016.
- [72] Roberto Guarneri, Marco Marino, Federico Barban, Marco Ganzetti, and Dante Mantini. Online eeg artifact removal for bci applications by adaptive spatial filtering. *Journal of neural engineering*, 15(5):056009, 2018.
- [73] Hoang-Anh T Nguyen, John Musson, Feng Li, Wei Wang, Guangfan Zhang, Roger Xu, Carl Richey, Tom Schnell, Frederic D McKenzie, and Jiang Li. Eog artifact removal using a wavelet neural network. *Neurocomputing*, 97:374–389, 2012.
- [74] JC Woestenburg, MN Verbaten, and JL Slangen. The removal of the eye-movement artifact from the eeg by regression analysis in the frequency domain. *Biological psychology*, 16(1-2):127–147, 1983.
- [75] Thomas Elbert, Werner Lutzenberger, Brigitte Rockstroh, and Niels Birbaumer. Removal of ocular artifacts from the eeg—a biophysical approach to the eog. *Electroencephalography and clinical neurophysiology*, 60(5):455–463, 1985.
- [76] Germán Gómez-Herrero, Wim De Clercq, Haroon Anwar, Olga Kara, Karen Egiazarian, Sabine Van Huffel, and Wim Van Paesschen. Automatic removal of ocular artifacts in the eeg without an eog reference channel. In *Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG 2006*, pages 130–133. IEEE, 2006.
- [77] Xun Chen, Hu Peng, Fengqiong Yu, and Kai Wang. Independent vector analysis applied to remove muscle artifacts in eeg data. *IEEE Transactions on Instrumentation and Measurement*, 66(7):1770–1779, 2017.
- [78] Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*, 2019.

Figures and Tables

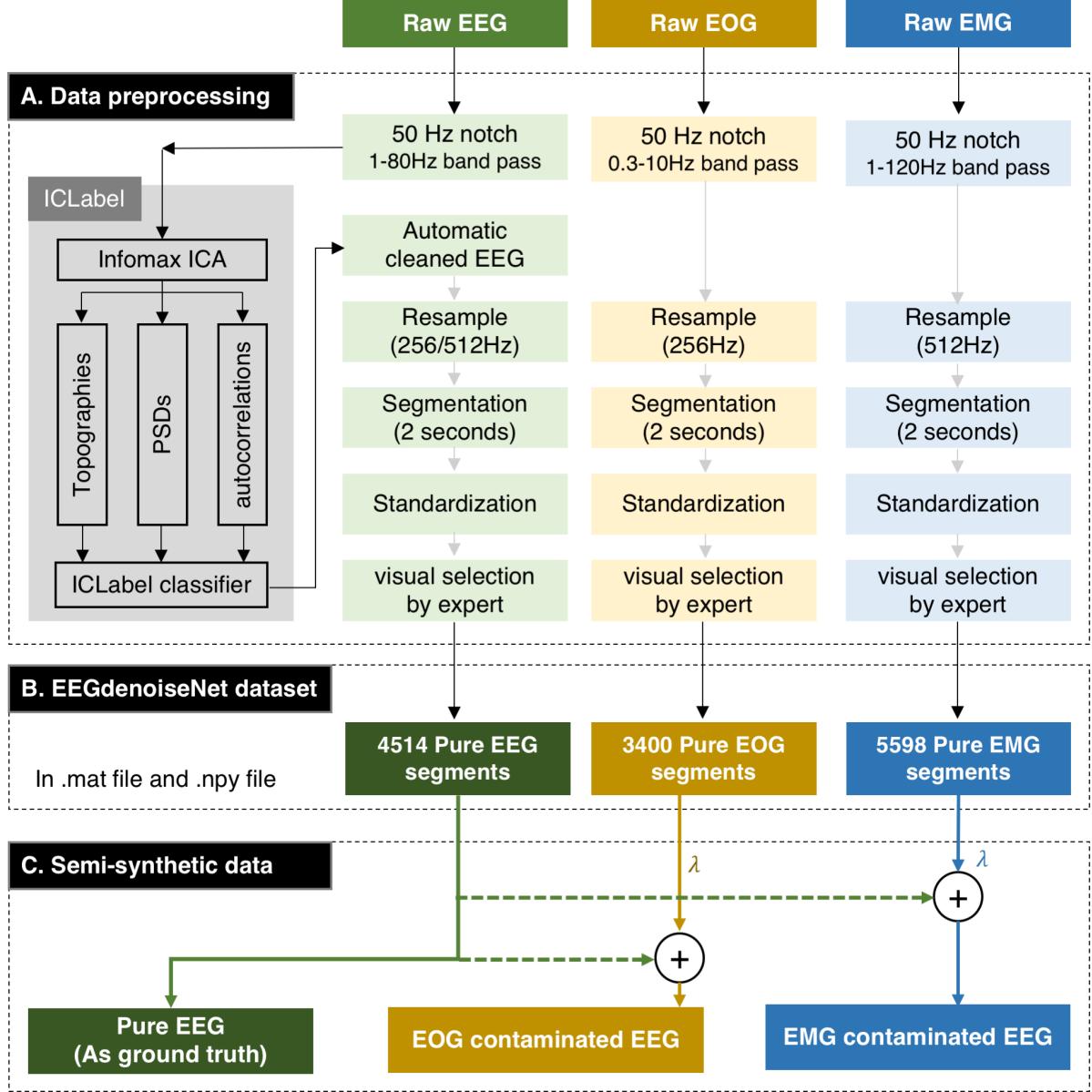


Figure 1: The pipeline for obtaining clean EEG, EOG and EMG. (A) To obtain the clean EEG, pure EOG and pure EMG segments, we firstly preprocess the raw data. The data preprocessing include filtering, ICA-based artifacts removal, resampling, standarization, and visual checked by an expert. (B) 4514 pure EEG segments, 3400 pure EOG segments and 5598 pure EMG segments are obtained. EEGdenoiseNet dataset include two data formats: .mat files and .npy files. (C) The semi-synthetic data is generated by mixing a pure EEG segment and an EOG/EMG segment.

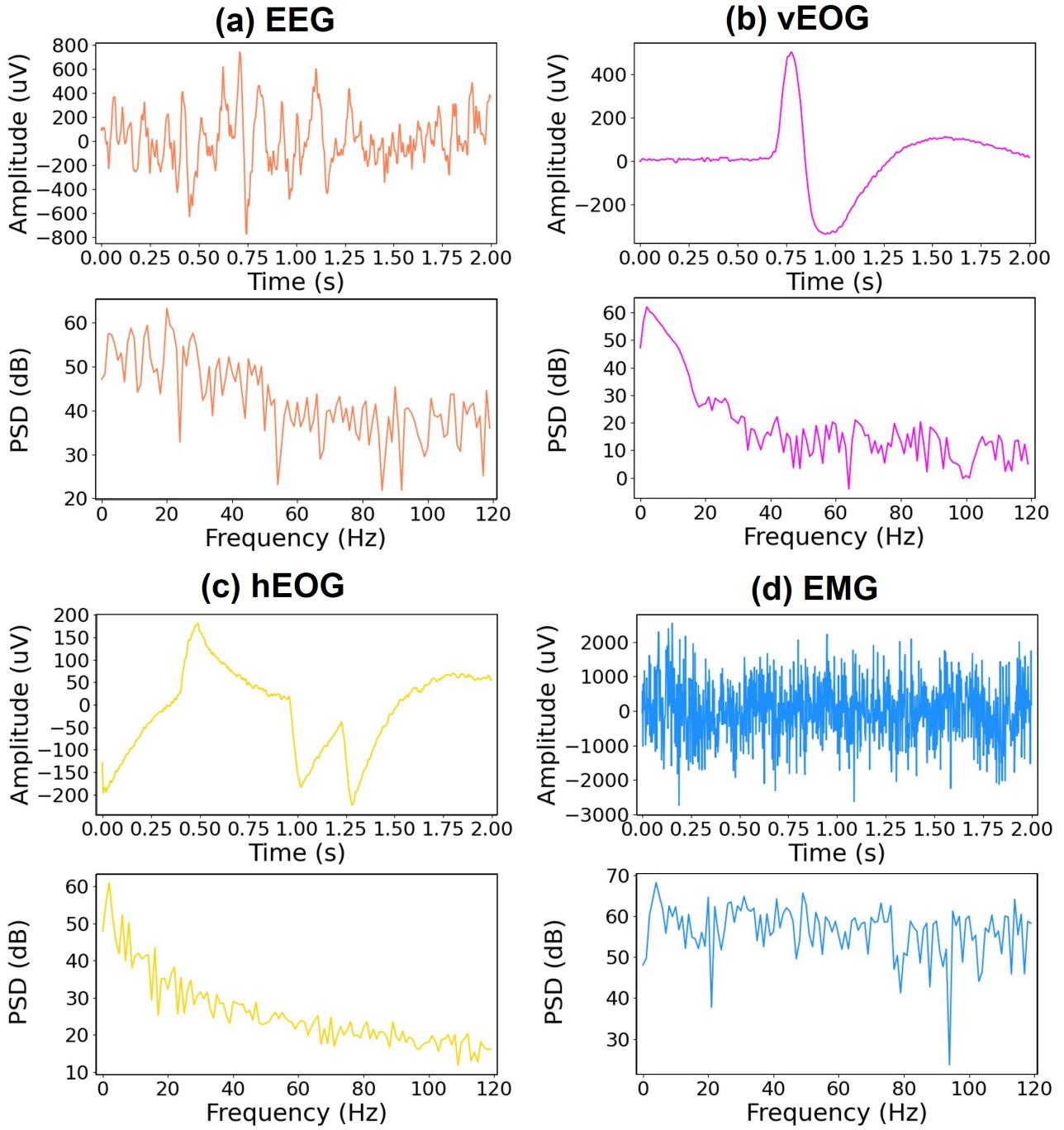


Figure 2: Examples of segments in EEGdenoiseNet dataset. (a) An EEG segment. (b) A vertical EOG (vEOG) segment. (c) A horizontal EOG (hEOG) segment. (d) An EMG segment. (upper) The time course. (bottom) The PSD.

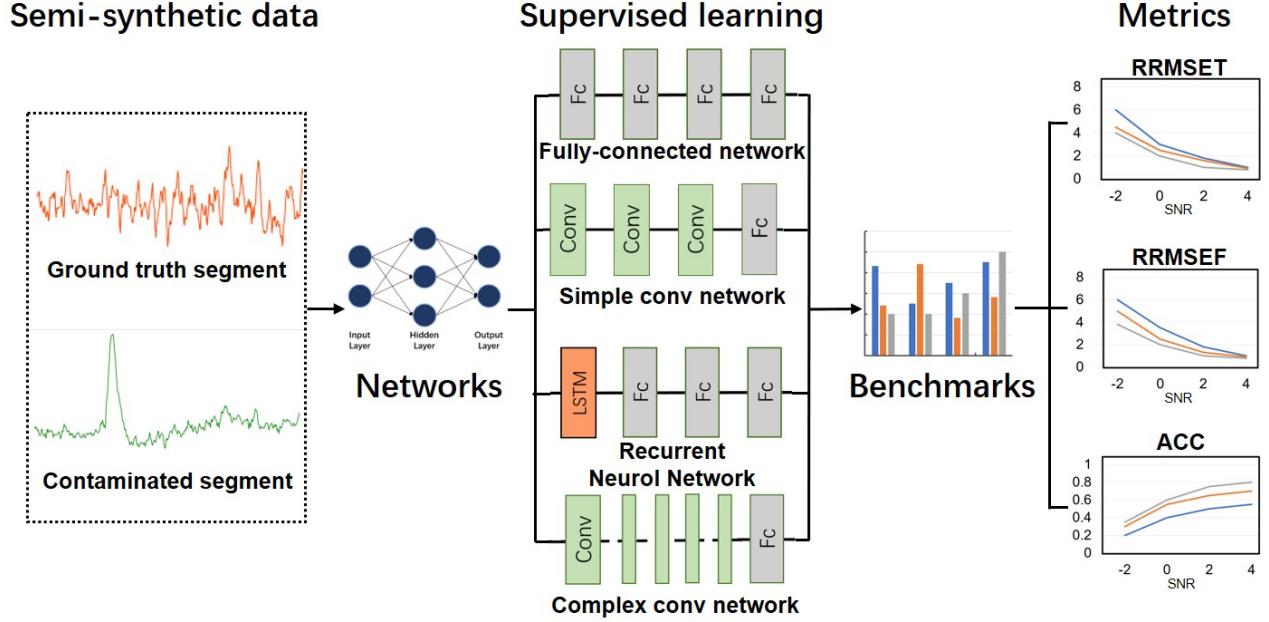


Figure 3: The framework of EEGdenoiseNet. The semi-synthetic data includes ground truth EEG segment and contaminated segment. The one-dimensional contaminated segment is fed into neural networks. The networks are trained in a supervised, end-to-end manner. The output of neural networks are the cleaned EEG signal. The performance of networks are quantified with multiple metrics as benchmarks. This process was performed separately for ocular artifact removal and myogenic artifact removal.

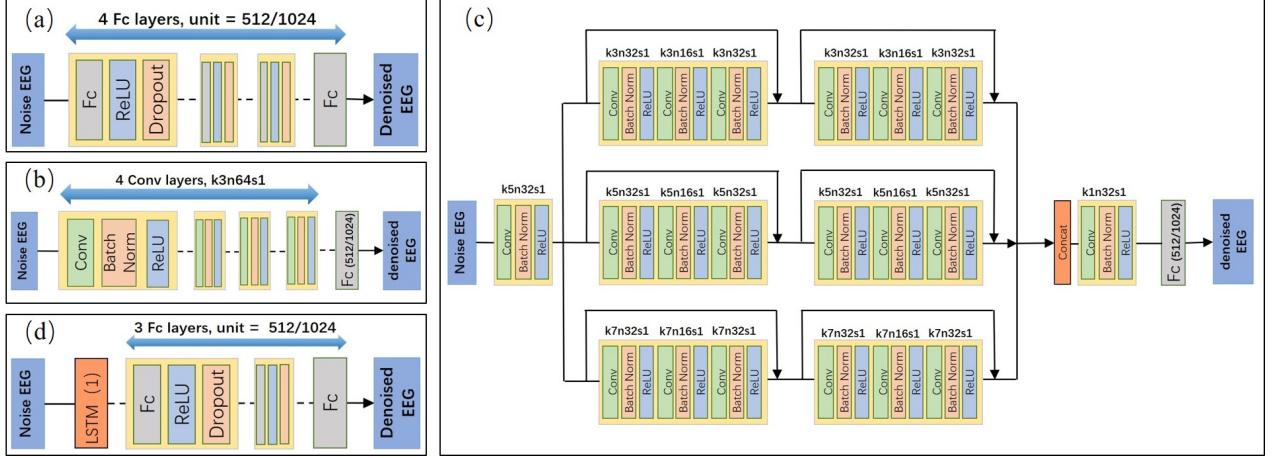


Figure 4: The structures of the four DL-based methods for benchmarking. (a)FCNN ; (b) Simple CNN; (c) Complex CNN; (d) RNN. For each of the networks, the input is the contaminated segments (1×512 for ocular artifact removal, and 1×1024 for myogenic artifact removal), and the output is the cleaned segments (1×512 for ocular artifact removal, and 1×1024 for myogenic artifact removal). The networks are trained to learn the genuine neural activities from contaminated signal segments according to contaminated segments and ground truth segments.

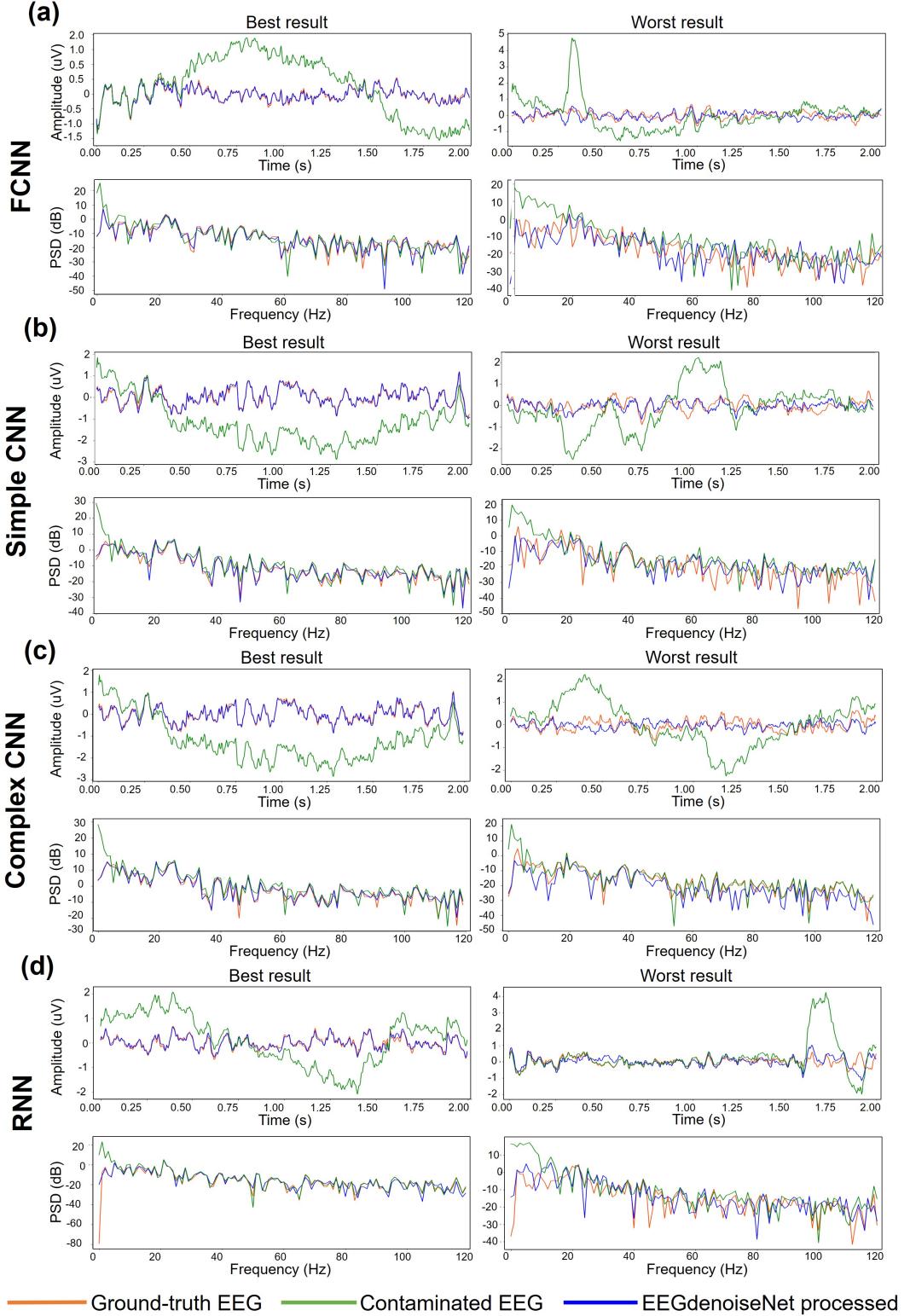


Figure 5: Some exemplary segments of the performance in temporal domain (upper) and spectral domain (bottom) for ocular artifact removal. (a) FCNN. (b) Simple CNN. (c) Complex CNN. (d) RNN. (left) The examples with the best denoising performance; (right) the examples with the worst denoising performance. The orange, green and blue line are the ground-truth EEG, the noisy EEG and the cleaned EEG by EEGdenoiseNet, respectively.

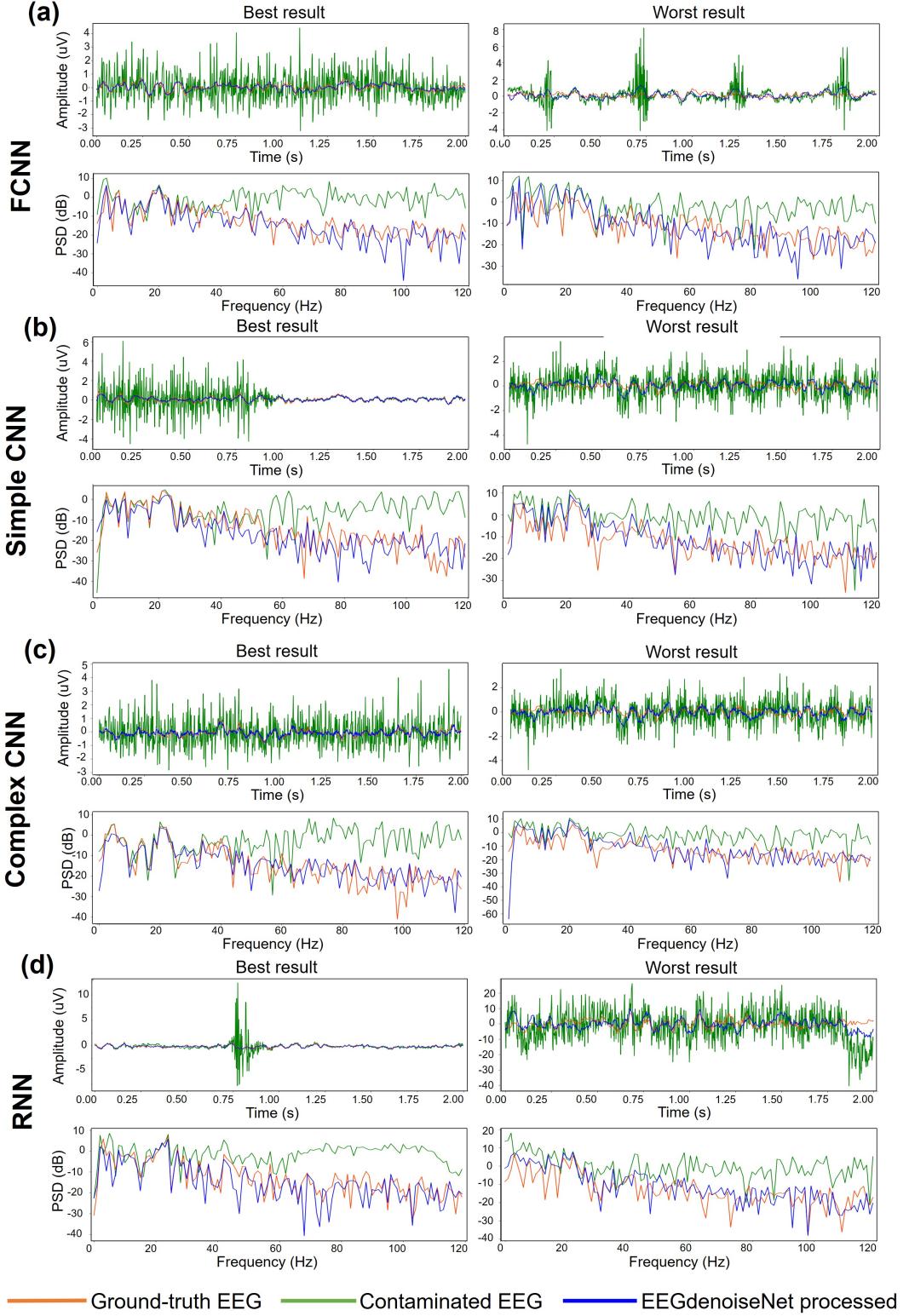
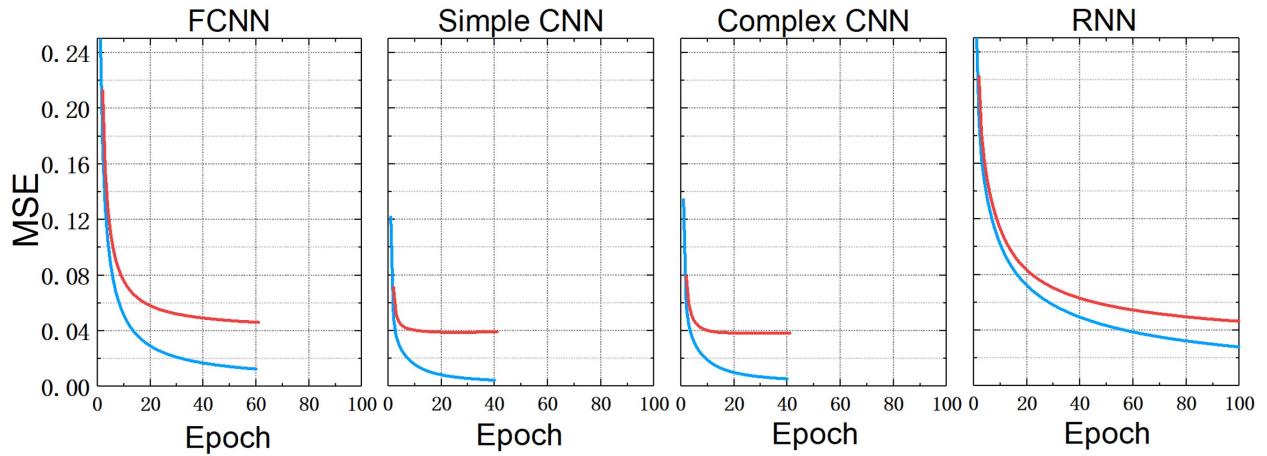


Figure 6: Some exemplary segments showing the performance in temporal (*upper*) and spectral (*bottom*) domains for myogenic artifact removal. (a) FCNN. (b) Simple CNN. (c) Complex CNN. (d) RNN. (*left*) The examples with the best denoising performance; (*right*) the examples with the worst denoising performance. The orange, green and blue lines are the ground-true EEG, the noisy EEG and the cleaned EEG by EEGdenoiseNet, respectively.

(a) Ocular artifact removal



(b) Myogenic artifact removal

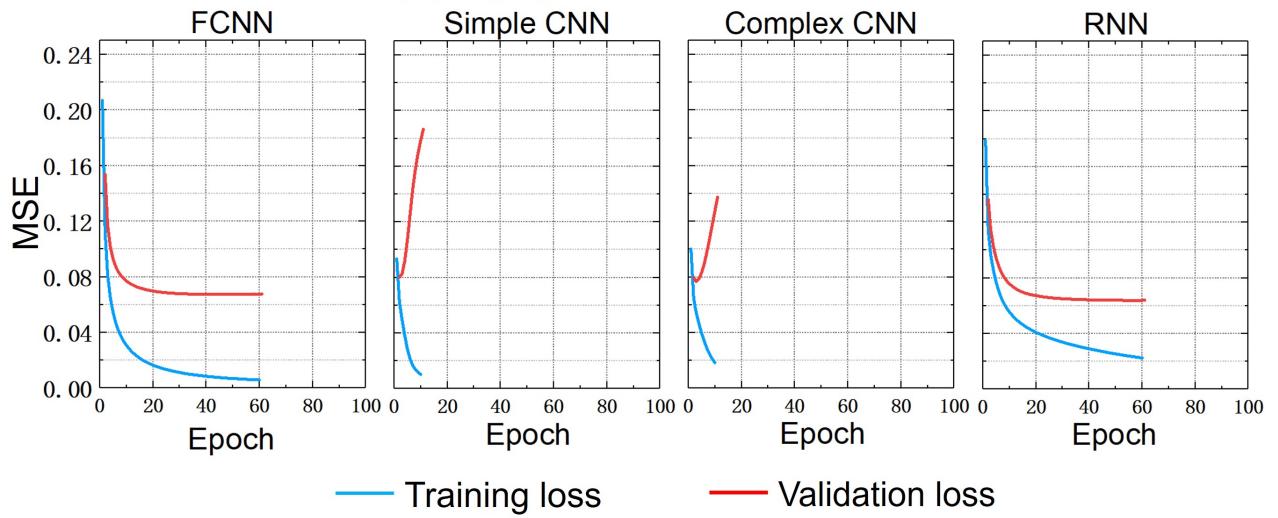
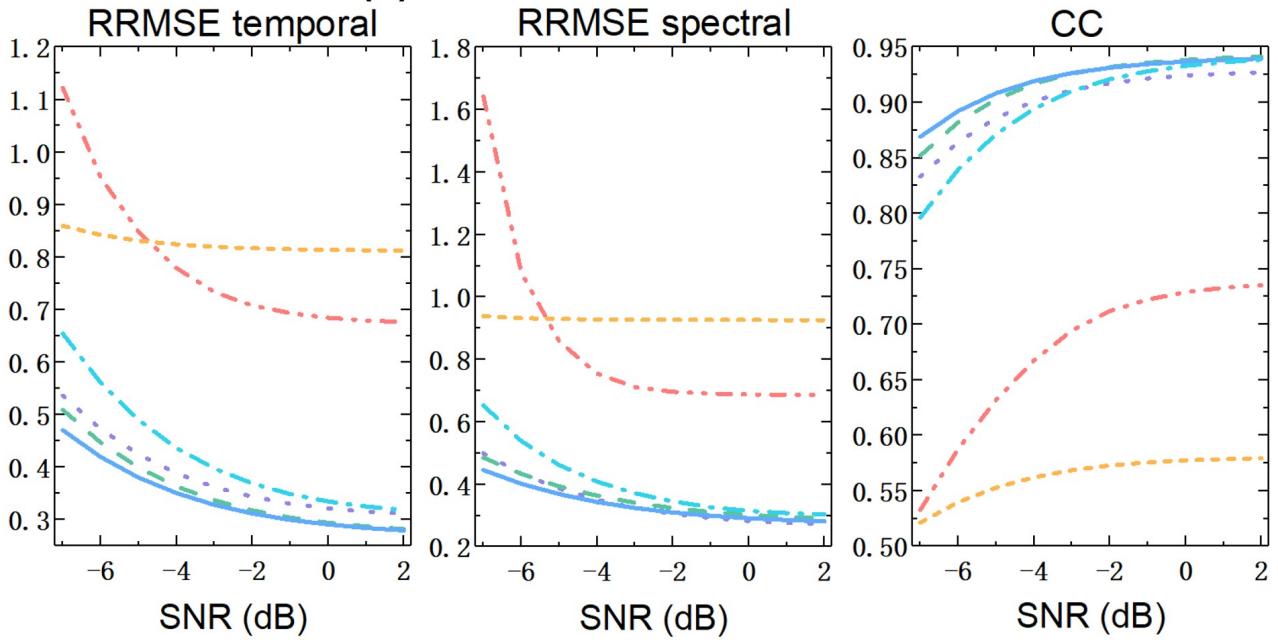


Figure 7: The MSE loss as a function of the number of epochs: (a) ocular artifact removal; (b) myogenic artifact removal. The red line is the learning curve for the training set and the blue line for the validation set.

(a) Ocular artifact removal



(b) Myogenic artifact removal

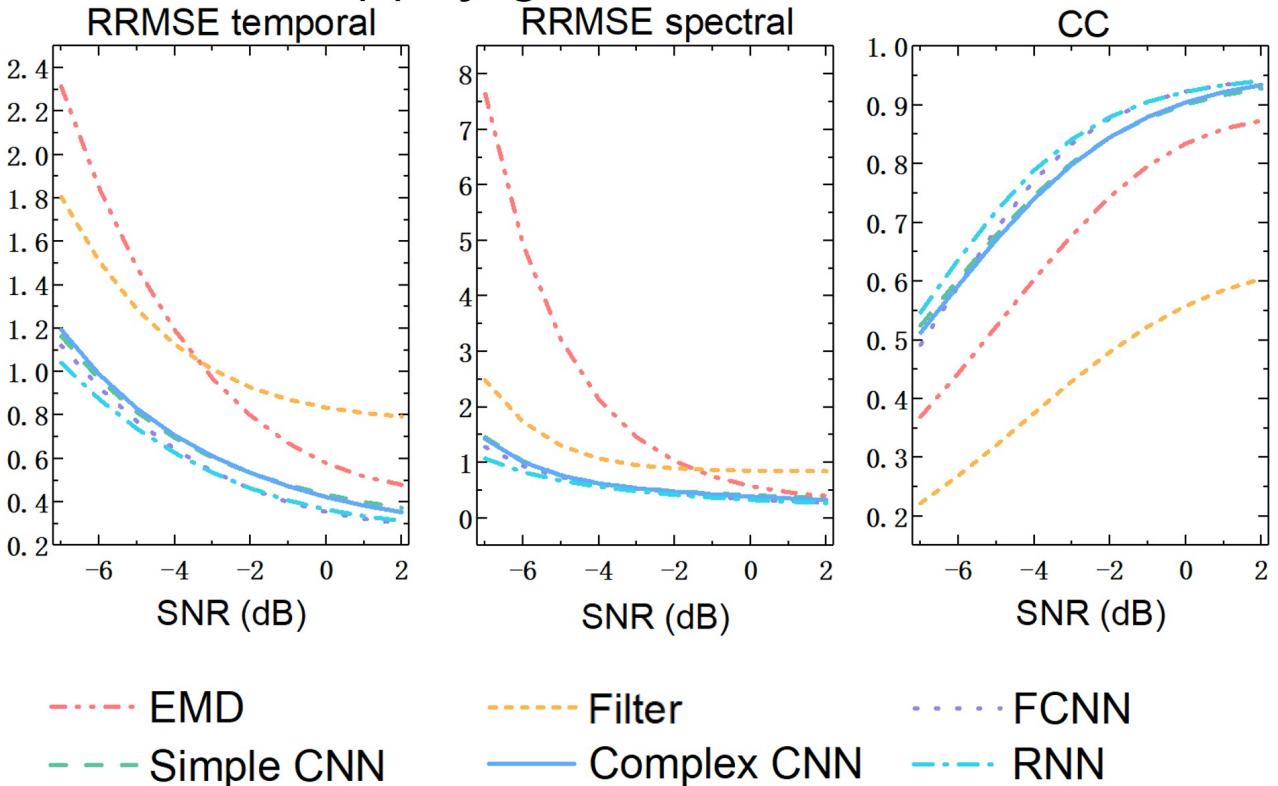
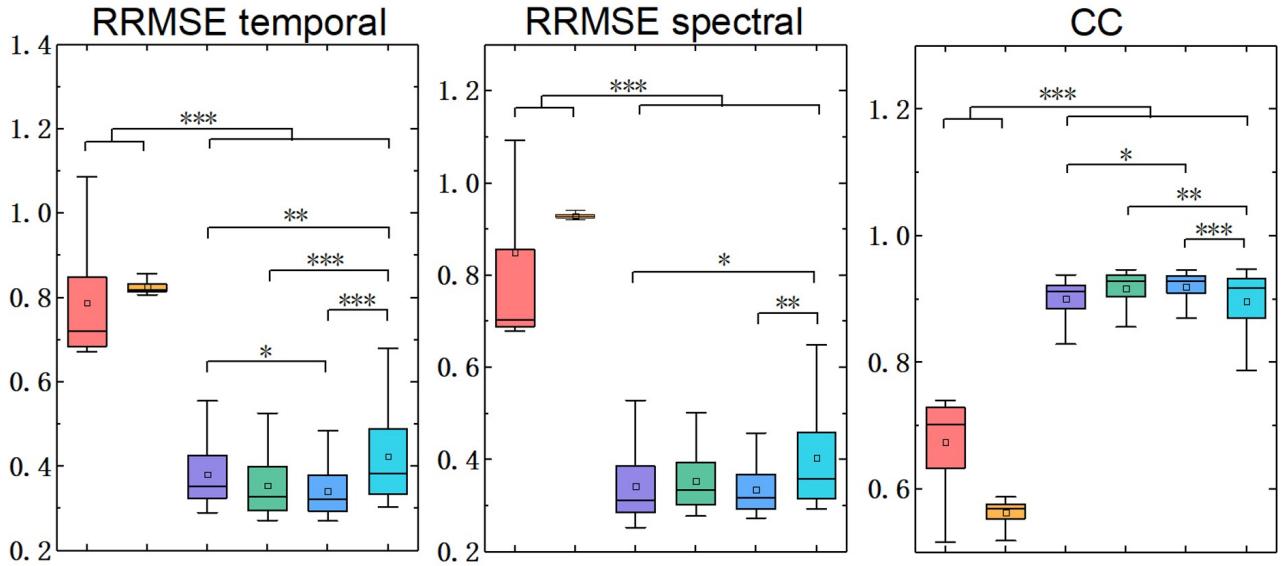
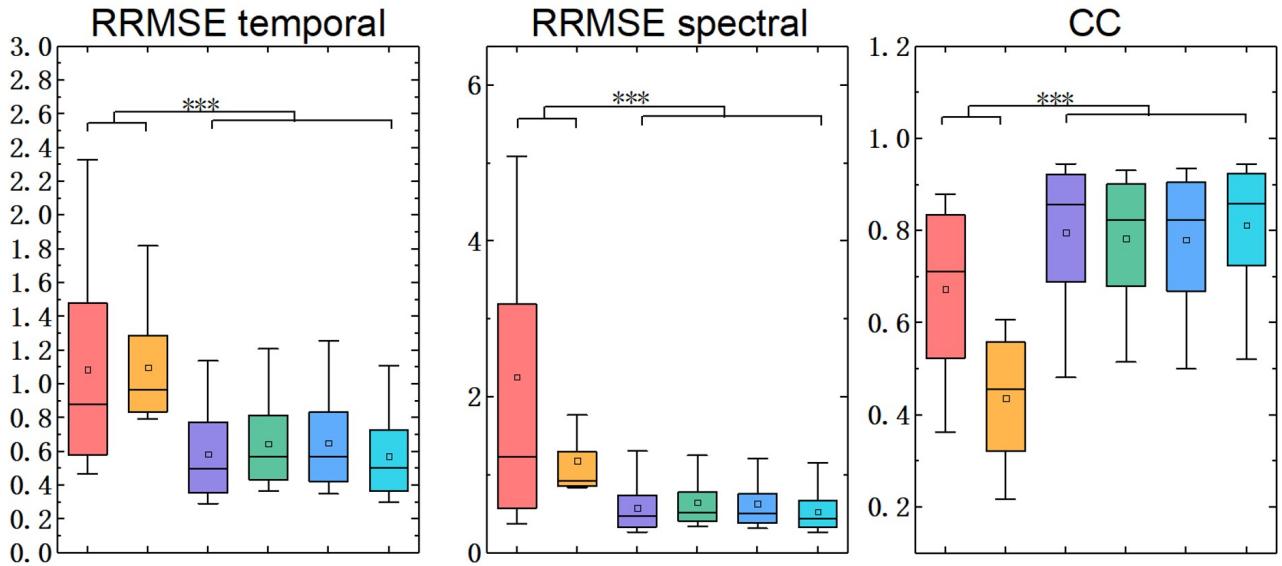


Figure 8: Performance of four deep-learning networks at different SNR levels: (a) ocular artifact removal; (b) myogenic artifact removal. The cold colored lines represent deep learning methods, while the warm colored lines represent traditional methods. The denoising performance increases as the SNR increases.

(a) Ocular artifact removal



(b) Myogenic artifact removal



EMD

Filter

FCNN

Simple CNN

Complex CNN

RNN

Figure 9: Performance of four DL networks (FCNN, simple CNN, complex CNN, RNN) and two traditional methods (EMD and filter): (a) Ocular artifact removal, (b) Myogenic artifact removal. Deep learning models robustly outperform EMD and filtering for EEG denoising.

Table 1: Summary of the data collections used in our dataset

Dataset	signal type	# of Subjects	mean age \pm SD	dataset website
Cho et al. (2017)	EEG	52	26 ± 3.86	http://gigadb.org/dataset/100295
Suguru et al. (2016)	EOG	20	22.75 ± 1.45	http://u4ag2kanosr1.blogspot.jp/
Naeem et al. (2006)	EOG	8	23.8 ± 2.5	http://www.bbci.de/competition/iv/
Schlögl et al. (2007)	EOG	10	age between 17 and 31	http://www.bbci.de/competition/iv/
Ville et al. (2015)	EMG	15	40.7 ± 9.6	http://urn.fi/URN:NBN:fi:tty-201611044685

Table 2: Power ratios of different frequency bands before and after ocular artifact removal

Denoising method	delta	theta	alpha	beta	gamma
EMD	0.025	0.042	0.096	0.585	0.252
Filter	0.000	0.000	0.000	0.405	0.595
FCNN	0.129	0.127	0.085	0.500	0.159
Simple CNN	0.131	0.127	0.085	0.492	0.165
Complex CNN	0.128	0.127	0.085	0.493	0.166
RNN	0.124	0.122	0.088	0.506	0.159
ground truth	0.143	0.141	0.093	0.467	0.157
contaminated signal	0.514	0.216	0.070	0.151	0.049

‘Ground truth’ refers to the clean EEG segments.

‘Contaminated signal’ refers to the mixed EEG signal generated by Eq. (1).

Abbreviation: EMD, empirical mode decomposition; FCNN, fully-connected neural network; CNN, covolutional neural network; RNN, recurrent neural network.

Table 3: Power ratios of different frequency bands before and after myogenic artifact removal

Denoising method	delta	theta	alpha	beta	gamma
EMD	0.227	0.162	0.093	0.330	0.188
Filter	0.000	0.000	0.000	0.312	0.687
FCNN	0.147	0.144	0.092	0.481	0.135
Simple CNN	0.119	0.138	0.096	0.506	0.142
Complex CNN	0.123	0.139	0.097	0.492	0.149
RNN	0.139	0.138	0.093	0.482	0.147
ground truth	0.142	0.140	0.093	0.464	0.160
contaminated signal	0.200	0.141	0.077	0.300	0.281

Open and free EEG datasets for epilepsy diagnosis

Palak Handa · Monika Mathur · Nidhi Goel

Received: date / Accepted: date

Abstract The Epilepsies are a common, chronic neurological disorder affecting more than 50 million individuals across the globe. It is characterized by unprovoked, recurring (similar or different type) seizures which are commonly diagnosed through clinical EEGs. Good-quality, open-access and free EEG data can act as a catalyst for on-going state-of-the-art (SOTA) research works for detection, prediction and management of epilepsy and seizures. They can also aid in improving the quality of life (QOL) of these diseased individuals and contribute research in healthcare multimedia, data analytics and Artificial Intelligence (AI) in personalized medicine. This paper presents widely used, available, open and free EEG datasets available for epilepsy and seizure diagnosis. A brief comparison and discussion of open and private datasets has also been done. Such datasets will help in development and evaluation of automatic computer-aided system in healthcare.

Keywords Open datasets · Biomedical multimedia · EEG signals · Epilepsy diagnosis · Seizure · Performance

1 Introduction

The Epilepsies are a chronic neurological disorder characterized by unprovoked, recurring (similar or differ-

P. Handa
Dept. of ECE, DTU, Delhi
E-mail: palakhanda97@gmail.com

M. Mathur
Dept. of ECE, IGDTUW, Delhi
E-mail: mathur.monika2007@gmail.com

N. Goel
Dept. of ECE, IGDTUW, Delhi
E-mail: nidhi.iitr1@gmail.com

ent type) seizures. A clinical EEG setting is used by doctors to observe different types of epileptic activity as it leaves distinct impressions in the form of interictal epileptiform discharges, peri-ictal activities and high frequency oscillations *etc* [1].

The annual economic burden of epilepsy is enormous in developing countries like India where it is estimated to be 88.2% of gross national product (GNP) per capita and 0.5% of the overall GNP [1]. Hence, early diagnosis through recent technologies like Artificial Intelligence (AI), feature engineering, data analytics and multimedia is vital and can aid in Quality Of Life (QOF) of patients and their associated caretakers.

Secured, reproducible AI algorithm, good quality data and efficient computing horse power are the major elements for development of early detection and prediction of epileptic wave forms through EEG signals. There are several types of EEGs such as intracranial, scalp, ambulatory, *etc*. They are recorded in video, image and signal format depending on their use and application in hospitals.

There is a huge demand of real-time biomedical multimedia tools for data analysis, and pattern recognition of such formats. Bonn EEG time series database [2] was the first EEG dataset to be publically available for research applications in this field. It remains as the benchmark dataset for most research works due to its availability. Several datasets discussed in this paper encourage scientific advances in this field.

The data quality of biomedical datasets is measured through various factors such as presence of artefacts and noise, missing values, descriptive information, annotations by health experts, pre-fined data structure, processing and robustness to outliers *etc*.

The datasets mentioned in the section 2 are freely available (except temple university which requires a lo-

gin), re-distributable for research purposes, previously freely available and or became private or removed from the portal based on adult and paediatric population. Table 1 shows a comparison of publically available, open-access (except temple university which requires a login) human EEG datasets for epilepsy diagnosis.

The source and availability of these were verified on 26-07-2021, which may change in the future. They were found using different keywords like ‘EEG datasets for epilepsy’, ‘datasets for epilepsy detection’, ‘EEG based epilepsy diagnosis’, and ‘open EEG datasets’ on Pub med and google scholar search engine.

2 EEG datasets for epilepsy diagnosis

There are several EEG datasets for epilepsy diagnosis which are freely available and private due to various reasons such as lack of ethical clearance. This section lists all the existing EEG datasets with their URLs for adult and paediatric population.

2.1 Adult datasets

The adult population consists of affected individuals above 20 years of age. There are several databases like American Epilepsy Society Seizure Prediction Challenge database [3], dataset of EEG recordings of pediatric patients with epilepsy based on the 10-20 system [4] and Karunya University [5] which contain both adult and paediatric EEG data. They are mentioned in section 2.2.

2.1.1 Bonn EEG time series database[2]

This database comprises of 100 single channels EEG of 23.6 seconds with sampling rate of 173.61 Hz. Its spectral bandwidth range is between 0.5 Hz and 85 Hz. It was taken from a 128 channel acquisition system. Five patients EEG sets were cut out from a multi-channel EEG recording and named A, B, C, D and E. Set A and B are the surface EEG recorded during eyes closed and open situation of healthy patients respectively. Set C and D are the intracranial EEG recorded during a seizure free from within seizure generating area and from outside seizure generating area of epileptic patients respectively. Set E is the intracranial EEG of an epileptic patient during epileptic seizures. Each set contains 100 text files wherein each text file has 4097 samples of 1 EEG time series in ASCII code. A band pass filter with cut off frequency as 0.53 Hz and 40 Hz has been applied on the data. It is an artifact free data and hence no prior pre-processing is required for the

classification of healthy (non-epileptic) and un-healthy (epileptic) signals. The strong eye movement’s artefacts were omitted. It was made available in 2001. The extended version of this data is now a part of EPILEPSIA project. Available link: [a](#) and [b](#)

2.1.2 Bern-Barcelona EEG database [6]

This multi channel EEG database was recorded using specialized electrodes and consists of five patients with longstanding pharmacoresistant temporal lobe epilepsy. The patients underwent epilepsy surgery. The sampling rate was either 512 or 1024 Hz based upon whether they were recorded with less or more than 64 channels of EEG system. Three out of five attained complete seizure freedom. Two types of EEG are present in the data *i.e.*, focal and non-focal. Each file has about 10240 samples for a time duration of 20 seconds. Available link: [c](#) and [d](#)

2.1.3 Temple University EEG corpus [7]

Temple University EEG corpus is the largest free EEG data available for epilepsy and seizure types diagnosis till date. It consists of data acquired from 2000 to 2013 using different EEG clinical settings for about 10,874 patients. This community has developed various software products such as annotation tools, toolboxes for seizure detection, and EDF browser for data analysis of EEG, EMG, and ECG *etc* signals. EDF browser helps to view EEG recording in a video form. There are various datasets available such as IBM Features For Seizure Detection (IBMFT), the TUH EEG epilepsy corpus, seizure corpus, slowing corpus, and events corpus *etc*. A user ID and password is required to get access to these datasets. [Available link](#).

2.1.4 Neurology and sleep center, New Delhi EEG dataset [8]

This database comprises of 5.12 seconds EEG data. It was recorded using 57 EEG channel Grass Tele-factor Comet AS40 Amplification System; sampled at 200 Hz. Its spectral bandwidth range is between 0.5 Hz and 70 Hz. Time series EEG datasets are categorized into three major MATLAB file folder namely ictal, pre-ictal and inter-ictal stages. Each MAT file has 1024 samples. A subset of this database is publically available. [Available link](#).

2.1.5 Epileptic Seizure Recognition Data Set [2]

The time series EEG dataset consists of 11500 instances of EEGs of 4 subjects suffering from epilepsy. This data

has been removed from the UCI machine learning repository recently and was released in 2017. It is a simplified version of the original data released by [2]. It consists of 5 subjects (4 unhealthy and 1 healthy) performing different activities and experience epileptic seizures except subject 1. The time duration for each EEG was 23.5 seconds.

2.1.6 Siena Scalp EEG Database [9, 10]

This multi-channel EEG database of 14 epileptic patients (9 males and 5 females) was recorded using specialized amplifiers, and reusable electrodes. The signals were recorded with a sampling rate of 512 Hz and stored in EDF files. The data has been acquired from Unit of Neurology and Neurophysiology of the University of Siena, Italy and focuses on seizure prediction. It is an integral part of national interdisciplinary research project PANACEE. This data contains 47 seizures from 128 hours of video EEG recording. The start and end time of a seizure was also recorded and contains the list of electrodes present on the scalp of a patient during event recording. Three types of seizures namely focal onset with and without impaired awareness, and focal to bilateral tonic-clonic (FBTC) were found and recorded in the diseased patients. [Available link](#).

2.1.7 Single electrode EEG data of healthy and epileptic patients [11, 12]

This dataset was generated with a motive to build predictive epilepsy diagnosis model and publically available since 2020. It was generated on a similar acquisition and settings *i.e.*, sampling frequency, bandpass filtering and number of signals and time duration as of University of Bonn. It has overcome the limitations faced by University of Bonn dataset such as different EEG recording (inter-cranial and scalp) for healthy and epileptic patients [11]. All the data were taken exclusively using surface EEG electrodes for 15 healthy and epileptic patients. [Available link](#).

2.1.8 Epileptic EEG Dataset [13]

This multi-channel, long term EEG database was recorded for 6 patients suffering from focal epilepsy. They were undergoing pre-surgical evaluation for possible epilepsy surgery. Different EEG segments of a seizure like ictal, pre-ictal, inter-ictal and its onsets have been included in the data. The signals were recorded with a sampling rate of 500 Hz and stored in EDF files. Labelled and classified data points (train and test set) have been mentioned for complex partial electrographic, and video-detected seizures. All the EEG signals underwent band

pass filtering of range 1-70 Hz where 50 Hz (utility frequency) was also removed. [Available link](#).

2.2 Paediatric datasets

The paediatric EEG database consists of affected individuals from age 1 month - 20 years.

2.2.1 Children's hospital Boston-MIT database [14]

This database comprises of 844 hour continuous EEG. 23 pediatric patients from age 1.5-19 who underwent scalp multi-channel EEG recording. It is the first paediatric EEG database available for epilepsy and seizure diagnosis. The patients were given anti-seizure medications. About 200 seizures were recorded in a universal bio-polar montage with about 24-27 EEG channels. Sampling frequency was kept to be 256 Hz. Each EEG segment is called as a record which usually is for duration of one hour. There are 9-42 edf files from a single subject. Additional vagal nerve stimulus signals are also present. Separate file name and montages have been mentioned for seizure v/s non seizure episode in EEG segments. [Available link](#).

2.2.2 Karunya University [5]

This database comprises of 18 channel EEG data with segments of normal, focal and generalized epileptic seizure activities from 1–107 years of patients. It was released in 2014 but the website is not available for research use now. Each segment has 2056 sample points. Sampling frequency was kept to be 256 Hz. The EEG recordings vary from 40 minutes to one hour. It was collected from a diagnostic center based in Coimbatore, India. [Available link](#).

2.2.3 A dataset of neonatal EEG recordings with seizures annotations [15]

This database consists of multi-channel, good quality EEG recordings of 79 term neonates where 39 of them suffered from neonatal seizures in the NICU of Helsinki University Hospital, Finland. The recordings were captured with NicOne EEG amplifier, and 19 EEG channel cap. The signals were recorded with a sampling rate of 256 Hz and stored in EDF files. It consists of seizure annotations by healthcare experts for seizure detection purpose. The data was pre-processed using butterworth high-pass filtering. The data also contains natural artefacts. [Available link](#).

Table 1 Comparison of existing EEG datasets for epilepsy diagnosis

Ref.	Availability	Type	Source	Year	Size	No. of channels	No. of patients	Sampling frequency	EEG segments
[2]	Freely available	Adult	e-repositori upf.	2001	3.05 MB	100 single	5	173.61 Hz	seizure states, healthy
[14]	Freely available	Paediatric	PhysioNet repository	2010	40 GB	23-26	22	256 Hz	Intractable seizures
[6]	Freely available	Adult	e-repositori upf.	2012	814 MB	64	5	512 Hz	Focal, Non-focal
[3]	Freely available	Dog and human	Kaggle	2014	105 GB	-	-	-	different types
[5]	Not available	Adult and Paediatric	Website	2014	-	-	-	256 Hz	normal, focal and generalized epileptic seizures
[7]	Free but requires login	Adult	Website	2015	572 GB	20-31	10,874	250, 256, 512 Hz	different types
[8]	Freely available	Adult	Researchgate	2016	604 KB	57	10	200 Hz	Ictal, inter-ictal, pre-ictal EEGs
[2]	Removed	Adult	UCI repository	2017	3 MB	100 single	5	173.61 Hz	seizure related, healthy
[15]	Freely available	Paediatric (neonates)	Zenedo	2018	4.3 GB	19	79	256 Hz	seizure onset
[16]	Private	Adult	-	2019	-	19	115	128 Hz	epileptic and healthy
[11]	Private	Adult	-	2019	-	-	50	250, 256 Hz	generalized and focal epilepsies
[17]	Private	Adult	-	2019	-	21	5	500 Hz	focal and tonic-clonic
[18]	Private	Pediatric	-	2019	-	-	29	200, 500 Hz	typical absence seizures
[19]	Private	Adult	-	2019	-	-	12	256 Hz	seizure events
[20]	Private	-	-	2019	-	21	25	200 Hz	seizure events
[21]	Private	-	-	2019	-	18	10	256 Hz	seizure states
[22]	Private	-	-	2019	-	22	22	250 Hz	ictal, non-ictal
[12]	Freely available	Adult	Zenedo	2020	20 MB	-	15	173.61 Hz	Inter-ictal
[23]	Private	-	-	2020	-	21	-	250 Hz	seizure onsets
[24]	Private	Adult	-	2020	-	21	150	256 Hz	seizure and normal
[10]	Freely available	Adult	PhysioNet repository	2020	20 GB	29	14	512 Hz	Epileptic seizures (focal onset, tonic-clonic)
[13]	Freely available	Adult	Mendeley repository	2021	3133 MB	21	6	500 Hz	Complex partial, electrographic and video-detected seizures
[4]	Freely available	Paediatric and Adult	Open neuro repository	2021	15 GB	52	30	2000 Hz	HFO markings

2.2.4 Dataset of EEG recordings of pediatric patients with epilepsy based on the 10-20 system [4]

This dataset consists of scalp EEG recordings to study the impact age on observed High Frequency Oscillations (HFO) in pediatric epileptic patients. Three hours of pediatric and adult EEG sleep data was recorded for 30 focal or generalised epileptic patients. The signals were recorded with a sampling rate of 2000 Hz and stored in EDF files. Different sleep stage annotations are available in this database. [Available link.](#)

2.3 Others

2.3.1 American Epilepsy Society Seizure Prediction Challenge database [3]

This database consists of intra-cranial EEG segments from dogs and humans with different acquisitions of sampling rate, duration of EEGs, and no. of electrodes etc. It was released as a part of the kaggle challenge hosted by the American Epilepsy Society in 2014 for development of seizure forecasting systems and witnessed about 504 teams. Different seizure segments of ictal, pre-ictal, post-ictal, inter-ictal were provided in MATLAB files. The data storage was about 105 GB. Additional annotated EEG data was also provided by the University of Pennsylvania and the Mayo Clinic. [Available link.](#)

2.3.2 Private databases

Several private databases have also been recorded for epilepsy diagnosis using EEG signals [11, 24, 16, 17, 18, 22, 23, 19, 20, 21]. The European Epilepsy database [25] is a private database which consists of high quality, annotated EEG signals from University of Bonn [2], Freiburg [26], Flint hills, and many multi-modal like MRI imaging data. The website of [5] is not available. EEG data in [27] was freely available till 2015 when its portal crashed.

3 Conclusion

Diagnosis, treatment and management of epilepsy is still a challenging task for the scientific and health-care community. It's detection by visual introspection of long hour EEG is not only time taking but a very tedious and subjective task. Artificial intelligence can help in escalating this process and lead to successful detection of different types of epilepsies through efficient, high quality and annotated EEG data. This paper has

presented all the existing EEG datasets for epilepsy diagnosis with its availability and brief comparisons. Such datasets motivate scientific research in early diagnosis of epilepsy through robust techniques.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. SV Thomas, PS Sarma, M Alexander, L Pandit, L Shekhar, C Trivedi, and B Vengamma. Economic burden of epilepsy in india. *Epilepsia*, 42(8):1052–1060, 2001.
2. Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):061907, 2001.
3. J Jeffry Howbert, Edward E Patterson, S Matt Stead, Ben Brinkmann, Vincent Vasoli, Daniel Crepeau, Charles H Vite, Beverly Sturges, Vanessa Rueybusch, Jaideep Mavoori, et al. Forecasting seizures in dogs with naturally occurring epilepsy. *PloS one*, 9(1):e81920, 2014.
4. Dorottya Cserpan et al. Dataset of eeg recordings of pediatric patients with epilepsy based on the 10-20 system, 2021.
5. Thomas George Selvaraj, Balakrishnan Ramasamy, Stanly Johnson Jeyaraj, and Easter Selvan Suvise-shamuthu. Eeg database of seizure disorders for experts and application developers. *Clinical EEG and neuroscience*, 45(4):304–309, 2014.
6. Ralph G Andrzejak, Kaspar Schindler, and Christian Rummel. Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients. *Physical Review E*, 86(4):046206, 2012.
7. Iyad Obeid and Joseph Picone. The temple university hospital eeg data corpus. *Frontiers in neuroscience*, 10:196, 2016.
8. P Swami, B Panigrahi, S Nara, M Bhatia, and T Gandhi. Eeg epilepsy datasets, 2016.
9. Paolo Detti, Giampaolo Vatti, and Garazi Zabalo Manrique de Lara. Eeg synchronization analysis for seizure prediction: A study on data of noninvasive recordings. *Processes*, 8(7):846, 2020.

10. P. Detti. Siena scalp eeg database (version 1.0.0), 2020. URL <https://doi.org/10.13026/5d4a-j060>.
11. Siddharth Panwar, Shiv Dutt Joshi, Anubha Gupta, and Puneet Agarwal. Automated epilepsy diagnosis using eeg with test set evaluation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(6):1106–1116, 2019.
12. Siddharth Panwar. Single electrode EEG data of healthy and epileptic patients, February 2020. URL <https://doi.org/10.5281/zenodo.3684992>.
13. Wassim Nasreddine. Epileptic eeg dataset, 2021. URL <https://doi.org/10.17632/5pc2j46cbc.1>.
14. Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
15. Nathan Stevenson, Karoliina Tapani, Leena Lauronen, and Sampsa Vanhatalo. A dataset of neonatal EEG recordings with seizures annotations, June 2018. URL <https://doi.org/10.5281/zenodo.2547147>.
16. Shivarudhrappa Raghu, Natarajan Sriram, Yasin Temel, Shyam Vasudeva Rao, Alangar Satyaranjandas Hegde, and Pieter L Kubben. Performance evaluation of dwt based sigmoid entropy in time and frequency domains for automated detection of epileptic seizures using svm classifier. *Computers in biology and medicine*, 110:127–143, 2019.
17. Duanpo Wu, Zimeng Wang, Lurong Jiang, Fang Dong, Xunyi Wu, Shuang Wang, and Yao Ding. Automatic epileptic seizures joint detection algorithm based on improved multi-domain feature of ceeg and spike feature of aeeg. *IEEE Access*, 7:41551–41564, 2019.
18. Mustafa Talha Avcu, Zhuo Zhang, and Derrick Wei Shih Chan. Seizure detection using least eeg channels by deep convolutional neural network. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1120–1124. IEEE, 2019.
19. Peter Z Yan, Fei Wang, Nathaniel Kwok, Baxter B Allen, Sotirios Keros, and Zachary Grinspan. Automated spectrographic seizure detection using convolutional neural networks. *Seizure*, 71:124–131, 2019.
20. Gwangho Choi, Chulkyun Park, Junkyung Kim, Kyungin Cho, Tae-Joon Kim, HwangSik Bae, Kyeongyuk Min, Ki-Young Jung, and Jongwha Chong. A novel multi-scale 3d cnn with deep neural network for epileptic seizure detection. In *2019 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–2. IEEE, 2019.
21. Jiuwen Cao, Jiahua Zhu, Wenbin Hu, and Anton Kummert. Epileptic signal classification with deep eeg features by stacked cnns. *IEEE Transactions on Cognitive and Developmental Systems*, 12(4):709–722, 2019.
22. Muhammad Bilal, Muhammad Rizwan, Sajid Saleem, Muhammad Murtaza Khan, Mohammed Saeed Alkatheir, and Mohammed Alqarni. Automatic seizure detection using multi-resolution dynamic mode decomposition. *IEEE Access*, 7:61180–61194, 2019.
23. Dhanalekshmi P Yedurkar and Shilpa P Metkar. Multiresolution approach for artifacts removal and localization of seizure onset zone in epileptic eeg signal. *Biomedical Signal Processing and Control*, 57:101794, 2020.
24. Khakon Das, Debasish Daschakladar, Partha Pratim Roy, Atri Chatterjee, and Shankar Prasad Saha. Epileptic seizure prediction by the detection of seizure waveform from the pre-ictal phase of eeg signal. *Biomedical Signal Processing and Control*, 57:101720, 2020.
25. Matthias Ihle, Hinnerk Feldwisch-Drentrup, César A Teixeira, Adrien Witon, Björn Schelter, Jens Timmer, and Andreas Schulze-Bonhage. Epilepsiae—a european epilepsy database. *Computer methods and programs in biomedicine*, 106(3):127–138, 2012.
26. Björn Schelter, Matthias Winterhalder, Thomas Maiwald, Armin Brandt, Ariane Schad, Jens Timmer, and Andreas Schulze-Bonhage. Do false predictions of seizures depend on the state of vigilance? a report from two seizure-prediction methods and proposed remedies. *Epilepsia*, 47(12):2058–2070, 2006.
27. Piotr Zwoliński, Marcin Roszkowski, Jarosław Źygierewicz, Stefan Haufe, Guido Nolte, and Piotr J Durka. Open database of epileptic eeg with mri and postoperative assessment of foci—a real world verification for the eeg inverse solutions. *Neuroinformatics*, 8(4):285–299, 2010.

Mobile BCI dataset of scalp- and ear-EEGs with ERP and SSVEP paradigms while standing, walking, and running

Young-Eun Lee¹, Gi-Hwan Shin¹, Minji Lee¹, and Seong-Whan Lee^{1,2,*}

¹Korea University, Department of Brain and Cognitive Engineering, Seoul, 02841, Republic of Korea

²Korea University, Department of Artificial Intelligence, Seoul, 02841, Republic of Korea

*corresponding author: Seong-Whan Lee (sw.lee@korea.ac.kr)

ABSTRACT

We present a mobile dataset obtained from electroencephalography (EEG) of the scalp and around the ear as well as from locomotion sensors by 24 participants moving at four different speeds while performing two brain-computer interface (BCI) tasks. The data were collected from 32-channel scalp-EEG, 14-channel ear-EEG, 4-channel electrooculography, and 9-channel inertial measurement units placed at the forehead, left ankle, and right ankle. The recording conditions were as follows: standing, slow walking, fast walking, and slight running at speeds of 0, 0.8, 1.6, and 2.0 m/s, respectively. For each speed, two different BCI paradigms, event-related potential and steady-state visual evoked potential, were recorded. To evaluate the signal quality, scalp- and ear-EEG data were qualitatively and quantitatively validated during each speed. We believe that the dataset will facilitate BCIs in diverse mobile environments to analyze brain activities and evaluate the performance quantitatively for expanding the use of practical BCIs.

figshare DOI: <https://doi.org/10.6084/m9.figshare.16669072>

Background & Summary

Human movement is a complex process that requires the integration of the central and peripheral nervous systems, and therefore, researchers have analyzed human locomotion using brain activity^{1,2}. Brain-computer interface (BCI) has been studied based on the communication between human thoughts and external devices to recover the motor sensory function of disabled patients and support daily life of healthy people.^{3–5}. In particular, electroencephalography (EEG) has been used as the most common method for measuring brain activity with high time resolution, portability, and ease of use⁶; in addition, several attempts have been made to increase its practicality^{7–10}. However, EEG recording in a mobile environment can cause artifacts and signal distortion, further resulting in loss of accuracy and signal quality^{11,12}. Therefore, research in the mobile environment is necessary to study the brain activity during movements to mitigate limitations such as loss of accuracy and signal quality and to improve the practical BCI technology^{9,12}. Moreover, several studies have developed software techniques to realize the practical BCI, such as preprocessing algorithm removing artifacts^{13–16} or novel classification algorithm improving user intention performance^{12,17}.

To recognize the human intention, two representative exogenous BCI paradigms, event-related potential (ERP)¹⁸ and steady-state visual evoked potential (SSVEP)¹⁹, are commonly used in the mobile environment owing to their strong responses to brain activity. ERP is a time-locked brain response to stimuli (i.e., visual, auditory, etc.), including a positive peak response (P300) that occurs 300 ms after the stimulus appears. The ERP has a relatively high performance in both scalp-EEG and ear-EEG, with accuracies of 85–95% for scalp-EEG^{20,21} and approximately 70% for ear-EEG⁹ in a static state. SSVEP is a period brain response in the occipital area to stimuli flickering at a particular frequency. The performance of SSVEP is reliable in terms of the accuracy and signal-to-noise ratio (SNR), with 80–95% accuracy for scalp-EEG^{17,20}, but 40–70% accuracy for ear-EEG as it is located far from the occipital cortex^{22,23}. Brain signal data obtained by performing BCI paradigms can be used to quantitatively evaluate signal quality in a mobile environment²⁴.

Portable and non-hair-bearing EEG have been frequently investigated to enhance the applicability of practical BCI in the real world^{25–29}. In particular, ear-EEG, which comprises electrodes placed inside or around the ear, has several advantages over conventional scalp-EEG in terms of stability, portability, and unobtrusiveness^{9,30}. Moreover, the signal quality of ear-EEG has been validated for recognizing human intention using several BCI paradigms, including ERP^{9,29,31}, SSVEP^{22,23}, and others³².

Recently, EEG datasets for mobile environments have been published, including motion information and different mobile environments. He et al.³³ recorded signals from 60-channel scalp-EEG, 4-channel electrooculography (EOG), and 6 goniometers from 8 participants while walking slowly at a constant speed of 0.45 m/s. They used a BCI paradigm, avatar control, by

predicting the joint angle of goniometers from the EEG while walking. Brantley et al.³⁴ collected a dataset consisting of full-body locomotion from 10 participants on stairs, ramps, and level grounds without BCI paradigms, by recording from 60-channel scalp-EEG, 4-channel EOG, 12-channel electromyogram (EMG), and 17 inertial measurement units (IMUs). Wagner et al.³⁵ recorded signals from 108-channel scalp-EEG, 2-channel EMG, 2 pressure sensors, and 3 goniometers from 20 participants without BCI paradigms while walking at a constant speed. Consequently, although these datasets were collected from the mobile environment, the movement condition was kept constant, and two of the datasets were performed without BCI paradigms. In addition, because only scalp-EEG signals were measured, the application of practical BCI was restricted.

In this study, we present a mobile BCI dataset with scalp- and ear-EEGs collected from 24 participants with BCI paradigms at different speeds. Data from 32-channel scalp-EEG, 14-channel ear-EEG, 4-channel EOG, and 27-channel IMUs were recorded simultaneously. The experimental environment involved movements of participants at different speeds of 0, 0.8, 1.6, and 2.0 m/s on a treadmill. For each speed, two BCI paradigms were used to evaluate signal quality, which facilitated diverse analysis, including time-domain analysis using ERP data and frequency-domain analysis using SSVEP data. Therefore, we believe that the dataset facilitates addressing the issues of brain dynamics in diverse mobile environments in terms of the cognitive level of multitasking, locomotion complexity, and quantitative evaluation of artifact removal methods or classifiers for BCI tasks in the mobile environment.

Methods

Participants

Twenty-four healthy individuals (14 men and 10 women, 24.5 ± 2.9 years of age) without any history of neurological or lower limb pathology participated in this experiment. In the ERP tasks, all of them participated, and 17 participants performed a slight running session at a speed of 2.0 m/s. In the SSVEP tasks, 23 of them participated, excluding one because of a personal problem unrelated to the experimental procedure; 16 participants performed a slight running session at a speed of 2.0 m/s. The participants were provided the option to perform a slight running session. This study was approved by the Institutional Review Board of Korea University (KUIRB-2019-0194-01), and all participants provided written informed consent before the experiments. All experiments were conducted in accordance with the Declaration of Helsinki.

Data Acquisition

For conducting the experiment, we simultaneously collected data from three different modalities: scalp-EEG, ear-EEG, and IMU (Figure 1a, b, and c). All data can be accessed from here³⁶. To synchronize the three devices, triggers were sent to the recording system of each device simultaneously while presenting a paradigm in MATLAB.

The head circumference of each participant was measured to select an appropriately sized cap of scalp-EEG. We obtained signals from the scalp with 32 EEG Ag/AgCl electrodes according to the 10/20 international system using BrainAmp (Brain Product GmbH). The ground and reference electrodes were placed on Fpz and FCz. In addition, we used four EOG channels to capture dynamically changing eye movements such as blinking. The EOG channels were placed above and below the left eye to measure the vertical eye artifacts ($VEOG_U$ and $VEOG_L$) as well as at the left and right temples to measure the horizontal eye artifacts ($HEOG_L$ and $HEOG_R$) (Figure 1b). The sampling rate of the scalp-EEG and EOG was 500 Hz with a resolution of 32 bits. All electrode impedances were maintained below $50\text{ }\Omega$, and most channels were reduced to below $20\text{ }\Omega$ ^{33,34}.

The ear-EEG consists of cEEGrids electrodes located around each ear of the participant, with eight channels on the left, six channels on the right, and ground and reference channels on the right at the center (Figure 1c)⁹. Two cEEGrids were connected to a wireless mobile DC EEG amplifier (SMARTING, mBrainTrain, Belgrade, Serbia). Data were recorded at a sampling rate of 500 Hz and a resolution of 24 bits. All impedances of ear-EEG were maintained below $50\text{ }\Omega$, and most channels were reduced to below $20\text{ }\Omega$ ^{33,34}.

To measure the locomotion of the participants, three wearable IMU sensors (APDM wearable technologies) were placed at the head, left ankle, and right ankle. An IMU consisted of 9-channel sensors, including a 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer. Therefore, 27-channel IMU signals were collected and recorded at a sampling rate of 128 Hz with a resolution of 32 bits.

Experimental Paradigm

They performed tasks under the ERP and SSVEP paradigms, in which stimuli were displayed on the monitor during each session at different speeds. Two BCI paradigms were developed based on the OpenBMI (<http://openbmi.org>)²⁰ and Psychtoolbox (<http://psychtoolbox.org>)³⁷ in MATLAB (The Mathworks, Natick, MA).

During the ERP task, target ('OOO') and non-target ('XXX') characters were presented on the monitor as visual stimuli. All characters were displayed with a black background at the center of the monitor screen. The proportion of the target was 0.2, and the total number of trials was 300. In a trial, one of the stimuli was presented for 0.5 s, and a fixation cross ('+') was presented to take a break for randomly 0.5–1.5 s (Figure 1d).

During the SSVEP task, three target SSVEP visual stimuli were displayed at three positions (left, center, and right) on an LCD monitor^{19,38}. The frequency range of stimuli containing 5–30 Hz is known to be appropriate for obtaining SSVEP responses³⁹. It is also known that movement artifacts have a significant impact on frequency spectrum below 12 Hz⁴⁰. Based on these studies, the stimuli were designed to flicker at 5.45, 8.57, and 12 Hz, which were calculated by dividing the monitor frame rate of 60 Hz by an integer (i.e., 60/11, 60/7, and 60/5)²⁰. The participants were asked to gaze in the direction of the target stimulus highlighted in yellow. In each trial, the target of a random sequence was noticed for 2 s, after which all stimuli blinked for 5 s, with a break time for 2 s. The SSVEP experiment consisted of 20 trials for each frequency, a total of 60 trials in a session (Figure 1e).

Experimental Protocol and Procedure

Figure 1a depicts the experimental setup of this study. The participants stood on the treadmill in a lab and were instructed to look at a 24-inch monitor (refresh rate: 60 Hz, resolution: 1920 × 1080 pixels) placed 80 (\pm 10) cm in front of the participants. The participants were monitored and instructed to minimize other movements, such as that of neck or arm, to avoid any artifacts that might occur by movements other than walking. The mobile environment included standing, slow walking, fast walking, and slight running at speeds of 0, 0.8, 1.6, and 2.0 m/s, respectively on the treadmill (0° inclination)^{41,42}.

To proceed with the experiment under the same conditions for each participant, the experimental procedures were sequentially performed (Figure 1f). They conducted two BCI tasks while standing, slow walking, fast walking, and slight running on the treadmill. Training sessions for the ERP task were conducted at a speed of 0 m/s to train the ERP classifier prior to all ERP tasks. Duration of a session of ERP and SSVEP tasks consisted of 7–8 min, with a total of 40 min and 32 min for all sessions. Each session involved the same procedure with a random sequence of targets. All sessions for one participant were performed on a single day. Moreover, due to the possibility of fatigue and habituation, which could be induced in the sequence of the experiment⁴³, the following actions were considered. At first, participants were allowed sufficient breaks between and within sessions when needed. Furthermore, the participants became familiar with the paradigm stimuli by being fully exposed to them before starting the experiment.

Preprocessing

We preprocessed the data using an open-source toolbox for EEG data, such as BBCI (https://github.com/bbci/bbci_public)⁴⁴, BCILAB (<https://github.com/sccn/BCILAB>)⁴⁵, and EEGLAB (<https://sccn.ucsd.edu/eeglab>)⁴⁶ in MATLAB. At first, the data were preprocessed using a high-pass filter that was set above 0.5 Hz using a fifth-order Butterworth filter. Thereafter, three procedures: EOG removal, line noise removal, and interpolation were performed. Vertical EOG components were removed from the scalp-EEG using the *flt_eog* function in BCILAB⁴⁷. The line noise removal method automatically removed artifacts that contained noise for extended periods of time with several parameters. This method removed bad channels that carried abnormal signals with standard deviations above the threshold of z-score using the function *flt_clean_channels* in BCILAB with threshold of 4 and window length of 5 s. The removed bad channels were interpolated using the super-fast spherical method to avoid losing any channel information. On average, 2.38 ± 1.94 channels in the scalp-EEG and 1.35 ± 1.18 channels in the ear-EEG were removed and interpolated for all participants in all sessions. All channels in scalp-EEG and ear-EEG were each re-referenced to a common average reference. We down-sampled the data from scalp-EEG, ear-EEG, and IMU sensors to 100 Hz. The continuous signal was segmented into epoched signals according to the time of each paradigm. For the ERP, each trial was segmented from -200–800 ms based on the stimulus presentation time. For the SSVEP, each trial was segmented from 0–5 s based on the starting time of the stimulus flickering.

Data Records

All data files are available in Open Science Framework repository³⁶ and are available under the terms of Attribution 4.0 International Creative Commons License (<http://creativecommons.org/licenses/by/4.0/>).

Data Format

All data are provided according to the standardized Brain Imaging Data Structure format for EEG data⁴⁸ as shown in Figure 2. The data format followed BrainVision Core Data Format, developed by Brain Products GmbH. The data file is organized with the following naming convention:

sub-XX_ses-YY_task-ZZ_eeg

where the session number includes 1–5, which session 1 indicates training session for ERP and session 2–5 indicates each speed of 0, 0.8, 1.6, and 2.0 m/s, respectively, and the task includes ERP and SSVEP. In ‘sourcedata’ folder, the data are separated into EEG and IMU because their sampling frequencies are different. In each subject folder, the sampling frequencies of data are downed to 100 Hz and data from all modalities are in a file. The number of channels for each modality is listed in Table 1.

Missing Data

Data Missing

The IMU data of participant 21 for ERP at 0 m/s and that of participant 12 for SSVEP at 0.8 m/s were missing because of a malfunction in the communication of the IMU during data collection.

Trials Missing

The number of trials for ERP data of participant 11 at every speed and participant 13 and 15 at a speed of 2.0 m/s, and SSVEP data of participant 14 at a speed of 2.0 m/s were approximately two-thirds of the normal number of trials because of the malfunction of device communication.

Excluded Data

The data of participant 17 at 2.0 m/s for SSVEP, participant 19 at 2.0 m/s for ERP and SSVEP, and participant 20 at 2.0 m/s for ERP were excluded since the electrodes did not adhere well during data recording, resulting in loss of more than 50% in a session.

Technical Validation

We evaluated the preprocessed data using the ERP and SSVEP paradigms in terms of accuracy and SNR. Statistical analysis was conducted using a one-tailed paired t-test to compare the performance at each moving speed to the performance at standing, as indicated by the asterisk at a confidence level of 95%. Moreover, the ERP waves and power spectral density (PSD) for SSVEP were used to evaluate the signal quality. Figure 3a depicts an example of the scalp-EEG, ear-EEG, and IMU signals for 5 s at speeds of 0, 0.8, 1.6, and 2.0 m/s. The amplitudes of the scalp-EEG, ear-EEG, and IMU increased as the speed increased^{41,49,50}. Figure 3b depicts an example of the topography for the scalp-EEG and ear-EEG at different speeds of 0, 0.8, 1.6, and 2.0 m/s. The powers of the scalp-EEG and ear-EEG increased as the speed increased.

Statistical Verification

To verify the dataset, we performed statistical verification to demonstrate significant differences between the speeds across every channel of scalp-EEG and ear-EEG. Figure 4a for the ERP and Figure 4b for the SSVEP depicts the topological map of t-values in particular frequency bands, including delta waves (0.5–3.5 Hz), theta waves (3.5–7.5 Hz), alpha waves (7.5–12.5 Hz), and beta waves (12.5–30 Hz)⁵¹. In particular, PSDs in each frequency band were analyzed using cluster-based correction with non-parametric permutation testing for multiple comparisons to verify the difference between the data at four speeds, including 0, 0.8, 1.6, and 2.0 m/s. The significance probabilities and critical values of permutation distribution are estimated using Monte-Carlo method with iterations of 10,000.

Significant channels could indicate that noise signals are included in corresponding frequency bands and speeds. The topography of the delta band depicts that step frequencies, which was mostly in the range of 0.5–3.5 Hz, affect most channels at all speeds. During slight running session, all channels, including scalp-EEG and ear-EEG, were significantly different in entire frequency band. In addition, paradigm-related areas such as the occipital area during SSVEP tasks and the central area during ERP tasks showed the large t-values in delta band, resulting in low concentration on the tasks due to the workload of multi-tasking.

Evaluation of ERP

The ERP dataset was evaluated by demonstrating ERP waves and metrics using the area under the receiver operating characteristic curve (AUC) and approximate SNR at each speed. AUC indicates the true positive rate over the false positive rate of the results. To acquire the AUC, the features of the ERP were extracted by the power over time intervals of every 50 ms from 200 ms to 450 ms. For the classification, we used a conventional classifier, regularized linear discriminant analysis, to evaluate the ERP performance. The data from the training session at a speed of 0 m/s were used for the training set, and the other dataset containing different speeds was used for the testing set. The SNR can indicate the quality of signals, and approximate SNR of ERP was calculated by the root mean square (RMS) of the amplitude of the peaks at P300 divided by the RMS of the average amplitude of the pre-stimulus baseline (-200–0 ms) at channel Pz^{52,53}.

Figure 5a depicts the ERP waves of the target and non-target stimuli in the scalp- and ear-EEGs at channels Pz and L10 at each speed. The higher the speed, the lower the amplitude of the P300 components of the target in both the scalp- and ear-EEGs. Tables 2 and 3 list the performance of ERP in the scalp-EEG and ear-EEG, respectively. The grand average AUCs of ERP for all participants were 0.90 ± 0.07 , and 0.67 ± 0.07 ($p < 0.05$) in the scalp-EEG at speeds of 0 and 1.6 m/s, respectively, and 0.72 ± 0.14 and 0.58 ± 0.06 ($p < 0.05$) in the ear-EEG at speeds of 0 and 1.6 m/s, respectively. The grand average SNRs of ERP for all participants were 0.95 ± 0.09 and 1.06 ± 0.14 ($p < 0.05$) for the scalp-EEG at speeds of 0 and 1.6 m/s, respectively, and 1.06 ± 0.27 and 0.98 ± 0.05 for the ear-EEG at speeds of 0 and 1.6 m/s, respectively.

Evaluation of SSVEP

The SSVEP dataset was evaluated by implementing statistical analysis to measure the signal properties using PSD, and the metrics using accuracy and approximate SNR at each speed. Accuracy was measured as the percentage of correct predictions in the total number of cases. A canonical correlation analysis was used for the classification that does not require the training datasets. The SNR of SSVEP was calculated using the ratio of the power of the target frequencies to the power of the neighboring frequencies (resolution: 0.25 Hz, number of neighbors: 12)⁵⁴.

Figure 5b depicts the PSD of the SSVEP for the scalp-EEG and ear-EEG at channels Oz and L10 at each speed. The higher the speed, the greater the power in all frequency spectra for both the scalp- and ear-EEGs. Tables 4 and 5 list the performance of the SSVEP for scalp-EEG and ear-EEG, respectively. The grand average accuracies of SSVEP for all participants were $88.70 \pm 19.52\%$ and $80.65 \pm 20.38\%$ ($p < 0.05$) for scalp-EEG at speeds of 0 and 1.6 m/s, respectively, and 53.19 ± 13.93 and 39.57 ± 6.39 ($p < 0.05$) for ear-EEG at speeds of 0 and 1.6 m/s, respectively. The grand average SNRs of SSVEP for all participants were 2.64 ± 0.99 and 1.92 ± 0.68 ($p < 0.05$) for the scalp-EEG at speeds of 0 and 1.6 m/s, respectively, and 1.21 ± 0.23 and 1.03 ± 0.10 ($p < 0.05$) for the ear-EEG at speeds of 0 and 1.6 m/s, respectively.

Usage Notes

This mobile dataset is available in the BrainVision Core Data Format. For analyzing the dataset, we recommend using a common open-source toolbox for EEG data, such as BBCI (https://github.com/bbci/bbci_public)⁴⁴, OpenBMI (<http://openbmi.org>)²⁰, and EEGLAB (<https://sccn.ucsd.edu/eeglab>)⁴⁶ in the MATLAB environment, or MNE (<https://martinos.org/mne>)⁵⁵ in the Python environment. The supporting code is available on GitHub (<https://github.com/DeepBCI/Deep-BCI>). For the preprocessing, we recommend performing down-sampling to give all signals equal sampling frequency, filtering out extremely low frequency below 0.1 Hz at least to remove the DC drift using a high-pass filter, and interpolating the high distributed channels among all channels. This dataset can be used for the performance evaluation of artifact removal methods and analysis of mental states with quantitative evaluation via BCI paradigms in a mobile environment.

Code Availability

The MATLAB scripts are available for loading data, for evaluating classification performance or signal quality, and for plotting figures at https://github.com/youngeun1209/MobileBCI_Data.

References

1. Barthélémy, D., Grey, M. J., Nielsen, J. B., & Bouyer, L. Involvement of the corticospinal tract in the control of human gait. *Prog. Brain Res.* **192**, 181–197 (2011).
2. Jahn, K. *et al.* Brain activation patterns during imagined stance and locomotion in functional magnetic resonance imaging. *Neuroimage* **22**, 1722–1731 (2004).
3. Lee, M.-H., Fazli, S., Mehnert, J., & Lee, S.-W. Subject-dependent classification for robust idle state detection using multi-modal neuroimaging and data-fusion techniques in BCI. *Pattern Recognit.* **48**, 2725–2737 (2015).
4. Jeong, J.-H., Shim, K.-H., Kim, D.-J., & Lee, S.-W. Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 1226–1238 (2020).
5. Kwon, O.-Y., Lee, M.-H., Guan, C., & Lee, S.-W. Subject-independent brain-computer interfaces based on deep convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **31**, 3839–3852 (2019).
6. Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* **113**, 767–791 (2002).
7. Artoni, F. *et al.* Unidirectional brain to muscle connectivity reveals motor cortex control of leg muscles during stereotyped walking. *Neuroimage* **159**, 403–416 (2017).
8. Luu, T. P., Nakagome, S., He, Y., & Contreras-Vidal, J. L. Real-time EEG-based brain-computer interface to a virtual avatar enhances cortical involvement in human treadmill walking. *Sci. Rep.* **7**, 8895 (2017).
9. Debener, S., Emkes, R., De Vos, M., & Bleichner, M. Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Sci. Rep.* **5**, 16743 (2015).
10. Jeong, J.-H., Kwak, N.-S., Guan, C., & Lee, S.-W. Decoding Movement-Related Cortical Potentials Based on Subject-Dependent and Section-Wise Spectral Filtering. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 687–698 (2020).
11. Bulea, T. C., Prasad, S., Kilicarslan, A., & Contreras-Vidal, J. L. Sitting and standing intention can be decoded from scalp EEG recorded prior to movement execution. *Front. Neurosci.* **8**, 376 (2014).

12. Kwak, N.-S., Müller, K.-R., & Lee, S.-W. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PLoS One* **12**, e0172578 (2017).
13. Gwin, J. T., Gramann, K., Makeig, S., & Ferris, D. P. Removal of movement artifact from high-density EEG recorded during walking and running. *J. Neurophysiol.* **103**, 3526–3534 (2010).
14. Castermans, T. *et al.* Optimizing the performances of a P300-based brain-computer interface in ambulatory conditions. *IEEE J. Emerg. Sel. Topics Circuits Syst.* **1**, 566–577 (2011).
15. Nordin, A. D., Hairston, W. D., & Ferris, D. P. Dual-electrode motion artifact cancellation for mobile electroencephalography. *J. Neural Eng.* **15**, 056024 (2018).
16. Lee, Y.-E., Kwak, N.-S., & Lee, S.-W. A real-time movement artifact removal method for ambulatory brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 2660–2670 (2020).
17. Kwak, N.-S., Müller, K.-R., & Lee, S.-W. A lower limb exoskeleton control system based on steady state visual evoked potentials. *J. Neural Eng.* **12**, 056009 (2015).
18. Lee, M.-H., Williamson, J., Won, D.-O., Fazli, S. & Lee, S.-W. A high performance spelling system based on EEG-EOG signals with visual feedback. *IEEE Trans. Neural Syst. Rehabil. Eng.* **26**, 1443–1459 (2018).
19. Won, D.-O., Hwang, H.-J., Dähne, S., Müller, K.-R. & Lee, S.-W. Effect of higher frequency on the classification of steady-state visual evoked potentials. *J. Neural Eng.* **13**, 016014 (2015).
20. Lee, M.-H. *et al.* EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience* **8**, giz002 (2019).
21. Yeom, S.-K., Fazli, S., Müller, K.-R., Lee, S.-W. An efficient ERP-based brain-computer interface using random set presentation and face familiarity. *PLoS One* **9**, e111157 (2014).
22. Floriano, A., Diez, P. F., & Bastos-Filho, T. F. Evaluating the influence of chromatic and luminance stimuli on SSVEPs from behind-the-ears and occipital areas. *Sensors* **18**, 615 (2018).
23. Kwak, N.-S., & Lee, S.-W. Error correction regression framework for enhancing the decoding accuracies of ear-EEG brain-computer interfaces. *IEEE Trans. Cybern.* **50**, 3654–3667 (2019).
24. Gramann, K., Gwin, J. T., Bigdely-ShamloDelorme, N., Ferris, D. P. & Makeig, S. Visual evoked responses during standing and walking. *Front. Hum. Neurosci.* **4**, 202 (2010).
25. Wang, Y.-T. *et al.* An online brain-computer interface based on SSVEPs measured from non-hair-bearing areas. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 14–21 (2016).
26. Chen, Y. *et al.* A high-security EEG-based login system with RSVP stimuli and dry electrodes. *IEEE Trans. Inf. Forensic Secur.* **11**, 2635–2647 (2016).
27. Wei, C.-S., Wang, Y.-T., Lin, C.-T., & Jung, T.-P. Toward drowsiness detection using non-hair-bearing EEG-based brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* **26**, 400–406 (2018).
28. Kidmose, P., Looney, D., Ungstrup, M., Rank, M. L., & Mandic, D. P. A study of evoked potentials from ear-EEG. *IEEE Trans. Biomed. Eng.* **60**, 2824–2830 (2013).
29. Bleichner, M. G., Mirkovic, B., & Debener, S. Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison. *J. Neural Eng.* **13**, 066004 (2016).
30. Goverdovsky, V., Looney, D., Kidmose, P., & Mandic, D. P. In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring. *IEEE Sens. J.* **16**, 271–277 (2015).
31. Bleichner, M. G., & Debener, S. Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG. *Front. Hum. Neurosci.* **11**, 163 (2017).
32. Mirkovic, B., Bleichner, M. G., De Vos, M., & Debener, S. Target speaker detection with concealed EEG around the ear. *Front. Neurosci.* **10**, 349 (2016).
33. He, Y., Luu, T. P., Nathan, K., Nakagome, S. & Contreras-Vidal, J. L. A mobile brain-body imaging dataset recorded during treadmill walking with a brain-computer interface. *Sci. Data* **5**, 180074 (2018).
34. Brantley, J. A., Luu, T. P., Nakagome, S., Zhu, F. & Contreras-Vidal, J. L. Full body mobile brain-body imaging data during unconstrained locomotion on stairs, ramps, and level ground. *Sci. Data* **5**, 180133 (2018).
35. Wagner, J. *et al.* High-density EEG mobile brain/body imaging data recorded during a challenging auditory gait pacing task. *Sci. Data* **6**, 211 (2019).

- 36.** Lee, Y.-E., Shin, G.-H., Lee, M., & Lee, S.-W. Mobile BCI dataset of scalp- and ear-EEGs with ERP and SSVEP paradigms while standing, walking, and running. *Open Science Framework* <https://doi.org/10.17605/OSF.IO/R7S9B> (2021).
- 37.** Kleiner, M., Brainard, D., & Pelli, D. What's new in Psychtoolbox-3?. *Perception* **36**, 14 (2007).
- 38.** Lee, M.-H., Williamson, J., Lee, Y.-E. & Lee, S.-W. Mental fatigue in central-field and peripheral-field steady-state visually evoked potential and its effects on event-related potential responses. *Neuroreport* **29**, 1301–1308 (2018).
- 39.** Parini, S., Maggi, L., Turconi, A. C., & Andreoni, G. A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication. *Comput. Intell. Neurosci.* **2009**, 864564 (2009).
- 40.** Snyder, K. L., Kline, J. E., Huang, H. J., & Ferris, D. P. Independent component analysis of gait-related movement artifact recorded using EEG electrodes during treadmill walking. *Front. Hum. Neurosci.* **9**, 639 (2015).
- 41.** Kline, J. E., Huang, H. J., Snyder, K. L., & Ferris, D. P. Isolating gait-related movement artifacts in electroencephalography during human walking. *J. Neural Eng.* **12**, 046022 (2015)
- 42.** Nordin, A. D., Hairston, W. D., & Ferris, D. P. Human electrocortical dynamics while stepping over obstacles. *Sci. Rep.* **9**, 4693 (2019).
- 43.** Edmunds, K. J. *et al.* Cortical recruitment and functional dynamics in postural control adaptation and habituation during vibratory proprioceptive stimulation. *J. Neural Eng.* **16**, 026037 (2019).
- 44.** Krepki, R., Blankertz, B., Curio, G., & Müller, K.-R. The Berlin Brain-Computer Interface (BBCI)—towards a new communication channel for online control in gaming applications. *Multimed. Tools Appl.* **33**, 73–90 (2007).
- 45.** Kothe, C. A. & Makeig, S. BCILAB: a platform for brain–computer interface development. *J. Neural Eng.* **10**, 056014 (2013).
- 46.** Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
- 47.** He, P., Wilson, G., & Russell, C., Removal of ocular artifacts from electro-encephalogram by adaptive filtering. *Med. and Biol. Eng. and Comput.* **42**, 407–412 (2004).
- 48.** Pernet, C. R. *et al.* EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci. Data* **6**, 1–5 (2019).
- 49.** Lin, Y.-P., Wang, Y., & Jung, T.-P. Assessing the feasibility of online SSVEP decoding in human walking using a consumer EEG headset. *J. NeuroEng. Rehabil.* **11**, 119 (2014).
- 50.** Zink, R., Hunyadi, B., Van Huffel, S., & De Vos, M. Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *J. Neural Eng.* **13**, 046017 (2016).
- 51.** Barollo, F. *et al.* Postural control adaptation and habituation during vibratory proprioceptive stimulation: an HD-EEG investigation of cortical recruitment and kinematics. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 1381–1388 (2020).
- 52.** Schimmel, H. The (\pm) reference: Accuracy of estimated mean components in average response studies. *Science* **157**, 92–94 (1967).
- 53.** Vos, M. D., Gandras, K. & Debener, S. Towards a truly mobile auditory brain-computer interface: Exploring the P300 to take away. *Int. J. Psychophysiol.* **91**, 46–53 (2014).
- 54.** Nakanishi, M., Wang, Y., Wang, Y.-T., Mitsukura, Y., & Jung, T.-P. Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate. *PLoS One* **9**, e99235 (2014).
- 55.** Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).

Acknowledgements

This work was supported by the Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean Government, No. 2017-0-00451: Development of BCI based Brain and Cognitive Computing Technology for Recognizing User's Intentions using Deep Learning, No. 2015-0-00185: Development of Intelligent Pattern Recognition Softwares for Ambulatory Brain-Computer Interface, and No. 2019-0-00079: Artificial Intelligence Graduate School Program (Korea University). We would like to express our sincere gratitude to Y.-H. Kang and D.-Y. Lee for their assistance in data collection, and N.-S. Kwak for his advice while designing the experiment.

Author Contributions Statement

Y.-E.L. contributed to the design of the experiment, data collection, software programming, data validation, and preparation of the manuscript. G.-H.S. contributed to data collection, software programming, data validation, and preparation of the manuscript. M.L. contributed to data validation and preparation of the manuscript. S.-W.L. contributed to supervision of the project and editing the manuscript.

Competing Interests

The authors declare no competing interests.

Figures & Tables

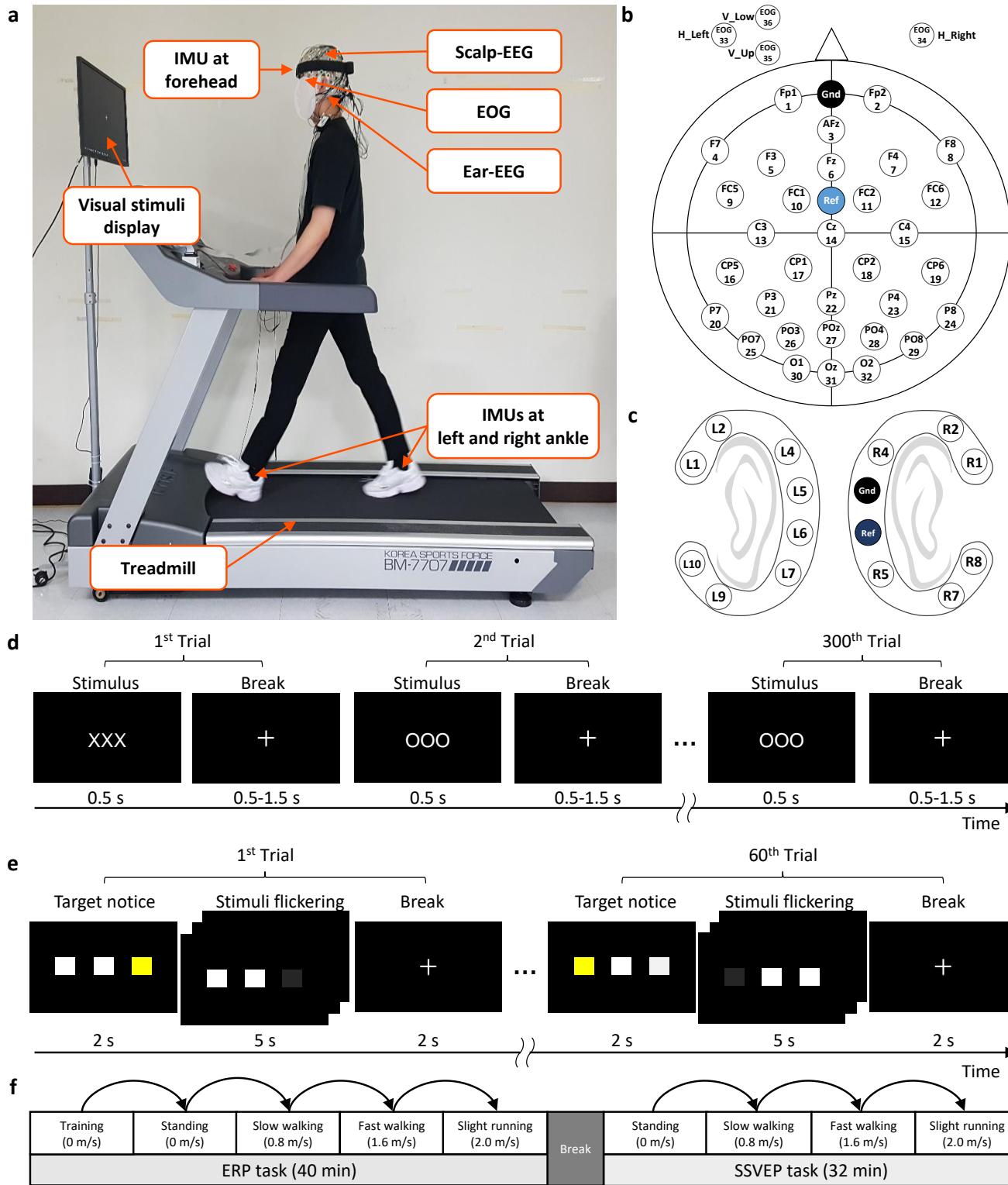


Figure 1. Experimental design. (a) Experimental setup while standing (0 m/s), slow walking (0.8 m/s), fast walking (1.6 m/s), and slight running (2.0 m/s) on the treadmill, wearing scalp-EEG, ear-EEG, EOG, and IMUs. Informed consent was obtained from the participant for publishing the figure. Channel placement of (b) scalp-EEG with EOG and (c) ear-EEG. Experimental paradigms for (d) ERP paradigm with 300 trials and (e) SSVEP paradigm with 60 trials. (f) Experimental procedure.

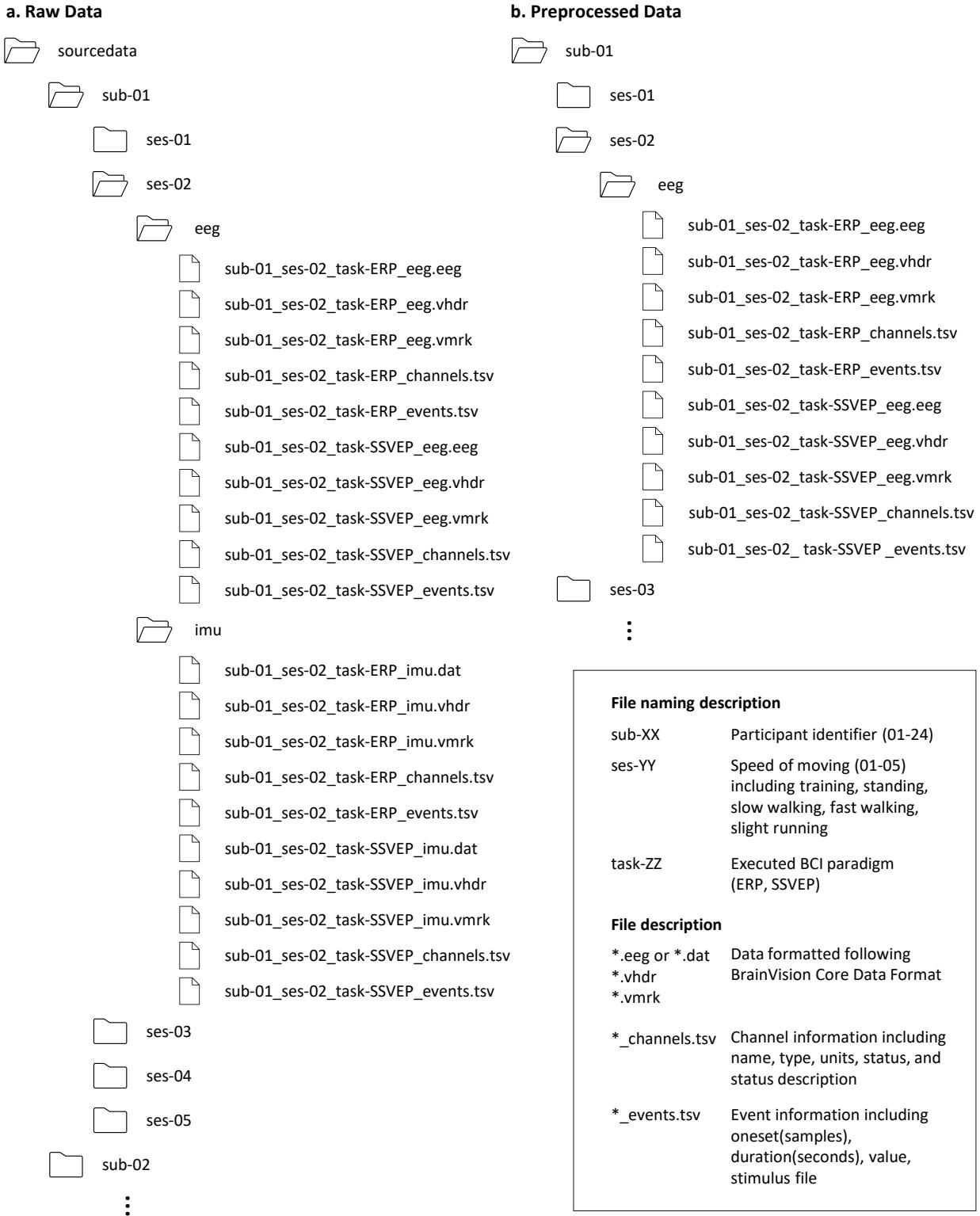


Figure 2. Data folder structure. The folders and files are described, including (a) raw data and (b) preprocessed data in the data repository. The folder was named with ‘sub-XX’, ‘ses-YY’, and modality, and the file was named with ‘sub-XX_ses-YY_task-ZZ_WW’. The ‘sub-XX’ indicated the participant identifier, including 1-24, the ‘ses-YY’ indicated the session number, including training(01), standing(02), slow walking(03), fast walking(04), and slight running(05), the ‘task-ZZ’ indicated executed BCI paradigms including ERP and SSVEP, and the modality ‘WW’ indicated the modality of each data, including EEG (scalp-EEG and ear-EEG) and IMU.

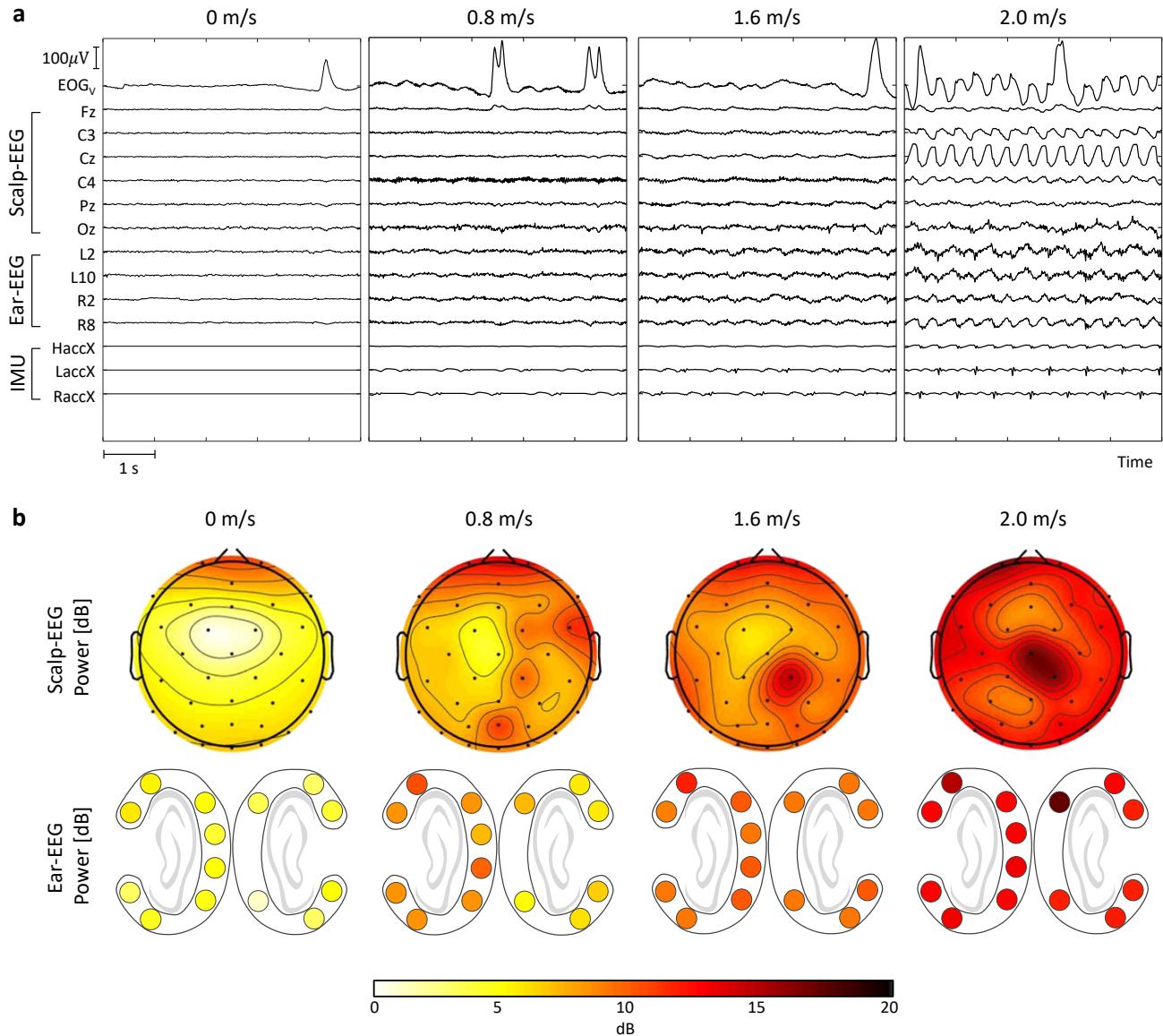


Figure 3. Examples of the signals and topography at different speeds. (a) Time-synchronized subset of scalp-EEG, ear-EEG, and IMU data for 5 s while moving at different speeds of 0, 0.8, 1.6, and 2.0 m/s. The EOG_v channel was calculated by subtracting lower VEOG from upper VEOG. (b) EEG power topography in each channel of scalp-EEG and ear-EEG.

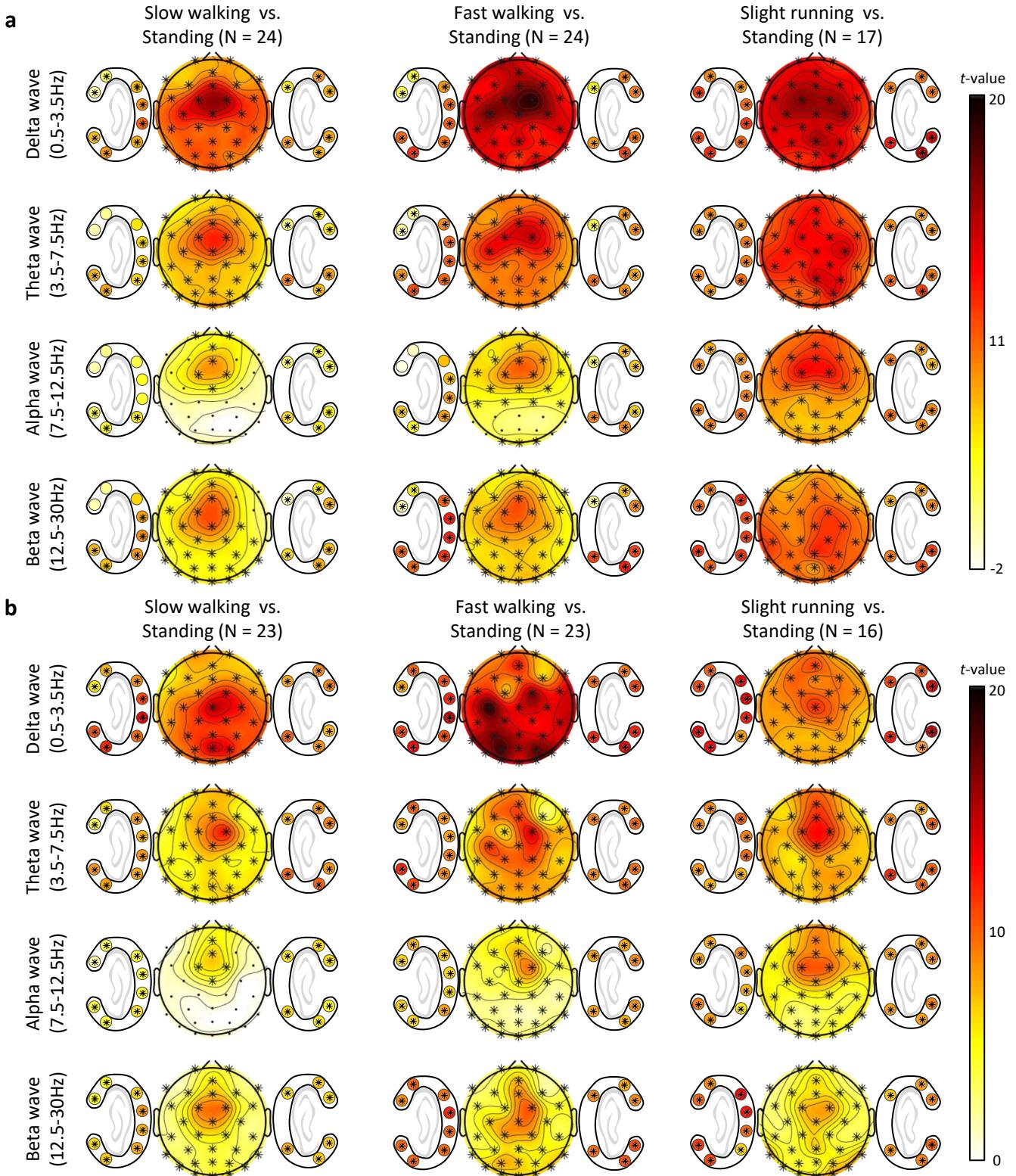


Figure 4. Statistical differences of PSD in each frequency band for scalp- and ear-EEGs between standing and other speeds while (a) ERP and (b) SSVEP. The colored topological maps indicate t-values and the electrodes in cluster showing a statistically significant effect on spectral power between the data of corresponding speeds are marked with black asterisk ($p < 0.05$, cluster-based correction for multiple comparison).

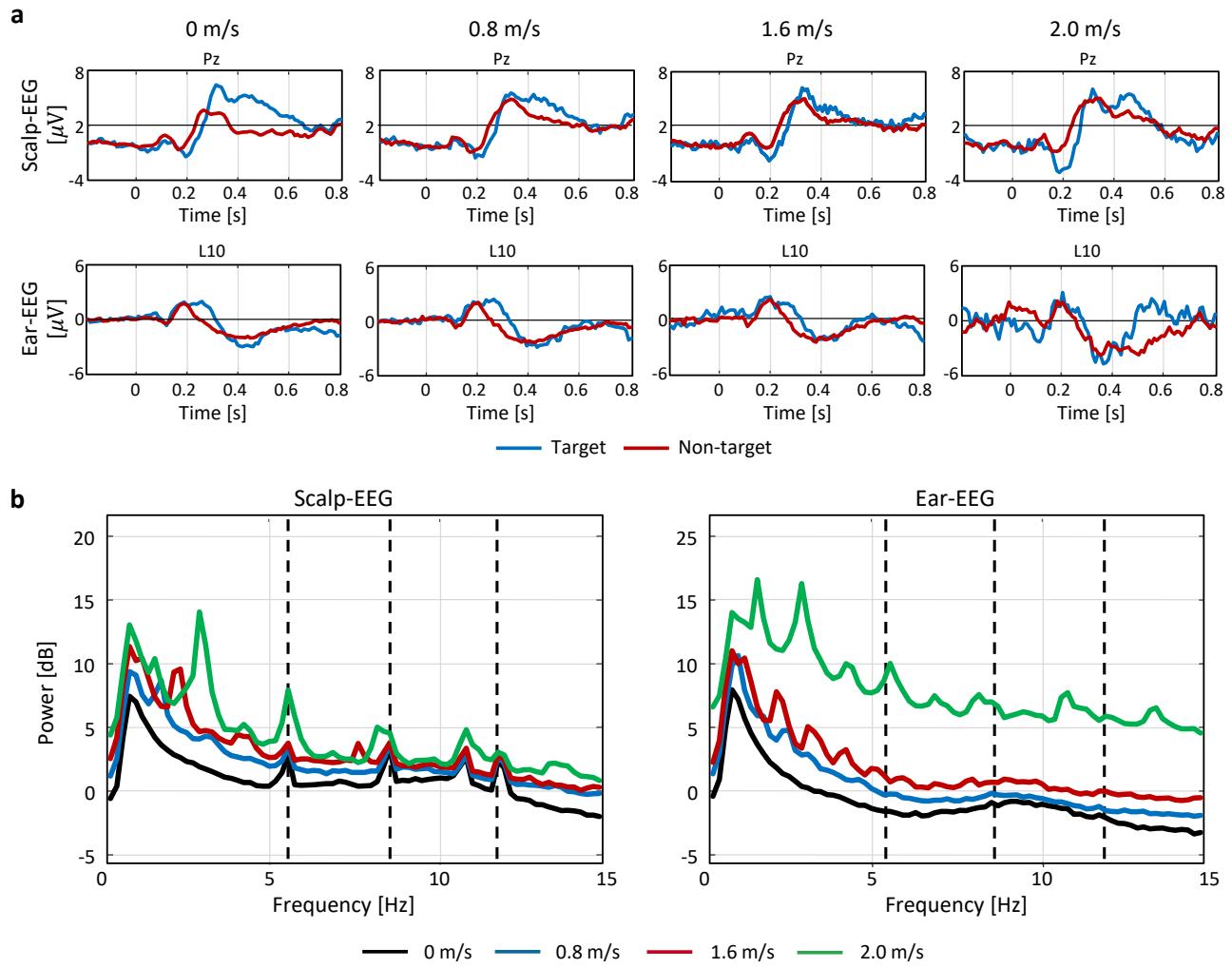


Figure 5. Grand average of all participants of ERP and SSVEP waveforms according to four different speeds of 0, 0.8, 1.6, and 2.0 m/s. (a) Grand average of all participants of ERP waves of target and non-target stimuli in scalp-EEG and ear-EEG for 1 s from -200–800 ms according to the trigger at different speeds. (b) Grand average PSD of all participants for SSVEP in scalp-EEG (left) and ear-EEG (right) at different speeds. The dash line indicated the target frequency, such as 5.45, 8.57, and 12 Hz.

Modality	The number of channels	Description
Scalp-EEG	36	32 EEG channels and 4 EOG channels
Ear-EEG	14	8 channels on the left side and 6 channels on the right side
IMU	27	3 devices (head, left ankle, and right ankle) 9-axis device containing accelerometers, gyroscopes, and magnetometers

Table 1. Description of channel types. Asterisk indicates significance levels of 1% between the performance at 0 m/s and corresponding speed.

Measure	AUC				SNR			
	Speed	0 m/s	0.8 m/s	1.6 m/s	2.0 m/s	0 m/s	0.8 m/s	1.6 m/s
s1	0.99	0.83	0.77	0.59	1.03	1.01	0.96	0.99
s2	0.85	0.68	0.58	0.61	0.89	0.91	0.95	0.97
s3	0.78	0.76	0.68	0.56	0.77	0.83	0.88	0.96
s4	0.86	0.84	0.69	0.44	0.93	0.92	0.96	0.95
s5	0.94	0.75	0.60	0.63	1.05	1.06	1.00	1.03
s6	0.98	0.86	0.80	0.63	1.15	1.60	1.47	1.11
s7	0.93	0.81	0.59	0.62	1.01	1.12	1.00	0.97
s8	0.89	0.62	0.68	0.59	0.99	1.18	1.35	1.11
s9	0.92	0.81	0.70	0.54	0.99	0.99	1.02	0.90
s10	0.97	0.70	0.58	0.56	1.05	1.18	1.13	1.04
s11	0.80	0.73	0.59	0.50	0.85	0.89	0.99	0.96
s12	0.80	0.81	0.64	0.54	1.01	0.92	1.01	0.99
s13	0.80	0.75	0.70	0.75	0.91	1.30	1.33	1.03
s14	0.91	0.66	0.57	0.59	1.08	1.25	1.05	0.93
s15	0.94	0.78	0.50	0.57	0.88	0.92	0.91	0.99
s16	0.94	0.85	0.73	0.55	0.88	1.06	1.04	0.95
s17	0.98	0.83	0.78	0.59	0.95	1.04	1.02	1.11
s18	0.92	0.84	0.69	-	1.07	1.01	1.04	-
s19	0.82	0.79	0.68	-	0.94	1.18	1.04	-
s20	0.90	0.71	0.68	-	0.92	1.23	1.10	-
s21	0.75	0.69	0.71	-	0.74	0.92	0.92	-
s22	0.95	0.85	0.72	-	0.93	0.98	0.99	-
s23	0.94	0.77	0.64	-	0.98	1.28	1.20	-
s24	1.00	0.86	0.72	-	0.90	1.00	1.00	-
AVG	0.90	0.77*	0.67*	0.58*	0.95	1.07*	1.06*	1.00
STD	0.07	0.07	0.07	0.06	0.09	0.17	0.14	0.06

Table 2. AUC and SNR of ERP in scalp-EEG. Asterisk indicates significance levels of 5% between the performance at 0 m/s and corresponding speed.

Measure	AUC				SNR			
	Speed	0 m/s	0.8 m/s	1.6 m/s	2.0 m/s	0 m/s	0.8 m/s	1.6 m/s
s1	0.52	0.55	0.52	0.49	1.28	0.92	1.06	0.97
s2	0.71	0.69	0.52	0.53	0.93	0.94	0.97	1.04
s3	0.66	0.57	0.57	0.54	0.77	0.92	0.86	0.97
s4	0.54	0.54	0.53	0.58	1.05	1.13	0.98	0.83
s5	0.94	0.75	0.63	0.53	0.97	0.99	1.01	0.87
s6	0.86	0.70	0.65	0.51	1.20	1.17	1.09	1.04
s7	0.85	0.74	0.57	0.56	1.06	1.08	1.01	1.01
s8	0.57	0.58	0.59	0.51	1.20	1.06	1.03	1.01
s9	0.81	0.76	0.70	0.60	0.93	1.01	1.01	0.95
s10	0.72	0.64	0.58	0.49	0.98	0.98	0.98	0.92
s11	0.69	0.47	0.55	0.47	1.07	0.94	0.96	1.11
s12	0.64	0.60	0.56	0.54	0.98	0.94	1.00	0.97
s13	0.46	0.46	0.62	0.49	0.97	1.08	1.06	1.02
s14	0.77	0.58	0.46	0.49	1.09	1.09	0.95	0.97
s15	0.76	0.56	0.51	0.55	0.97	0.97	0.94	0.96
s16	0.79	0.71	0.57	0.51	1.01	1.06	1.00	0.99
s17	0.87	0.80	0.67	0.64	1.01	0.97	1.00	0.99
s18	0.51	0.68	0.58	-	0.99	1.03	0.99	-
s19	0.72	0.69	0.73	-	0.88	1.24	0.94	-
s20	0.51	0.41	0.58	-	2.24	1.08	0.88	-
s21	0.69	0.64	0.52	-	0.88	1.01	0.97	-
s22	0.88	0.68	0.55	-	0.95	1.01	0.98	-
s23	0.80	0.69	0.56	-	0.94	0.97	0.97	-
s24	0.90	0.73	0.64	-	0.95	0.96	0.96	-
AVG	0.72	0.63*	0.58*	0.53*	1.06	1.02	0.98	0.98
STD	0.14	0.10	0.06	0.04	0.27	0.08	0.05	0.06

Table 3. AUC and SNR of ERP in ear-EEG. Asterisk indicates significance levels of 5% between the performance at 0 m/s and corresponding speed.

Measure	Accuracy (%)				SNR			
Speed	0 m/s	0.8 m/s	1.6 m/s	2.0 m/s	0 m/s	0.8 m/s	1.6 m/s	2.0 m/s
s1	100	98.33	98.33	100	4.60	3.82	2.27	2.48
s2	100	100	98.33	90.00	3.93	3.42	3.01	2.91
s3	96.67	90.00	83.33	33.33	1.52	1.69	1.35	1.59
s4	100	100	100	93.33	2.43	2.09	2.06	1.79
s5	96.67	96.67	96.67	31.67	3.79	3.42	3.42	1.63
s6	100	81.67	75.00	35.00	2.35	2.18	1.88	1.36
s7	100	98.33	96.67	35.00	4.12	3.68	2.87	2.66
s8	96.67	81.67	65.00	41.67	1.14	1.31	1.01	1.20
s9	98.33	81.67	98.33	33.33	2.46	1.66	1.90	2.31
s10	100	86.67	78.33	35.00	3.42	1.88	1.90	1.74
s11	100	86.67	81.67	50.00	2.19	1.71	1.99	2.05
s12	100	96.67	90.00	60.00	2.33	2.16	1.91	1.45
s13	80.00	66.67	80.00	71.67	1.32	1.23	1.11	1.45
s14	100	96.67	90.00	36.17	2.87	2.55	2.22	2.19
s15	100	100	100	98.33	2.87	2.82	2.39	2.57
s16	78.33	93.33	96.67	-	2.72	2.22	2.09	-
s17	80.00	70.00	56.67	-	1.84	1.35	1.13	-
s18	91.67	63.33	48.33	31.67	1.43	1.32	1.17	1.63
s19	33.33	33.33	31.67	-	1.17	1.15	0.99	-
s20	45.00	48.33	50.00	-	1.32	1.27	1.19	-
s21	96.67	91.67	98.33	-	2.74	2.60	2.81	-
s22	100	100	95.00	-	2.81	2.67	2.29	-
s23	46.67	50.00	46.67	-	1.24	1.02	1.13	-
AVG	88.70	83.12*	80.65*	54.76*	2.64	2.14*	1.92*	1.94*
STD	19.52	18.68	20.38	25.84	0.99	0.84	0.68	0.51

Table 4. Accuracy and SNR of SSVEP in scalp-EEG. Asterisk indicates significance levels of 5% between the performance at 0 m/s and corresponding speed.

Measure	Accuracy (%)				SNR			
Speed	0 m/s	0.8 m/s	1.6 m/s	2.0 m/s	0 m/s	0.8 m/s	1.6 m/s	2.0 m/s
s1	96.67	73.33	46.67	26.67	2.06	1.63	1.32	1.09
s2	61.67	41.67	36.67	33.33	1.37	1.10	0.86	1.10
s3	53.33	46.67	36.67	33.33	1.15	1.25	0.97	2.04
s4	55.00	41.67	33.33	43.33	1.11	0.96	0.94	1.50
s5	68.33	68.33	53.33	28.33	1.57	1.27	1.08	1.31
s6	41.67	38.33	40.00	33.33	1.04	1.08	1.07	1.19
s7	58.33	48.33	41.67	33.33	1.18	1.21	1.11	1.29
s8	58.33	55.00	41.67	25.00	1.08	1.11	1.14	1.12
s9	60.00	51.67	38.33	35.00	1.17	1.02	0.92	1.55
s10	51.67	40.00	38.33	31.67	1.11	1.06	1.05	1.14
s11	35.00	33.33	36.67	35.00	1.07	0.98	0.95	1.21
s12	75.00	55.00	38.33	36.67	1.46	1.40	1.12	1.10
s13	46.67	38.33	50.00	36.67	1.09	1.09	1.03	1.22
s14	51.67	41.67	35.00	36.17	1.19	0.94	0.95	1.24
s15	50.00	38.33	35.00	36.67	1.42	1.18	0.97	1.01
s16	48.33	48.33	46.67	-	1.14	1.08	1.06	-
s17	46.67	38.33	35.00	-	1.05	0.93	0.93	-
s18	48.33	45.00	30.00	28.33	1.13	0.87	1.13	1.31
s19	40.00	25.00	35.00	-	0.98	0.89	1.04	-
s20	40.00	30.00	35.00	-	1.02	0.93	0.93	-
s21	63.33	50.00	55.00	-	1.06	1.14	1.11	-
s22	43.33	50.00	36.67	-	1.27	1.05	1.01	-
s23	30.00	31.67	35.00	-	1.05	1.00	1.08	-
AVG	53.19	44.78*	39.57*	33.30*	1.21	1.09*	1.03*	1.27
STD	13.93	11.11	6.39	4.43	0.23	0.17	0.10	0.24

Table 5. Accuracy and SNR of SSVEP in ear-EEG. Asterisk indicates significance levels of 5% between the performance at 0 m/s and corresponding speed.

A streamable large-scale clinical EEG dataset for Deep Learning

Dung Truong¹, Manisha Sinha², Kannan Umadevi Venkataramu², Michael Milham³, Arnaud Delorme^{1,4}

Abstract— Deep Learning has revolutionized various fields, including Computer Vision, Natural Language Processing, as well as Biomedical research. Within the field of neuroscience, specifically in electrophysiological neuroimaging, researchers are starting to explore leveraging deep learning to make predictions on their data without extensive feature engineering. The availability of large-scale datasets is a crucial aspect of allowing the experimentation of Deep Learning models. We are publishing the first large-scale clinical EEG dataset that simplifies data access and management for Deep Learning. This dataset contains eyes-closed EEG data prepared from a collection of 1,574 juvenile participants from the Healthy Brain Network. We demonstrate a use case integrating this framework, and discuss why providing such neuroinformatics infrastructure to the community is critical for future scientific discoveries.

Keywords—Deep Learning, EEG, Neuroinformatics

I. INTRODUCTION

Electroencephalography (EEG) is a neuroimaging technique that measures global-scale synchronous brain activities with high temporal resolution. The portability of EEG enables the neural underpinning of cognition in both typical laboratory settings and natural environments [1]. EEG analysis has significant implications for the understanding of human cognition, neurotechnology, as well as translational research on neuropsychiatric and neurological disorders.

Deep learning (DL) applied to large datasets is a powerful method for extracting abstract patterns from complex digital signals. DL has revolutionized fields such as Computer Vision and Natural Language Processing, and is making impacts on biomedical research [2, 3]. As a data-intensive machine learning paradigm, a particularly important aspect for the training of DL is data availability.

It is not surprising, therefore, that the rise in DL model performance in different communities comes at the point of release of large-scale datasets for that specific community [4]. We believe there is tremendous potential in applying DL directly to EEG data, and that availability of DL-ready large-scale EEG datasets for EEG can accelerate research in this field. Yet, such datasets, when available, are typically not formatted in a way that they can readily be used for DL applications. The preprocessing of such datasets often requires extensive knowledge of EEG processing, therefore limiting the pool of potential DL users.

Here we report the publication of a large-scale EEG dataset from more than a thousand subjects prepared in a format that is readily used by current DL models. This dataset is not only large in scale but also rich in metadata. We provide

several labels associated with the data samples, and users may also leverage clinical information associated with this data – available upon request from the Child Mind Institute (see Methods). We set up a framework that makes the training of models straightforward, as no local data download is needed, and all data is stored on Amazon Web Service. We finally illustrate the use of this dataset in Deep Learning research via a biological sex classification task using Python and MATLAB. To the best of our knowledge, this is the first large-scale EEG dataset formatted for Deep Learning.

II. METHODS

Dataset. The dataset comes from the larger data sharing project Healthy Brain Network (HBN) by the Child Mind Institute [5]. HBN is a continuing initiative focused on creating and sharing a biobank of community data from up to ten thousands of children and adolescents (ages 5-21) enriched by psychopathology assessments. Participants performed various tasks, while multimodal brain imaging (MRI, fMRI, and EEG), eye-tracking, actigraphy, voice and video data were recorded. The full dataset also includes behavioral and cognitive phenotypes, as well as genetics. We selected a subset of EEG data of 1,574 subjects during the resting state task in which participants viewed a fixation cross on the center of the screen and were instructed to open or close their eyes at various points throughout the recording period, which lasted 5 minutes. High-density EEG data were recorded in a sound-shielded room at a sampling rate of 500 Hz with a bandpass of 0.1 to 100 Hz, using a 128-channel EEG geodesic hydrogel system by Electrical Geodesics Inc. (EGI).

Data preprocessing. Although DL may be applied to raw EEG data without any preprocessing [2], we minimally preprocessed the data using a commonly applied data preprocessing pipeline using EEGLAB v2021 [6] functions running on MATLAB 2020b. To avoid eye saccade artifacts, we used only eye-closed data segments (~170s per subject), ignoring the first and last 3 seconds of each eye-closed period (resulting in five periods of 34 seconds). We removed the mean baseline for each data channel, down-sampled the data to 128 Hz, and subsequently band-pass filtered the data between 0.25–25 Hz (FIR filter of order 6601; 0.125 Hz and 25.125 Hz cutoff frequencies (-6 dB); zero phase, non causal). Data were re-referenced to the averaged mastoids and cleaned using the Artifact Subspace Reconstruction EEGLAB plug-in *clean_rawdata* (v2.3) [7], with the options to automatically remove artifact-dominated channels (parameters used were 5 for FlatLineCriterion, 0.7 for ChannelCriterion, and 4 for LineNoiseCriterion). Removed channels were then interpolated using 3-D spline interpolation (EEGLAB *interp.m* function). No bad portions of data were removed. Finally, we segmented eye-closed data periods into non-

¹ Swartz Center for Computational Neuroscience, Institute for Neural Computation, UC San Diego, La Jolla, CA, USA

² Appasamy Associates Private Limited, Azhiyur, Tamil Nadu, India

³ Center for the Developing Brain, Child Mind Institute, New York, 10022, NY, USA

⁴ CerCo CNRS, Paul Sabatier University, Toulouse, France

overlapping 2-s windows: each preprocessed 2-s epoch was used as a sample for our final dataset. Each subject provided about 81 2-s samples (mean 80.8 ± 3.32). To set up our dataset for the benchmark experiment (see Results) we sub-selected 24 channels (out of the 128 available) closest to a montage used in similar work [8]. Each sample in our dataset thus had dimension 24x256 (24 channels and 2(s) x 128(Hz) time points). No further preprocessing was performed. EEGLAB preprocessing scripts are available on the GitHub repository created for this data release (<https://github.com/sccn/HBN-rest-DL>).

Training, validation, and test sets. We split the dataset into training, validation, and test sets in size ratio 60:30:10. This gave 71,300 samples (885 participants) for training, 39,868 samples (492 subjects) for validation, and 16,006 samples (197 subjects) for testing.

Labels. We provide all available non-clinical subject information as the labels, including sex, age, and handedness. Users can select which subject information to use as labels for their models by using the custom label selection script provided in Fig. 1 and Fig. 2. Biological sex includes binary values (0 for Male and 1 for Female), with roughly equal distribution (50%) between the two sexes in train, validation, and test sets. Summary statistics of age and handedness are provided in the respective Table 1 and Table 2 below.

	Train	Validation	Test
Min	5.02	5.04	5.06
Max	21.9	21.7	21.2
Mean	10.5	10.5	10.7
Median	9.74	9.88	10.3
Stdev	3.61	3.68	3.37

Table 1. Summary of age information

	Train	Validation	Test
Min	-100 (Fully Left-Handed)		
Max	100 (Fully Right-Handed)		
Mean	57.9	58.6	54.3
Median	75.6	75.6	76.7
Stdev	50.0	49.4	56.1

Table 2. Summary of handedness information

We also provide the within-subject position of the segment in the continuous data as labels. This is particularly relevant for self-supervised learning tasks where Deep Learning models learn latent representations of the data using pretext tasks such as temporal context prediction where the models learn to predict whether two EEG segments are adjacent to each other in time [9]. Learned representations of EEG samples can then be used for downstream classification tasks and still achieve high prediction accuracy without being trained directly on the categorical labels.

The original subject IDs are also provided so that interested users can submit requests for clinical information to the child mind institute and match them with the subjects' IDs. Clinical phenotypes may be obtained through the COllaborative Informatics and Neuroimaging Suite (COINS) Data Exchange (<http://coins.mrn.org/dx>) or an HBN-dedicated instance of the Longitudinal Online Research and

Imaging System (LORIS) located at <http://data.healthybrainnetwork.org/>. As the HBN phenotypic data are protected by a Data Usage Agreement (DUA), users must complete the agreement, which can be found at http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/sharing.html, and have approved by an authorized institutional official before receiving access. The intent of the HBN DUA is to ensure that data users agree to protect participant confidentiality when handling the high dimensional HBN phenotypic data and that they will agree to take the necessary measures to prevent breaches of privacy.

Deep Learning frameworks. We test our dataset with the Deep Learning frameworks of MATLAB and Python, two popular programming languages in the neuroscientific community. For Python, we focus on the PyTorch framework for its modular, flexible, and customizable design, conducive to experimental Deep Learning projects. Each framework comes equipped with a data handling library. Thus we designed our dataset to be easily loaded using the MATLAB *imageDatastore* [10] and PyTorch *WebDataset* [11] libraries. Using these libraries makes further data processing and augmentation straightforward. Another advantage of these libraries is the possibility of directly streaming the dataset from cloud storage, making the handling and local storing of the dataset minimal.

Direct data streaming from cloud storage. The dataset is hosted on Amazon Web Service (AWS) cloud resource in a public Simple Storage Service (S3) bucket (Fig. 1). Storage and access fees are covered through an agreement with AWS for freely and openly accessible scientific data. A benefit of storing data on S3 buckets is the support for online data streaming. To simplify data loading as well as to support users with limited storage and memory capacity (e.g. where the entire dataset might not fit into memory), we provide a mechanism to stream the dataset individual samples directly from the S3 bucket, using the features of MATLAB *imageDatastore* and PyTorch *WebDataset* libraries. Data users only need to provide the Unified Resource Locator (URL) of the S3 bucket and data will be streamed online during model training. Only a small portion of the data (e.g. a training batch) will be loaded into memory at a time and no local storage of the data is required. MATLAB and Python implementations to support such online data streaming are different, since their respective frameworks require different file and directory configurations, as well as different file extensions to allow speedy data parsing when streaming during model training. As such, the dataset is organized in two subdirectories *matlab* and *python* in the AWS S3 bucket. The Results section provides the specific URLs and example data loading code for the two programming languages.

For MATLAB, each sample is saved in a *.mat* file and the dataset can be streamed using the *imageDatastore* object starting from version 2021b. After creating the training, validation, and/or test set *imageDatastore* objects using their respective S3 URLs, users will need to load the corresponding label files (either from a local download or directly from S3) containing labels for all samples in the sets. We note that model training when streaming the data suffers from a significant slowdown as compared to loading the entire

dataset locally. MathWorks has been notified of the issue and is working on a solution.

For the Python/PyTorch framework, the *WebDataset* library enables online data streaming. Individual samples are saved in *.npy* file format and sample-label matching is done automatically by having a separate label file for each sample, sharing the same file name as the sample file but with a different file extension *.cls*. The training, validation, and test sets are then packaged into POSIX tar archives, and only the URLs to these archives on S3 are needed to create the dataset with complete labels. Note that even though all the samples are contained in a single tarball, the *WebDataset* Python object will dynamically download individual samples as needed.

Public computational resources. For all computation, we used The Neuroscience Gateway (NSG). NSG provides neuroscience researchers with free high-performance computing resources of the Extreme Science and Discoveries Engineering (XSEDE) network through the San Diego Supercomputer Center [12]. NSG comes installed with popular EEG processing, Machine Learning, and Deep Learning toolboxes and libraries, including MATLAB, EEGLAB, Jupyter Notebook, PyTorch, and Tensorflow. NSG comes equipped with V100 NVidia GPU with 32 GB memory and multicore CPUs with more memory space. NSG comes with both web interface and programmatic access through RESTful APIs, with easy-to-management job submission and management scheme. Leveraging this, EEGLAB toolbox comes equipped with a plugin to submit NSG jobs directly on the EEGLAB/MATLAB environment *nsgportal* [13]. This is a public resource and we recommend users use it for DL applications to EEG data.

III. RESULTS

Below we provide access URLs and examples for the version of the dataset prepared for Python and MATLAB. Comprehensive example scripts can be found at the GitHub repository <https://github.com/sccn/HBN-rest-DL>.

A. Python example

Python version of the dataset can be downloaded at https://childmind.s3.us-west-1.amazonaws.com/python/childmind_python.zip. The training, validation, and test set can be streamed individually via PyTorch's Webdataset library after installation using the command *pip install webdataset*. Below we provide examples of Python scripts to stream and work with the training set directly from the S3 bucket. Data users can generalize it for the validation and test set by modifying the file name accordingly.

```
import torch
import webdataset as wds
def selectLabel(x, lbl):
    # function to select desired label
    lbl_idx =
    ["id", "sex", "age", "handedness", "index"].index(lbl.lower())
    x = x.decode("utf-8").split(",")[:lbl_idx]
    return x if lbl_idx == 0 else float(x)
```

```
s3_url = 'https://childmind.s3.us-west-1.amazonaws.com/python/childmind_train.tar' # replace 'train' with 'val' and 'test' accordingly

train_data =
wds.WebDataset(s3_url).decode().map_dict(cls=lambda
x: selectLabel(x, 'ID')).to_tuple("npy", "cls")

# Check out first sample and its label
sample, label = next(iter(train_data))
print(f'Sample size: {sample.shape}') # (24, 256)
print(f'Label: {label}') # NDARFB908HVX
```

Figure 1. Example Python script to stream the dataset from AWS S3 bucket. Users select the desired label by providing the appropriate label name for the second argument of the *selectLabel* function. See GitHub repository <https://github.com/sccn/HBN-rest-DL> for the latest version of the code.

B. MATLAB example

MATLAB version of the dataset can be downloaded at https://childmind.s3.us-west-1.amazonaws.com/matlab/childmind_matlab.zip. Direct data streaming from S3 bucket is demonstrated below in Fig. 2 for the training set and can be generalized for the validation and test set accordingly.

```
setenv('AWS_DEFAULT_REGION', 'us-west-1')

% Custom reader function
load_sample = @(x) x.sample;
readfun = @(x) load_sample(load(x));

% Load training labels
load('s3://childmind/matlab/train_labels.mat', '-mat', 'train_label_info');
label_col = 3; % column index containing sex label
train_labels =
categorical(cell2mat(train_label_info(:, label_col)));

% Create datastore and assign labels
load('s3://childmind/matlab/FileSet_train.mat', '-mat', 'fs_train');
train_imds = imageDatastore(fs_train,
'FileExtensions', '.mat', 'ReadFcn', readfun, 'Labels', train_labels);

% Preview the first sample of the datastore
sample = preview(train_imds);
fprintf('Sample size = %d, %d \n', size(sample));
fprintf('Label: Sex (1 = Female, 0 = Male) of the subject is %s \n', train_imds.Labels(1));
plot(sample');
```

Figure 2. Example MATLAB script to stream the dataset from AWS S3 bucket. Label is selected by providing the column index in *label_col*. See GitHub repository <https://github.com/sccn/HBN-rest-DL> for the latest version of the code.

Benchmark experiment. We ran the dataset on a model modified from VGG-16, a high-performing Computer Vision Convolutional Neural Network, for the biological sex classification task [14]. Per-subject classification accuracy was 87% (95% confidence interval 86.6% - 87.4%), achieving state-of-the-art performance. Examples of model training and evaluation through data streaming can be found in the mentioned GitHub repository.

IV. DISCUSSION

We provided a large EEG dataset that is Deep Learning ready. Researchers can readily download the dataset from the publicly available AWS S3 bucket. We also implemented a mechanism so that no local download is required and data can be used by direct streaming. We demonstrated the integrated framework and a biological sex classification task using the dataset. To the best of our knowledge, this is the first large-scale EEG dataset that is formatted in a Deep Learning ready way, allowing non-EEG experts to process such data.

Caching vs streaming. Based on the size of the data, the MATLAB or Python library may cache the data, so it can be reused across training steps. Even if this is the case, our solution offers advantages over the standard model where the data is downloaded for offline processing. The main advantage is that the data can only be partially downloaded (for example female-only samples). Another advantage is that the data can be aggregated across different datasets.

Future work will focus on further conversion of archived EEG data. The standardization of data organization and metadata annotation formats such as the Brain Imaging Data Format (BIDS) [15] and Hierarchical Event Descriptor (HED) [16] will facilitate the sharing, discoveries, and integration of large-scale datasets. We will work on automated conversion tools from such formats to DL-ready applications.

ACKNOWLEDGMENTS

This work was supported by NIH (R24MH120037-03 and R01NS047293-17) and by MathWorks through the MATLAB Community Toolbox Program.

Expanse supercomputer time was provided via XSEDE allocations and NSG (the Neuroscience Gateway). We thank Amitava Majumdar, Subhashini Sivagnanam, and Kenneth Yoshimoto for providing computational resources.

REFERENCES

- [1] Makeig, S., Gramann, K., Jung, T.P., Sejnowski, T.J., Poizner, H. Linking brain, mind and behavior. *International Journal of Psychophysiology*, 2009. <https://doi.org/10.1016/j.ijpsycho.2008.11.008>
- [2] Roy, Y., et al., Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 2019. **16** 051001.
- [3] LeCun, Y., Bengio, Y., Hinton, G. *Deep learning*. Nature, 2015. 521, 436–444. <https://doi.org/10.1038/nature14539>
- [4] Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. Data and its (dis)contents: A survey of dataset development and use in machine learning research. Patterns, 2021, 2(11), 100336. <https://doi.org/10.1016/j.patter.2021.100336>
- [5] Alexander, L., Escalera, J., Ai, L. et al., An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, 2017. **4**, 170181.
- [6] Delorme, A. and S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 2004. **134**(1): pp. 9-21.
- [7] Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al., Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Trans. Bio-Med. Eng*, 2015. **62**(11), 2553–2567.
- [8] Van Putten, M. J. A. M., Olbrich, S. and Arns, M., Predicting sex from brain rhythms with deep learning. *Scientific Reports*, 2018. **8**.
- [9] Banville, H., Chehab, O., Hyvärinen, A., Engemann, D. A., Gramfort, A., Uncovering the structure of clinical EEG signals with self-supervised learning. *J. Neural Eng.* 2021, 18 046020
- [10] *Getting Started with Datastore*, MATLAB & Simulink. Accessed on: Jan. 25, 2022. [Online]. Available: https://www.mathworks.com/help/matlab/import_export/what-is-a-datastore.html.
- [11] Aizman, A., Maltby, G., Breuel, T. *Efficient PyTorch I/O library for Large Datasets, Many Files, Many GPUs*, PyTorch. Aug. 11, 2020. Accessed on: Jan. 25, 2022. [Online]. Available: <https://pytorch.org/blog/efficient-pytorch-io-library-for-large-datasets-many-files-many-gpus/>.
- [12] Sivagnanam, S., Yoshimoto, K., Carnevale, N., Nadeau, D., Kandes, M., Petersen, T.J., Truong, D., Martínez-Cancino, R., Delorme, A., Makeig, S., & Majumdar, A. Neuroscience Gateway Enabling Large Scale Modeling and Data Processing in Neuroscience Research. Practice and Experience in Advanced Research Computing. 2020. doi.org/10.1145/3311790.3399625.
- [13] Martínez-Cancino, M., Delorme, A., Truong, D., Artoni, F., Kreutz-Delgado, K., Sivagnanam, S., Yoshimoto, K., Majumdar, A., Makeig, S. The Open EEGLAB Portal Interface:High-Performance Computing with EEGLAB. *Neuroimage*, 2021. Volume 224, 116778.
- [14] Truong, D., Milham, M., Makeig, S., and Delorme, A. Deep Convolutional Neural Network Applied to Electroencephalography: Raw Data vs Spectral Feature. *Annu Int Conf IEEE Eng Med Biol Soc*, 2021, pp. 1039-1042, doi: 10.1109/EMBC46164.2021.9630708.
- [15] Gorgolewski, K., Auer, T., Calhoun, V. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data*, 2016. <https://doi.org/10.1038/sdata.2016.44>
- [16] Robbins, K., Truong, D., Jones, A., Callanan, I., and Makeig, S. Building FAIR functionality: Annotating event-related imaging data using Hierarchical Event Descriptors (HED). *Neuroinformatics*, 2021. <https://doi.org/10.31219/osf.io/5fg73>

EEG & EMG dataset for the detection of errors introduced by an active orthosis device

Niklas Kueper^{1,*†}, Kartik Chari^{1,*†}, Judith Bütefür², Julia Habenicht², Su Kyoung Kim¹, Tobias Rossol¹, Marc Tabie¹, Frank Kirchner^{1,3}, and Elsa Andrea Kirchner^{1,2}

¹ Robotics Innovation Centre, German Research Centre for Artificial Intelligence (DFKI), 28359 Bremen, Germany

² Institute of Medical Technology Systems, University of Duisburg-Essen, 47057 Duisburg, Germany

³ Robotics Lab, University of Bremen, 28359 Bremen, Germany

† These authors contributed equally to this work and share first authorship

Correspondence*:

Niklas Kueper
niklas.kueper@dfki.de

Kartik Chari
kartik.chari@dfki.de

ABSTRACT

This paper presents a dataset containing recordings of the electroencephalogram (EEG) and the electromyogram (EMG) from eight subjects who were assisted in moving their right arm by an active orthosis device. The supported movements were elbow joint movements, i.e., flexion and extension of the right arm. While the orthosis was actively moving the subject's arm, some errors were deliberately introduced for a short duration of time. During this time, the orthosis moved in the opposite direction. In this paper, we explain the experimental setup and present some behavioral analyses across all subjects. Additionally, we present an average event-related potential analysis for one subject to offer insights into the data quality and the EEG activity caused by the error introduction. The dataset described herein is openly accessible. The aim of this study was to provide a dataset to the research community, particularly for the development of new methods in the asynchronous detection of erroneous events from the EEG. We are especially interested in the tactile and haptic-mediated recognition of errors, which has not yet been sufficiently investigated in the literature. We hope that the detailed description of the orthosis and the experiment will enable its reproduction and facilitate a systematic investigation of the influencing factors in the detection of erroneous behavior of assistive systems by a large community.

Keywords: EEG, event-related potential, EMG, orthosis, dataset, robotics, BCI, error detection

1 INTRODUCTION

Exoskeletons and orthoses are frequently used to assist or enable human movement (see Kirchner & Bütfür (2022)). They are even able to augment classical therapy approaches such as mirror therapy (Kirchner et al. 2013). The electroencephalogram (EEG) can be used not only to infer the intention to move but also to trigger the assistance provided by an exoskeleton (Kirchner & Bütfür 2022). This has been shown to be very important for successful neuro-rehabilitation (Noda et al. 2012, Hortal et al. 2015). It can also be used to infer the subjective correctness of the behavior of a robot that the human observes or interacts with, as shown in several works such as by Kim et al. (2017, 2020), Chavarriaga et al. (2014). EEG activity can also be used to enable the teaching of subjective preferences to a prosthesis, as proposed by Iturrate et al. (2015). Furthermore, misbehavior of an assistive device can be detected using the error-related potentials (*ErrPs*) that occur when the brain observes errors (van Schie et al. 2004).

Inferring errors from EEG is challenging because it requires asynchronous classification of relevant patterns in the EEG, which often leads to many false positives due to long interaction times with the system or long task times for the system (Omedes et al. 2015, Lopes-Dias et al. 2021a, Spüler & Niethammer 2015, Lopes-Dias et al. 2021b). While most studies focus on how the brain evaluates erroneous behavior using visual stimuli, according to our literature review, there is only one study that uses tactile stimuli to elicit an ErrP (Schiatti et al. 2018). Moreover, it has only been shown that rather complex errors in behavior elicit the ErrP or correlated changes in the frequency domain. It has not yet been studied how the brain responds to simpler errors that are quite obvious and may not require extensive evaluation to assess correctness.

Based on the available literature, we believe that there is a gap in the research on error-related activity. Modalities, other than the visual modality, should also be studied more closely during the interaction or as a source of feedback in terms of what activities

they induce in the brain. It is also necessary to better understand what complexity in task, response, or interaction errors is required to elicit the ErrP. Furthermore, methods for continuous classification of error-related activity and methods that allow to distinguish between different, partially overlapping EEG activities, should be explored more intensively. In this regard, there is a particular lack of openly accessible data that would allow a larger research community to contribute. Furthermore, the use of robots limits the number of research groups that can conduct research on such problems. For this reason, we recorded a dataset of 8 subjects wearing an active orthosis device that introduces simple errors in its behavior. In addition, details about active elbow orthosis are provided, as it is a fairly simple robotic system that could allow other groups to replicate the work or extend it. Furthermore, the experimental procedure has also been described which would give relevant information about the error complexity, subject instructions, and whether or not subjects should respond explicitly to errors.

We hope that this first open-access dataset will enable broader research on how assistive technology can be improved by using EEG to provide more natural and individualized support for activities of daily living. Such support is very important for rehabilitation (Kornhaber et al. 2018).

The rest of the paper is structured as follows. Section 2 provides detailed information about the experimental design and methods used to record the dataset. It also describes the data format and the folder structure for a better understanding of the dataset. Furthermore, Section 3 presents a preliminary quality analysis of the recorded data in the form of response-time analysis and event-related potential analysis. Finally, in Section 4, we provide an overview of this experiment and discuss future possibilities.

2 METHODS AND EXPERIMENTAL DESIGN

This section provides information about the experimental design including details about

subjects' informed consent, experimental setup & procedure, methods used for data acquisition, and the formatting of the recorded dataset.

2.1 Participants

Eight healthy subjects (four male and four female; average age 21.8 ± 2.4 years; right-handed; students) voluntarily participated in the study. Some days before the start of experiments, all subjects were invited to the lab for a basic introduction and preliminary testing. This included checking the fit of the used orthosis and measuring the head circumference for determining their EEG cap size. All the subjects gave their written informed consent and were told that they could stop the experiment at any time without consequences. The experiment lasted for 4.9 ± 0.6 hours on average and all the subjects received a monetary compensation of 10€ per hour.

2.2 Experimental Setup and Procedure

An overview of all the protocols followed throughout the experiment is provided in this section.

2.2.1 Subject Preparation

Before the start of the experiments, the subjects were prepared with a 64-channel EEG system and an eight-channel EMG system (see Section 2.3.1 and 2.3.2 for details). Furthermore, each subject was fitted with an active orthosis (see Section A) on their right arm as shown in Figure 1-(a), and held a small air-filled ball in their left hand. To trigger support from the orthosis, the subjects were required to express their intention to move by applying a force greater than a start threshold in the movement direction. The required force varied among subjects, depending on their strength. It was ensured that the start thresholds were large enough to prevent unintended starts (refer to Section C). After indicating their intention to move, the subjects were instructed to ease their arm muscles as the orthosis took control of the movement and applied adequate torque at the elbow joint. A

Table 1. Start thresholds for each subject.

Subject Code	Start Threshold (Flexion)	Start Threshold (Extension)
AQ59D	1.0 N · m	1.2 N · m
BY74D	0.8 N · m	1.2 N · m
AC17D	0.8 N · m	1.2 N · m
AW59D	0.7 N · m	1.2 N · m
AY63D	1.0 N · m	1.2 N · m
BS34D	1.2 N · m	1.4 N · m
AJ05D	1.0 N · m	1.2 N · m
AA56D	1.0 N · m	1.2 N · m

comprehensive list of the different thresholds for each subject can be found in Table 1.

2.2.2 Experimental Procedure

The subjects' task was to identify errors in the behavior of the orthosis. These errors were deliberately introduced during flexion or extension movements. Here, the term *error* refers to a momentary change in the direction of orthosis movement for a short duration of time (see Table 2). Furthermore, the term *movement trial* will be used to indicate a complete range of flexion or extension movement.

In the first experimental run, subjects were asked to perform 30 movement trials (15 flexions and 15 extensions) with no errors to obtain a baseline. The experimental run began only after the subjects heard a start phrase from the experimenter. This was followed by a training session where the subjects got familiarized with how the errors felt and were instructed to squeeze the ball in their left hand as soon as they felt an error. In each of the following runs, six errors were introduced in a randomized sequence within 30 movement trials. The order of occurrence of errors was varied individually for each run (see Section 2.2.3 for more details). Overall, as part of the experiment, each subject performed 10 experimental runs with six errors each and one baseline run without any errors.

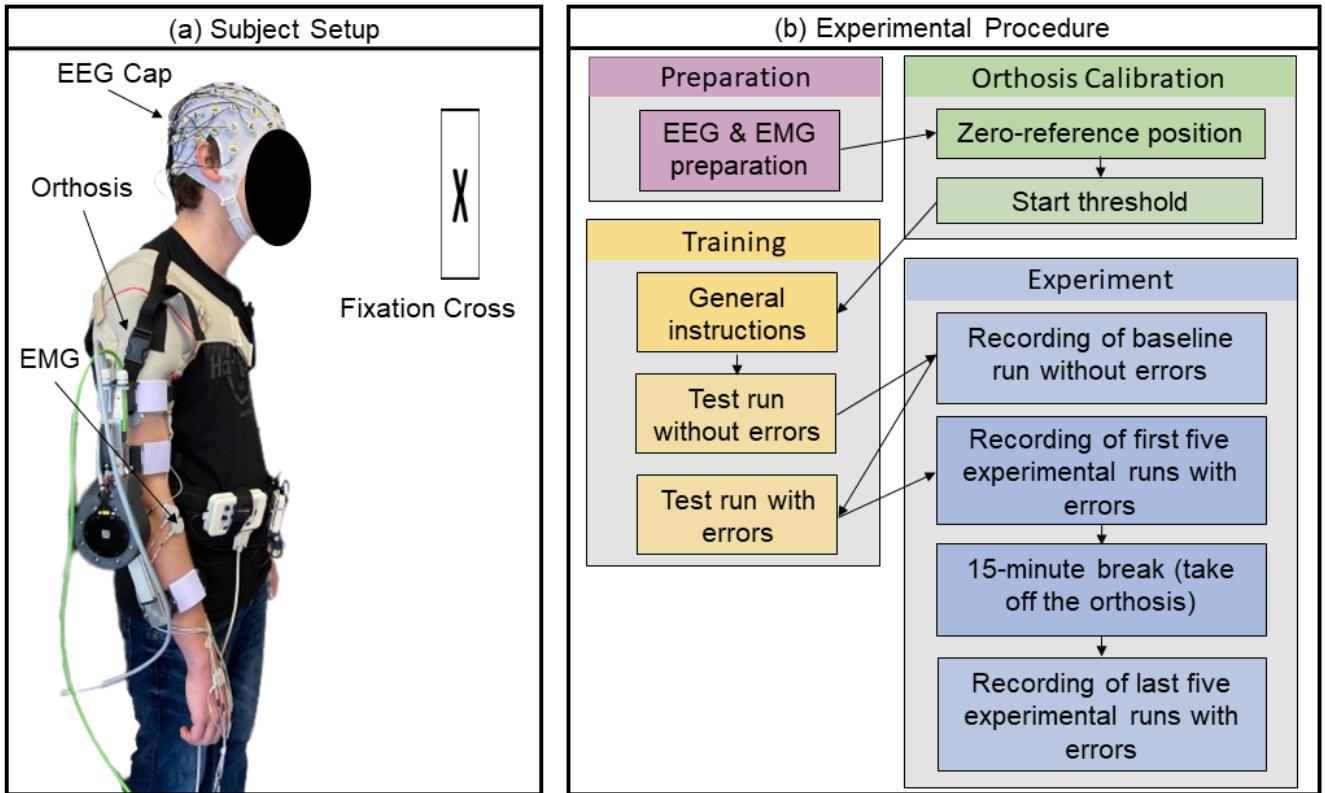


Figure 1. (a) Subject prepared with EEG and EMG electrodes wearing orthosis on their right arm (b) Visualization of the different steps in the experimental procedure.

Before each run, the subjects were reminded to stand still to avoid motion artifacts in the EEG and EMG data. They were also asked to fixate their eyes upon a black cross against the white wall in front of them to avoid eye artifacts in the EEG data. It was also brought to their notice that if, for some reason, they felt an error but forgot to squeeze the ball, they shall just move forward with the run. The subjects were not informed about missed errors during the experiment. At the end of a run, the orthosis motor automatically disabled itself and the subjects were informed about this via a stop phrase. After five runs, a 15-minute break was given during which the subjects could relax and take off the orthosis. A visual summary of the whole experimental procedure is provided in Figure 1-(b).

2.2.3 Error Introduction

At the start of each experimental run, a distinct list of six random values between 1 and 30 (*number of*

Table 2. Orthosis operating parameters.

Parameters	Value
Number of errors	6
Duration of errors	250 ms
Fully Extended position	-10°
Fully Flexed position	-90°
Maximum deviation	0.3°
Mean error position (Flexion)	-42°
Mean error position (Extension)	-58°

errors) was generated. This list, termed as the *error sequence*, followed the following conditions:

- Values 1 and 2 must not be included.
- There must be a gap of at least two between two consecutive numbers from the list.
- The list must be sorted in ascending order.

The values in the list corresponded to the trial numbers in which errors were introduced. The Python3 *random* library (Van Rossum & Drake 2009) was used to generate the random sequence.

Furthermore, whenever a movement trial began, the trial number was matched with all numbers from the generated error sequence. If a match was found, an error was deliberately introduced near the *Mean error position* within the movement trial. The error position varied for flexion and extension with a *Maximum deviation* of 0.3° from the *Mean error position* as mentioned in Table 2.

In practice, if the orthosis were executing flexion before the introduction of the error, it would transition into extension for the specified duration of error (see Table 2) and then resume flexion until the end of the trial and vice-versa.

2.3 Data Acquisition

This section provides detailed information about the methods used for recording the EEG and EMG data. Additionally, it also describes the process of synchronization of these two types of data.

2.3.1 EEG Recording

The EEG data were recorded using the 64-channel LiveAmp64 system from Brain Products GmbH¹ and an ActiCap slim electrode system² with an extended 10-20 layout. The reference electrode was placed at FCz and the GND at AFz.

Great efforts were made to record high-quality EEG data and minimize the noise in the data by keeping the impedances of all 64 electrodes below a threshold of 5 kΩ. This impedance check was performed both prior to and after each experimental run. The EEG data were recorded using the *Recorder* software³ (version 1.25.0001) from Brain Products GmbH. The sampling rate was 500 Hz and the measurement system used hardware filters that limited the bandwidth of the data to a passband of 0.0 Hz - 131.0 Hz.

2.3.2 EMG Recording

To record bipolar EMG data, the ANT mini eego amplifier⁴ was used. The EMG data were recorded with a sampling rate of 1000 Hz using an adapted eego SDK⁵ for Python. Eight channels were used, each measuring the muscle activity of the following muscles on both the arms:

- M. biceps brachii
- M. triceps brachii lateral
- M. triceps brachii long head
- M. flexor digitorum superficialis

Before placing the electrodes, the skin was prepared with Isopropyl alcohol (70% V/V). The electrodes were placed on the muscle belly in accordance with the SENIAM guidelines (Hermens et al. 2000).

2.3.3 Synchronization of EMG & EEG Data

To enable synchronization of EEG and EMG data for offline analysis, an Arduino Nano board and the Sensor & Trigger extension from the EEG system were used to mark start and end time points of the EMG data recordings within the EEG data. The EEG system was used as the main device to enable an alignment of both data streams with respect to each other. Despite the EEG data recordings starting before the EMG, the marked events serve as reference points to align both data streams. With this approach, an average time difference below 8.5 ms between both data streams was achieved after evaluating the synchronicity for all recorded data sets. This result was arrived at by comparing the amount of recorded EMG data against the marked events recorded by the EEG system. Please refer to section 2.4.3 for the specification of the marked events in the EEG data.

¹ <https://www.brainproducts.com/solutions/liveamp/>

² <https://www.brainproducts.com/solutions/acticap/>

³ <https://www.brainproducts.com/downloads/recorder/>

⁴ https://www.ant-neuro.com/products/eego_8

⁵ <https://gitlab.com/smeeze/eego-sdk-pybind11/-/tree/0ace9b329b7cf5f6d1da5d387d0f2a5c07e87ee7>

2.4 Dataset and Format

This section provides a description of the data format, along with detailed information about the dataset and the recorded events.

2.4.1 Data Format

The recorded EEG data follows the BrainVision Core Data Format 1.0, consisting of a binary data file (*.eeg*), a header file (*.vhdr*), and a marker file (*.vmrk*)⁶. For ease of use, the data can be exported into the widely adopted BIDS format as described in Gorgolewski et al. (2016). Furthermore, for data analysis, processing and classification, two popular options are available - MNE (Python)⁷ and EEGLAB (MATLAB)⁸. In contrast, the EMG data is stored in the *.txt* format, where each column represents a separate EMG channel.

2.4.2 Dataset Description

In this section, the dataset's folder structure is explained along with the convention used for naming the files.

2.4.2.1 Folder Structure

This section describes the hierarchical folder structure of the recorded dataset. At the highest level, there are three folders, namely *EEG*, *EMG*, and *Metadata*. The *Metadata* folder contains a *.txt* file for each subject, segregated by a unique code, which consists of meta-information about the subject as well as the measurement sets. In addition to these files, there is also a *short_description.txt* file with some general information about the whole study.

Furthermore, within each of the modality folders (*EEG* or *EMG*), there is an additional level of folders segregated by subject codes. Inside the *EEG* folder, each subject sub-folder is further divided into two sub-folders viz. *data* and *imp*. The *data* folder consists of the actual measurement files as described in section 2.4.1. In total, there is one

baseline set without any errors stored inside another sub-folder named *baseline_without_error* and 10 sets with deliberate errors introduced. Each header file (*.vhdr*) also contains the impedance values of every electrode before the set. Conversely, each header file inside the *imp* folder contains impedance values after each set. All in all, all impedance values, before and after the set, are available within the header files (*.vhdr*). It has to be noted that, for some subjects, an additional set was recorded for safety purposes and included in this dataset under a sub-folder named *additional sets*. For more detailed information, please refer to the Metadata readme files included within the dataset.

2.4.2.2 Naming Convention for Data Files

A consistent naming convention was followed for all our files dividing the filename into five parts. The first part was the date of acquisition in *yyyymmdd* format (e.g. 20230424), followed by the subject code (e.g. AC17D). The third part included the experiment identifier, in this case, *orthosisErrorIjcai*, followed by *multi* indicating that subjects could hear and see (multiple modalities) during the experiments. For baseline runs, the suffix *baseline_set* plus the set number (e.g. 1 or 2) was added while for experimental runs with errors, only the run number was appended at the end (e.g. *set5*). As an example, a filename would look like 20230424_AC17D_orthosisErrorIjcai_multi_set1.txt. It is important to note that, the term *set* was used to represent the data files associated with the corresponding experimental run.

2.4.3 Recorded Events

To keep track of all the events occurring during the set, these events were recorded and stored in marker files (*.vmrk*). The marker files are located within the *data* folder of each subject under *EEG* (see section 2.4.2 for data structure). The first recorded event (after the start of a set) was named *S1* and it marked the start of the EMG recording for synchronization purposes (see section 2.3.3 for detailed information). The event *S1* also occurred at the end of the EMG measurement. The next

⁶ <https://www.brainproducts.com/support-resources/brainvision-core-data-format-1-0/>

⁷ https://mne.tools/stable/reading_raw_data.html

⁸ <https://sccn.ucsd.edu/eeglab/index.php>

recorded event was *S64* which marked the start of flexion movement. Similarly, the start of an extension movement was marked by the event *S32*. In order to mark a trial without errors, the event *S48* was added around the *Mean error position* as mentioned in Table 2. The event *S96* occurred as soon as an error was introduced in the trial. Additionally, if the subject squeezed the ball, the recorded event *S80* was written to the marker file.

3 ANALYSIS OF DATA QUALITY

In the following, we performed some basic analysis on the recorded data to give proof of data quality and to briefly describe evoked EEG activity as well as muscle activity. Furthermore, to ensure data quality, invalid measurement sets were excluded from the data repository. The excluded sets of all subjects are listed below in the form of *subject code, set number* followed by the reason for excluding the set. The excluded sets are as follows:

- AA56D, *set8*: orthosis shutdown by max. current limit.
- AC17D, *baseline_set1*: too much noise in EEG.
- AJ05D, *baseline_set1*: error in event recording.
- AJ05D, *set8*: too much noise in the EEG.
- AJ05D, *set9*: too much noise in the EEG.
- AQ59D, *set1*: orthosis did not start.
- AW59D, *set1*: too much noise in the EEG.
- AY63D, *baseline_set1*: subject played with the air-filled ball during experiment.

Each of the rejected sets was excluded and supplemented by an additional measurement set (as mentioned in Section 2.4.2.1). Although the data quality was kept as high as possible, we observed a 50 Hz noise for some measurement sets and EEG channels, introduced by the orthosis. The 50 Hz noise was also observed in the EMG data.

3.1 Behavioral analysis

For response-time analysis, we analyzed the response times for the incorrect events (error events). As mentioned above, the subjects were

Table 3. Results of response-time (RT) analysis. (A) Median RT for each subject over 10 datasets. (B) Median RT for each set over 8 subjects. $\mu \pm \sigma$: mean \pm standard deviation.

(A) Analysis 1		(B) Analysis 2	
subject	RT	dataset	RT
AQ59D	0.67 s	set 1	0.72 s
BY74D	0.61 s	set 2	0.73 s
AC17D	0.83 s	set 3	0.70 s
AW59D	0.72 s	set 4	0.71 s
AY64D	0.67 s	set 5	0.70 s
BS34D	0.68 s	set 7	0.74 s
AJ05D	0.91 s	set 6	0.80 s
AA56D	0.89 s	set 8	0.76 s
		set 9	0.68 s
		set 10	0.74 s
$\mu \pm \sigma$		0.75 ± 0.11	$\mu \pm \sigma$
			0.73 ± 0.03

instructed to squeeze an air-filled ball after recognizing an error. The time between the error event (*S96*, true label) and response to the event (*S80*) was calculated for all events.

According to the experimental design, we expected a total of 480 responses to error events (6 error events \times 10 datasets \times 8 subjects = 480 error events). However, we found 9 false negative cases (i.e., the ball was not squeezed, even after an error event occurred) and 5 false positive cases (i.e., the ball was squeezed, even when the error event did not occur). Hence, a total of 471 error event-response pairs (480 error events - 9 false negatives) were used to compute response time. We obtained a median value of 0.72 s over 471 error event-response pairs.

We also performed two additional analyses. First, we calculated the response time averaged over all 10 sets for each subject (see, Table 3-(A)). We also calculated the response time averaged over all 8 subjects for each set (see, Table 3-(B)).

3.2 Event Related Potentials (ERP) Analysis

For ERP analysis, we analyzed the EEG data using EEGLAB⁹. We preprocessed the data as follows. The raw EEGs were downsampled to 250 Hz, re-referenced to an average reference, and filtered between 0.1 Hz and 15 Hz. The *FCz* channel, used as a reference in the EEG recording, was recalculated as an EEG channel for ERP analysis. After preprocessing, the artifacts were rejected by visual inspections which means that only clean EEG data were used for EEG segmentation. The artifact-free EEG data were segmented into epochs from 0.1 s to 1 s after each event type (correct/incorrect). Epochs were averaged within each event type with a baseline correction (−0.1 s until stimulus onset).

For averaging epochs, we only used the error events with correct responses i.e., true positive cases (the ball was squeezed when error events occurred). Figure 2 shows the ERPs averaged over all epochs for each event type (S48: correct event, S96: incorrect event) for Subject AQ59D. The ERP morphology, i.e., the shape and distribution on the scalp suggest that the introduction of errors elicits a P300 component, more specifically a P3b (Polich 2007) component. This may be elicited by infrequently occurring odd events to which subjects respond, i.e., task-relevant events (Kirchner et al. 2016). In the case of this subject, we could not observe an error-related potential, which is usually evoked by the recognition of errors.

4 CONCLUSION

We presented and described an open-access dataset of EEG and EMG data obtained from eight subjects who were assisted by an active orthosis device in moving their right arm. Behavioral analysis showed that the errors were very well recognized by the subjects. Introduced errors were a momentary change in the direction of movement of the orthosis for a short amount of time. The errors were simple

and easy to detect. The appearance of an average event-related potential (ErrP) in the form of *P3b* indicates that the subject recognized the erroneous events as odd events. One reason that the error introduction did not elicit an error-related potential in the EEG could be due to the simplicity of the error. On the other hand, the very dominant *P3b* could also overlay the ErrP. These conclusions are very preliminary due to the analysis of only one subject. We hope that the dataset provided and the detailed information about the experimental setup would allow its replication. This would enable the research community to systematically investigate the relationship between odd-event detection and erroneous event evaluation, evoked in the brain. A better understanding of this relationship would help to develop future approaches that could allow automatic adaptation of an assistive device to a subject's subjective needs.

DATA AVAILABILITY STATEMENT

The dataset recorded in this study can be found online on Zenodo^{10 11}.

ETHICS STATEMENT

The studies involving human participants were approved by the local Ethical Committee of the University of Duisburg-Essen, Germany. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTION

EK and SK: designed the study. FK: provided feedback to the study design. NK, JH, JB, and KC: prepared the subjects and conducted the experiments. KC and NK: developed the software for operating the experimental setup. TR: responsible for the mechatronic design of the active orthosis device. TR and MT: set up the low-level control and interface of the elbow motor.

⁹ <https://sccn.ucsd.edu/eeglab/index.php>

¹⁰ Training data: <https://doi.org/10.5281/zenodo.7951044>

¹¹ Test data: <https://doi.org/10.5281/zenodo.7966275>

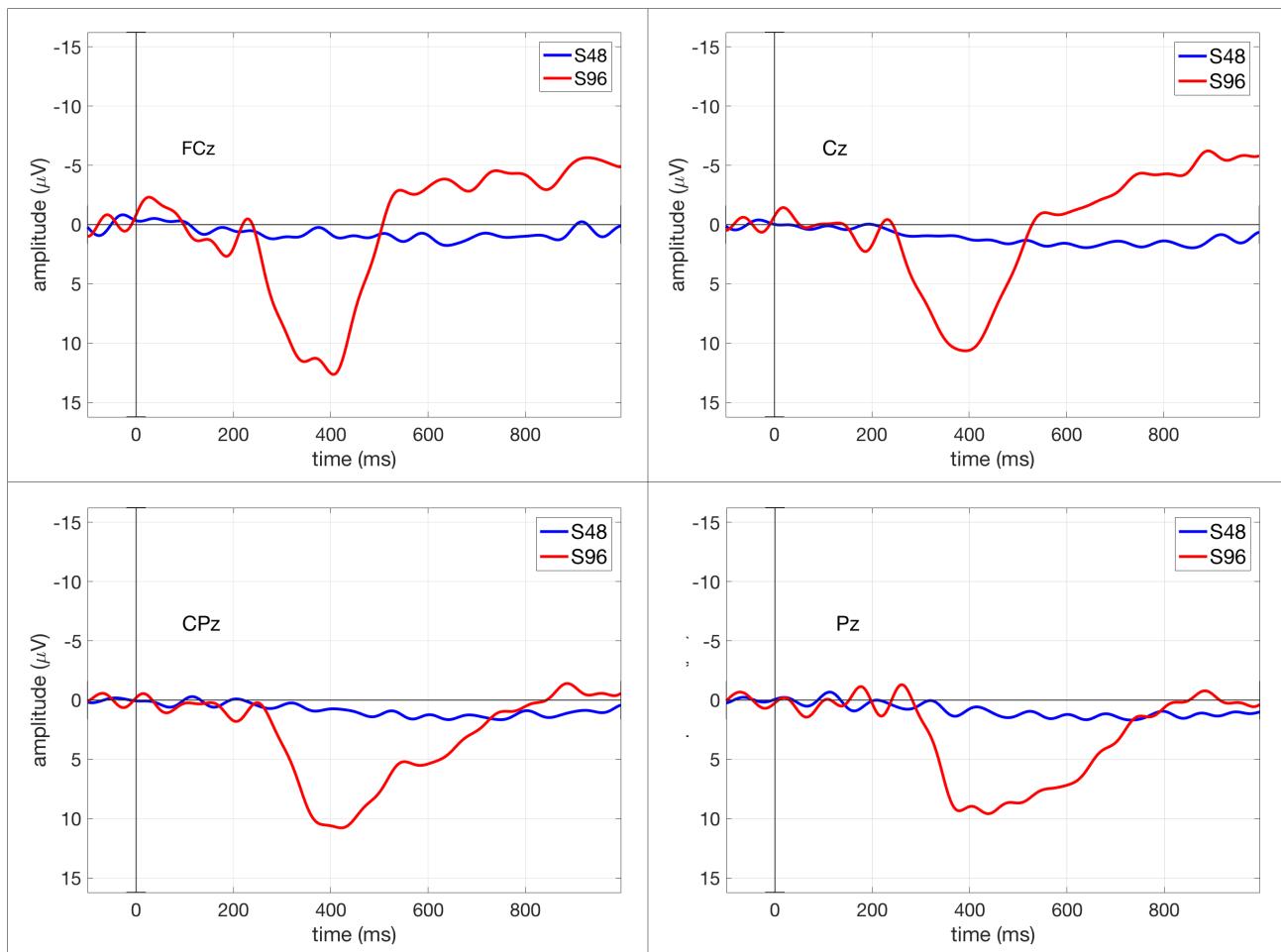


Figure 2. ERPs averaged over all epochs (trials) within each event type: correct event (S48) and incorrect event (S96) for Subject AQ59D.

SK and NK: analyzed the data quality (response-time analysis and ERP analysis) and wrote the results. NK and JB: managed the dataset. NK, KC, JB, JH, and TR: drafted the manuscript. EK: wrote the abstract, introduction and conclusion. EK, FK and KC: provided critical revisions. KC: handled the overall formatting of the manuscript. EK: supervised the study. All authors contributed to the report and approved the submitted version.

FUNDING

This work was funded by the German Federal Ministry of Education and Research (BMBF) within the project M-RoCK (Grant number: 01IW21002) and by the Institute of Medical Technology Systems at the University of Duisburg-Essen.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all the subjects who participated in the study. Furthermore, we would like to thank the students from the University of Duisburg-Essen, namely Variscia Permata Putri, Patrick Fugmann, and Sophia Theisinger who conducted preliminary experiments and provided valuable insights that helped us greatly with this study. Additionally, we would like to acknowledge

Jan-Philipp Brettschneider for his work on the mechanical design of orthosis while interning at DFKI.

APPENDIX

A Active Orthosis Device

The orthosis device (Figure 3) was mainly built from off-the-shelf components and in-house manufactured parts (e.g. 3D printed or water jet cut) resulting in a simple and cost-effective design. The orthosis consists of an upper arm and a forearm structure connected and driven by an actuator. To compensate for the weight of the device and thereby enhance wearer comfort, the upper arm structure is attached to a strap that is sewn on a shoulder orthosis (see Figure 1-(a)).

To transmit forces to the wearer's arm, the elbow orthosis connects to the human body via an upper arm and a forearm interface. The cuffs consist of silicone pads cast on bent, water-jet-cut aluminum sheet parts. The sheet parts transmit the interaction forces between the orthosis and the wearer and can be deformed elastically to an individual arm shape. The silicone pads not only provide soft contact points to the human tissue but also friction between the arm and interfaces to keep the orthosis in place.

To accommodate different wearers, both upper arm and forearm structure lengths can be adjusted. While the upper arm length is fixed after adjustment, the forearm interface is attached to a slider. This prismatic joint adds a passive degree of freedom to the system that not only adjusts the forearm length but also compensates for misalignments between the axes of rotation of the elbow and orthosis. For further individual adjustment, the forearm interface can be inclined.

The actuator is a T-MOTOR AK80-6 with a rated torque of $6 \text{ N}\cdot\text{m}$ ($12 \text{ N}\cdot\text{m}$ peak). This drive belongs to the category of so-called quasi-direct-drives or proprioceptive actuators (Wensing et al. (2017)). These comprise a high torque density motor combined with a low gear ratio transmission. Such actuators feature high intrinsic back-drivability and

simultaneously allow for accurate high-bandwidth force control. The quasi-direct-drive actuation paradigm has been introduced in rehabilitation robotics throughout the last years, e.g., in Lv et al. (2018), Yu et al. (2020). For communication with the AK80-6, we used a USB to CAN interface and the driver software of Shubham Vyas (2023). We operated the drive in position control mode using the onboard low-level controller. In addition to the software limits, a mechanical stop constrains the range of motion of the orthosis device from the hard limit (upper arm and forearm structure in parallel) to a flexion of -130° , for safety reasons.

The orthosis was powered by a laboratory power supply set to a voltage of 24 V and a current of up to 3 A. An emergency stop was connected in series with the actuator in case of malfunctions, and a $2400 \mu\text{F}$, 100 V rated capacitor was also added in parallel with the actuator for transient voltage suppression. Cable glands were used as strain reliefs to protect the data and power cable connectors from damage.

B Response Listener and Event Trigger Board

Whenever the subject felt an error in the orthosis device, they would press the air-filled ball in their left hand. The state of this ball was recorded by the *Response Listener*, a microcontroller that continuously read 16-bit analog inputs and transmitted the pin states, through a serial connection, to the Python script. This script then converted the pin states into single-byte values and sent them serially to the *Event Trigger Board* which is an ATmega328-based microcontroller board (Arduino Nano¹²).

Additionally, other events such as the error introduction, start of flexion or extension, and no-error movement trial were also sent as unique single-byte arrays to the *Event Trigger Board*. This board read the byte values serially and mapped them into transistor-transistor logic (TTL) output signals. These signals were eventually recorded into

¹² <https://store.arduino.cc/products/arduino-nano>

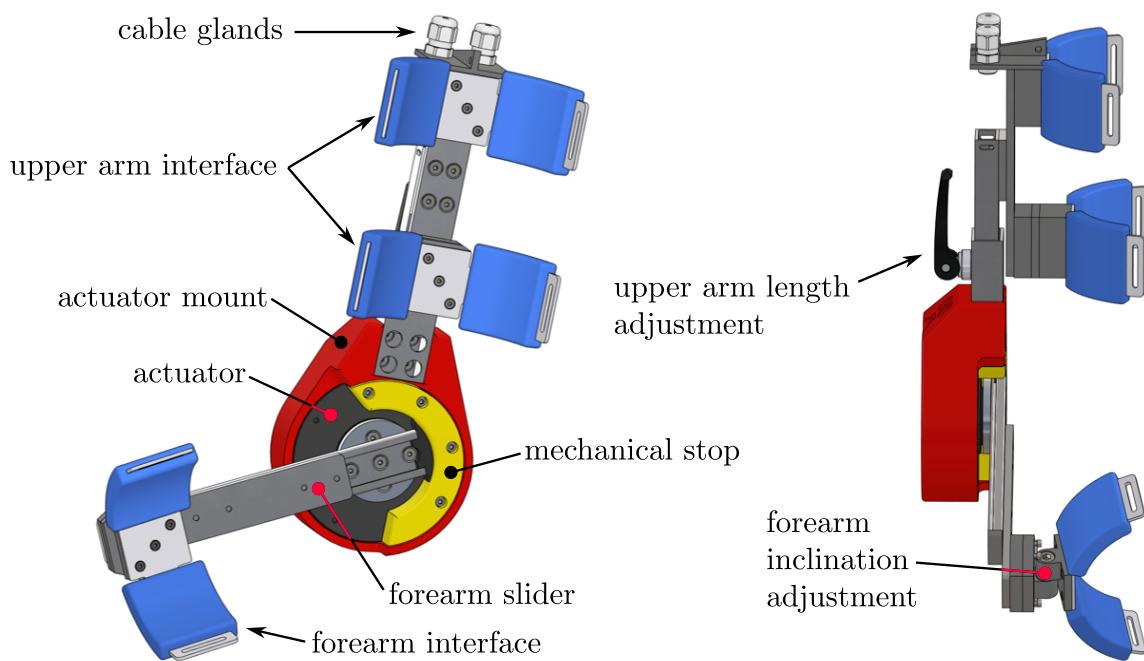


Figure 3. Computer Aided Design (CAD) model side and front view of the active orthosis device. The actuator mount and mechanical stop are colored for better visibility. The connection to the shoulder orthosis can be seen in Figure 1-(a).

the electroencephalogram (EEG) marker files, as described in the *Recorded Events* section of the paper.

C Calibration Sequence and Safety Measures

For calibrating the zero-reference position of the orthosis, the subjects were asked to fully extend their arms to the hardware limit of the elbow motor. After this calibration, the orthosis would move to a more comfortable start position which was termed the *Fully Extended position* (see Figure 4). As a safety measure, the range of motion of the orthosis was restricted between the *Fully Extended position* and *Fully Flexed position* (both inclusive). On reaching the range limits, the elbow motor stopped automatically and held the position until the subjects crossed the start threshold. Moreover, another layer of safety was added by restricting the maximum supply current and forcing motor torque limits.

Furthermore, the start thresholds for flexion and extension were calibrated individually. Initially, the

threshold values were set to $1.0 \text{ N} \cdot \text{m}$ for flexion and $1.2 \text{ N} \cdot \text{m}$ for extension. To determine the appropriate threshold value for flexion, multiple flexion movement trials were performed by the subjects until they arrived at a force that felt neither too weak nor too strong. In contrast, to establish the optimal start threshold for each subject for extension, they were asked to rest their arms in the *Fully Flexed position* and the lowest value at which no unwanted extensions occurred was selected.

REFERENCES

- Chavarriaga, R., Sobolewski, A. & Millán, J. D. R. (2014), ‘Errare machinale est: The use of error-related potentials in brain-machine interfaces’, *Frontiers in Neuroscience* **8**, 208.
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O. et al. (2016), ‘The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments’, *Scientific data* **3**(1), 1–9.

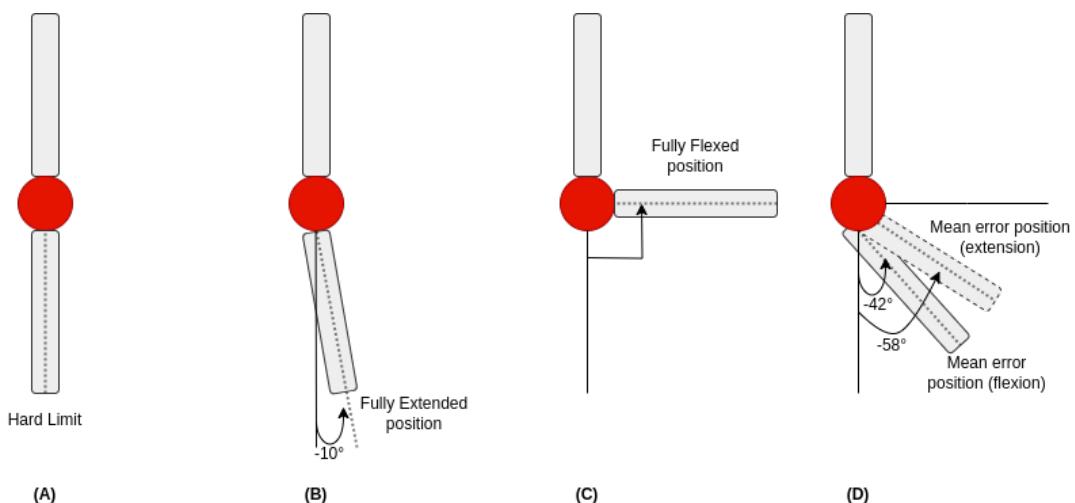


Figure 4. Different operating positions of the orthosis. **(A)** Zero-reference position for the experimental run. **(B)** Comfortable start position and extension limit during movement trial. **(C)** Flexion limit during movement trial. **(D)** Positions around which errors were introduced.

Hermens, H. J., Freriks, B., Disselhorst-Klug, C. & Rau, G. (2000), ‘Development of recommendations for SEMG sensors and sensor placement procedures’, *Electromyography and Kinesiology* **10**(5), 361–374.

Hortal, E., Planelles, D., Resquin, F., Climent, J. M., Azorín, J. M. & Pons, J. L. (2015), ‘Using a brain-machine interface to control a hybrid upper limb exoskeleton during rehabilitation of patients with neurological conditions’, *Journal of neuroengineering and rehabilitation* **12**(1), 1–16.

Iturrate, I., Chavarriaga, R., Montesano, L., Minguez, J. & Millán, J. d. R. (2015), ‘Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control’, *Scientific Reports* **5**, 13893.

Kim, S. K., Kirchner, E. A. & Kirchner, F. (2020), Flexible online adaptation of learning strategy using EEG-based reinforcement signals in real-world robotic applications, in ‘IEEE International Conference on Robotics and Automation (ICRA)’, pp. 4885–4891.

Kim, S. K., Kirchner, E. A., Stefes, A. & Kirchner, F. (2017), ‘Intrinsic interactive reinforcement learning—Using error-related potentials for real world human-robot interaction’, *Scientific reports* **7**(1), 1–16.

Kirchner, E. A. & Bütfür, J. (2022), ‘Towards bidirectional and coadaptive robotic exoskeletons for neuromotor rehabilitation and assisted daily living: a review’, *Current Robotics Reports* pp. 1–12.

Kirchner, E. A., Kim, S. K., Straube, S., Seeland, A., Wöhrle, H., Krell, M. M., Tabie, M. & Fahle, M. (2013), ‘On the applicability of brain reading for predictive human-machine interfaces in robotics’, *PLoS One* **8**(12), e81732.

Kirchner, E. A., Kim, S. K., Tabie, M., Wöhrle, H., Maurus, M. & Kirchner, F. (2016), ‘An intelligent man-machine interface—multi-robot control adapted for task engagement based on single-trial detectability of p300’, *Frontiers in Human Neuroscience* **10**.

URL: <https://www.frontiersin.org/articles/10.3389/fnhum.2016.00291>

Kornhaber, R., Mclean, L., Betihavas, V. & Cleary, M. (2018), ‘Resilience and the rehabilitation of adult spinal cord injury survivors: a qualitative systematic review’, *Journal of advanced nursing* **74**(1), 23–33.

Lopes-Dias, C., Sburlea, A. I., Breitegger, K., Wyss, D., Drescher, H., Wildburger, R. & Müller-Putz, G. R. (2021a), ‘Online asynchronous detection of error-related potentials in participants with a spinal cord injury using a generic classifier’, *Journal of Neural Engineering* **18**(4), 046022.

Lopes-Dias, C., Sburlea, A. I., Breitegger, K., Wyss, D., Drescher, H., Wildburger, R. & Müller-Putz, G. R. (2021b), ‘Online asynchronous detection of error-related potentials in participants with a spinal cord injury using a generic classifier’, *Journal of Neural Engineering* **18**(4), 046022.

- URL:** <https://dx.doi.org/10.1088/1741-2552/abd1eb>
- Lv, G., Zhu, H. & Gregg, R. D. (2018), ‘On the design and control of highly backdrivable lower-limb exoskeletons: A discussion of past and ongoing work’, *IEEE Control Systems Magazine* **38**(6), 88–113.
- Noda, T., Sugimoto, N., Furukawa, J., Sato, M.-a., Hyon, S.-H. & Morimoto, J. (2012), Brain-controlled exoskeleton robot for bmi rehabilitation, in ‘2012 12th ieee-ras international conference on humanoid robots (humanoids 2012)’, IEEE, pp. 21–27.
- Omedes, J., Iturrate, I., Minguez, J. & Montesano, L. (2015), ‘Analysis and asynchronous detection of gradually unfolding errors during monitoring tasks’, *Journal of neural engineering* **12**, 056001.
- Polich, J. (2007), ‘Updating p300: An integrative theory of p3a and p3b’, *Clinical neurophysiology* **118**(10), 2128–2148.
- Schiatti, L., Tessadori, J., Deshpande, N., Barresi, G., King, L. C. & Mattos, L. S. (2018), Human in the Loop of Robot Learning: EEG-Based Reward Signal for Target Identification and Reaching Task, in ‘2018 IEEE International Conference on Robotics and Automation (ICRA)’, pp. 4473–4480.
- Shubham Vyas, Shivesh Kumar, J. B. (2023), ‘dfki-ric-underactuated-lab/mini-cheetah-tmotor- python-can’, *Zenodo*.
- URL:** <https://doi.org/10.5281/zenodo.7671321>
- Spüler, M. & Niethammer, C. (2015), ‘Error-related potentials during continuous feedback: using eeg to detect errors of different type and severity’, *Frontiers in Human Neuroscience* **9**.
- URL:** <https://www.frontiersin.org/articles/10.3389/fnhum.2015.00155>
- Van Rossum, G. & Drake, F. L. (2009), *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA.
- van Schie, H., Mars, R., Coles, M. & Bekkering, H. (2004), ‘Modulation of activity in medial frontal and motor cortices during error observation’, *Nature Neuroscience* **7**(5), 549–554.
- Wensing, P. M., Wang, A., Seok, S., Otten, D., Lang, J. & Kim, S. (2017), ‘Proprioceptive actuator design in the mit cheetah: Impact mitigation and high-bandwidth physical interaction for dynamic legged robots’, *IEEE Transactions on Robotics* **33**(3), 509–522.
- Yu, S., Huang, T.-H., Yang, X., Jiao, C., Yang, J., Chen, Y., Yi, J. & Su, H. (2020), ‘Quasi-direct drive actuation for a lightweight hip exoskeleton with high backdrivability and high bandwidth’, *IEEE/ASME Transactions on Mechatronics* **25**(4), 1794–1802.

EEG Decoding for Datasets with Heterogenous Electrode Configurations using Transfer Learning Graph Neural Networks

Jinpei Han ¹, Xiaoxi Wei ¹, A. Aldo Faisal ^{1,2}

¹ Brain & Behaviour Lab, Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom.

² Chair in Digital Health & Data Science, University of Bayreuth, 95447 Bayreuth, Germany.

E-mail: aldo.faisal@imperial.ac.uk

March 2023

Abstract.

Objective. Brain-Machine Interfacing (BMI) has greatly benefited from adopting machine learning methods for feature learning that require extensive data for training, which are often unavailable from a single dataset. Yet, it is difficult to combine data across labs or even data within the same lab collected over the years due to the variation in recording equipment and electrode layouts resulting in shifts in data distribution, changes in data dimensionality, and altered identity of data dimensions. Our objective is to overcome this limitation and learn from many different and diverse datasets across labs with different experimental protocols. *Approach.* To tackle the domain adaptation problem, we developed a novel machine learning framework combining graph neural networks (GNNs) and transfer learning methodologies for non-invasive Motor Imagery (MI) EEG decoding, as an example of BMI. Empirically, we focus on the challenges of learning from EEG data with different electrode layouts and varying numbers of electrodes. We utilise three MI EEG databases collected using very different numbers of EEG sensors (from 22 channels to 64) and layouts (from custom layouts to 10-20). *Main Results.* Our model achieved the highest accuracy with lower standard deviations on the testing datasets. This indicates that the GNN-based transfer learning framework can effectively aggregate knowledge from multiple datasets with different electrode layouts, leading to improved generalization in subject-independent MI EEG classification. *Significance.* The findings of this study have important implications for Brain-Computer-Interface (BCI) research, as they highlight a promising method for overcoming the limitations posed by non-unified experimental setups. By enabling the integration of diverse datasets with varying electrode layouts, our proposed approach can help advance the development and application of BMI technologies.

Keywords: Brain-Computer Interface, EEG Signal, Motor Imagery, Heterogenous Datasets, Transfer Learning, Graph Neural Network, Domain Adaptation

1. Introduction

Integrating large disparate datasets collected over years and many different sites has led to rapid progress in large deep learning models in computer vision [1] and natural language processing [2]. However, in many bio-signal and medical data processing domains, data collection is not as standardized in terms of equipment, sensor arrangements, and protocols for data collection. A prime example is Electroencephalography (EEG), which use spans a broad range of medical and technological applications from sleep analysis [3, 4] to non-invasive brain-computer interfaces (BCI). In the following, we focus on EEG decoding for non-invasive BCI [5] as a challenging example of machine learning.

In EEG-based BCI decoding, one of the most challenging areas is motor imagery (MI) decoding, where a user imagines certain actions, which trigger changes in brain signals. The computer decodes their intention from these changes in the signal in real-time [5, 6]. Existing work based on MI EEG provides a non-invasive and portable BCI solution for the users to control external devices such as rehabilitation robotics, mind-controlled wheelchairs etc., by mental execution [7–9]. However, the MI EEG methods commonly suffer from data scarcity and heterogeneity of the training datasets, which result in poor generalization performance on new test subjects. Extensive research efforts have been devoted to developing machine learning algorithms to enable automatic MI classification from EEG signals.

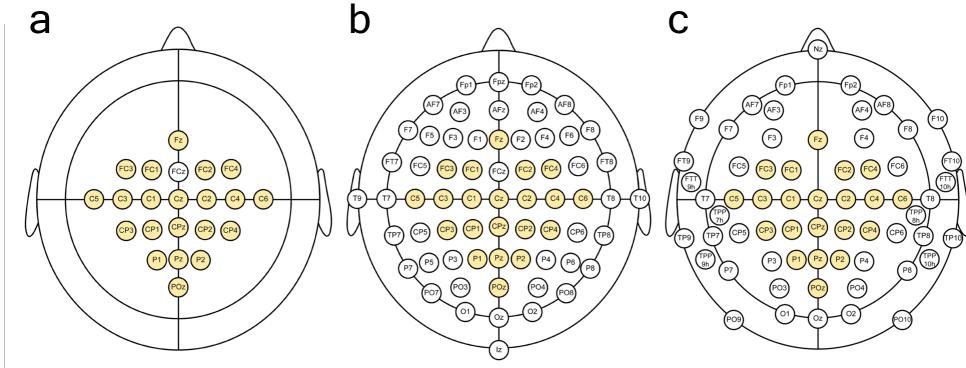


Figure 1: The 3 motor imagery EEG datasets showing different numbers of electrodes and their arrangements, visualised in the EEG 10-20 system: From left to right: BCIC 2a [10], PhysioNet MI [11], OpenBMI [12] (from left). A naive strategy to combine and utilise these datasets by only using the data recorded from these common sensors (yellow circles) discards a lot of data recorded from the other sensors (white circles).

Deep learning advances have enabled end-to-end learning without prior feature extraction from raw EEG signals in recent years. Most existing MI-EEG studies employ deep learning methods based on two kinds of basic computational models, Convolutional Neural Network (CNN) [13–17] and Recurrent Neural Network (RNN) [18, 19], that outperformed conventional Brain-Computer Interface (BCI) classifiers [20, 21]. More recently, spatial and temporal attention mechanisms [4, 22] have also been introduced to EEG deep learning models to encourage the model to focus on more

discriminative parts instead of processing the entire input equally. Feature learning requires deep learning models to use large datasets and delivers in turn the ability to learn intricate high-dimensional features, that often would have not been conceived by human-designed signal processing [23, 24].

Despite the success of deep learning methods in BCI decoding, and in most other domains of medical and biological signal analysis, there remain several issues concerning developing robust and accurate BCI decoding models using conventional end-to-end supervised training methods [25]. Deep learning methods are very data-hungry, and although efforts were made to improve their data efficiency in EEG BCI [14, 26, 27], they still require a substantial amount of data. Moreover, data-efficient methods cannot address the heterogeneity issue caused by inter-subject variation. Inter-subject variations, in motor imagery BCI systems, are caused by different mental task conditions, and motivational or psychological states. Without a very large amount of training samples, existing deep learning methods are subject to overfitting which can lead to limited generalization performance on new subjects [25, 28, 29]. However, since EEG signals are time-intensive to collect, the number of subjects and trials, even in publicly available datasets is relatively small [30]. This is especially an issue if we are considering BCI interfaces for patients, where data collection burdens end-users and may impede the long-term adoption of technology [31]. In conclusion, the use of deep learning methods in BCI decoding has been successful. Still, it is data-intensive and requires a substantial amount of data to avoid overfitting caused by inter-subject variability. However, collecting a large amount of EEG data is time-intensive and can burden end-users, especially patients. To address these challenges, better and more principled methods are needed to improve the accuracy and robustness of BCI decoding models.

One of the most promising approaches to overcome these challenges is transfer learning. Transfer learning has been used successfully in many domains, including computer vision [32], and natural language processing [33], to overcome the challenges of limited data and inter-domain variations by leveraging the knowledge learned from a related task [34]. Increasing attention has been paid to addressing the challenging inter-subject variability and data scarcity issue using transfer learning [35–37], especially domain adaptation [34]. We focus here on supervised domain adaptation as a special case of transfer learning to address the domain shift between different subjects and sensor layouts across multiple datasets. Supervised domain adaptation is a sub-field of transfer learning that aims to address the problem of training machine learning models in a source domain and adapting them to a target domain with different data distributions [34]. This approach enables the transfer of knowledge from a well-labelled source domain to a poorly labelled target domain, thereby reducing data requirements and mitigating the problems of data scarcity and insufficient labelling as the model learned in a similar domain can be manually adjusted in the target domain with only a few trials [34].

The EEG Transfer Learning benchmark BEETL [38], was initiated to evaluate the performance of current transfer learning algorithms across 5 motor imagery (MI) datasets with different numbers of channels, electrode locations and MI tasks. The top three winning teams used latent subject alignment, deep set alignment, label alignment, Maximum Classifier Discrepancy, or Riemannian geometry to achieve superior performance on the target dataset. While combining multiple datasets for training can facilitate the models' generalization performance, it is difficult to use data from multiple datasets for training since electrode layouts are often different. The

collected data suffers from significant domain shifts resulting from a variety of devices, and experiment setups, which hindered the use of existing EEG deep learning models. While selecting common electrodes between datasets might be a practical solution, it may exclude information from the unique channels of that dataset [38]. The above solutions either used 17 common channels across all three source datasets to include more subjects or just used one source dataset to keep more common channels with the target dataset. The trade-off between keeping more common electrode channels or including more subjects from more datasets limited the effectiveness of transfer learning.

Unlike image data, where pixels are arranged in a grid, neural data has sensor arrangements better described by generic graphs. Existing CNN and RNN-based deep learning methods extract spatial features of the channels by stacking the sensor readings together in a grid, without taking into account the electrode positions [39], which neglects functional neural connections within the brain which oversimplifies the EEG feature extraction process [39]. Thus, more sophisticated solutions are needed to deal more effectively with the variable sensor numbers, and their spatial relationships, which are seldom covered in the existing literature [40].

Inspired by the graph theory, Graph Neural Network (GNN) has increasingly attracted attention in deep learning research recently. It has the ability to capture complex relationships between objects and make inferences based on the graph structure data. Some studies have attempted to use GNN for EEG decoding to learn the neural connectivity of different parts of the brain [4, 41, 42]. For example, Song *et al.* [41] and Jia *et al.* [4] both proposed adjacency-matrix learning strategy to derive the intrinsic inter-channel relationships dynamically. However, their works need engineering some domain-specific features like differential entropy (DE) and power spectral density (PSD) from raw data and therefore are not end-to-end one-stage approaches. Zhong *et al.* [42] applied a regularized graph neural network architecture to learn the emotion-related functional connectivity with a node-wise domain adversarial training method to address the cross-subject variation during EEG-based emotion recognition. In addition, Demir *et al.* [39, 43] proposed EEG-GNN and EEG-GAT to classify Errp and RSVP EEG signals. The latter EEG-GAT model involves a multi-head attention mechanism to enhance the functional neural connectivity subject to specific cognitive tasks between different electrode sites. However, empirical results showed GNNs are prone to suffer from overfitting and poor convergence for small datasets, the issue of data scarcity and heterogeneity in MI EEG datasets might be more prominent [39]. Special architecture designs are needed for GNN models to adapt to EEG data from unseen subjects or datasets.

In this work, we propose a framework combining GNNs with transfer learning that can aggregate EEG data with different electrode layouts and facilitate data generalisation from the target dataset. Our framework consists of separate GNN blocks to learn the spatial connectivity of each individual EEG dataset with different electrode layouts, followed by a shared latent alignment block to handle both subject-level and dataset-level heterogeneity. Our proposed framework can effectively be trained from multiple datasets simultaneously. To compare with the experimental settings in existing GNN research and transfer learning studies for EEG signals, here we highlight the novelty of our learning paradigm. First, conventional GNN models for EEG decoding are validated under subject-dependent settings. In this study, the training and test sets are different groups of subjects thus making the evaluation of the algorithm subject-independent, which test the generality of the algorithm. Secondly,

existing works on transfer learning in the EEG domain mostly focus on inter-subject adaptation within the same database. A portion of the target subjects' EEG data is generally available for training in the experiment setting of supervised transfer learning works, and with both the data from source subjects and target subjects for training, the model can be quickly and accurately adapted. In contrast, our work focuses on utilising the data from multiple small EEG datasets with different sensor layouts and making our GNN model adapt to EEG data of fully unseen subjects in the target dataset. We present an early work that uses GNN and domain adaptation to simultaneously address the inter-subject variation and the cross-dataset channel variations.

2. Materials and Methods

To tackle the real-world challenges, we apply supervised domain adaptation to a GNN framework to aggregate information from multiple small datasets with different electrode layouts. The overall pipeline of the proposed framework is summarised in Figure 2(a). The proposed framework consists of separate encoders E_d, G_d for each dataset, a shared latent alignment layer P and a shared classifier C . Each encoder branch inputs the fixed dimensionality of the EEG signal to the particular dataset. The shared latent alignment block P uses distance metric to simultaneously address the distribution shifts across different subjects and different electrode layouts. During inference time, only the corresponding encoder branch is used to decode the data from the target dataset.

2.1. Problem Formulation

To formulate the problem in mathematical terms, consider EEG segments from D different datasets. We denote the input into our neural networks from dataset $d \in D$ as $\mathbf{X} \in \mathbb{R}^{S_d \times T}$, where S_d represent the number of sensors of EEG recordings from dataset d and T represent EEG samples with size of recording frequency times window size in seconds. The MI class output is represented as $y \in [0, 1]$, where class 0,1 represents left/right-hand imagination, respectively. An adjacency matrix that defines the graph structure and the edge weights of EEG recordings from dataset d , is represented as $\mathbf{A}_d = S_d \times S_d$, where an undirected edge between two sensors i, j is represented as $(\mathbf{A}_{ij} > 0)$. In this work, we focus on utilising the data from multiple small EEG datasets to form different graph representations and use transfer learning to make our GNN model adapt to EEG data with different sensor layouts.

2.2. Graph Neural Network Background

To apply our proposed framework, we utilize GNNs to process the EEG data represented as graphs. A GNN is a type of artificial neural network that can process data that is represented as graphs. It is widely used to learn the relationships between elements in a graph structure, such as in knowledge graphs, social networks, and chemical compounds.

Forming graph-structured data for a GNN typically involves representing the graph as a combination of two main components: the graph structure and the node features. The graph structure is typically represented using an adjacency matrix or an edge list. An adjacency matrix is a square matrix of size $S \times S$, where S is the

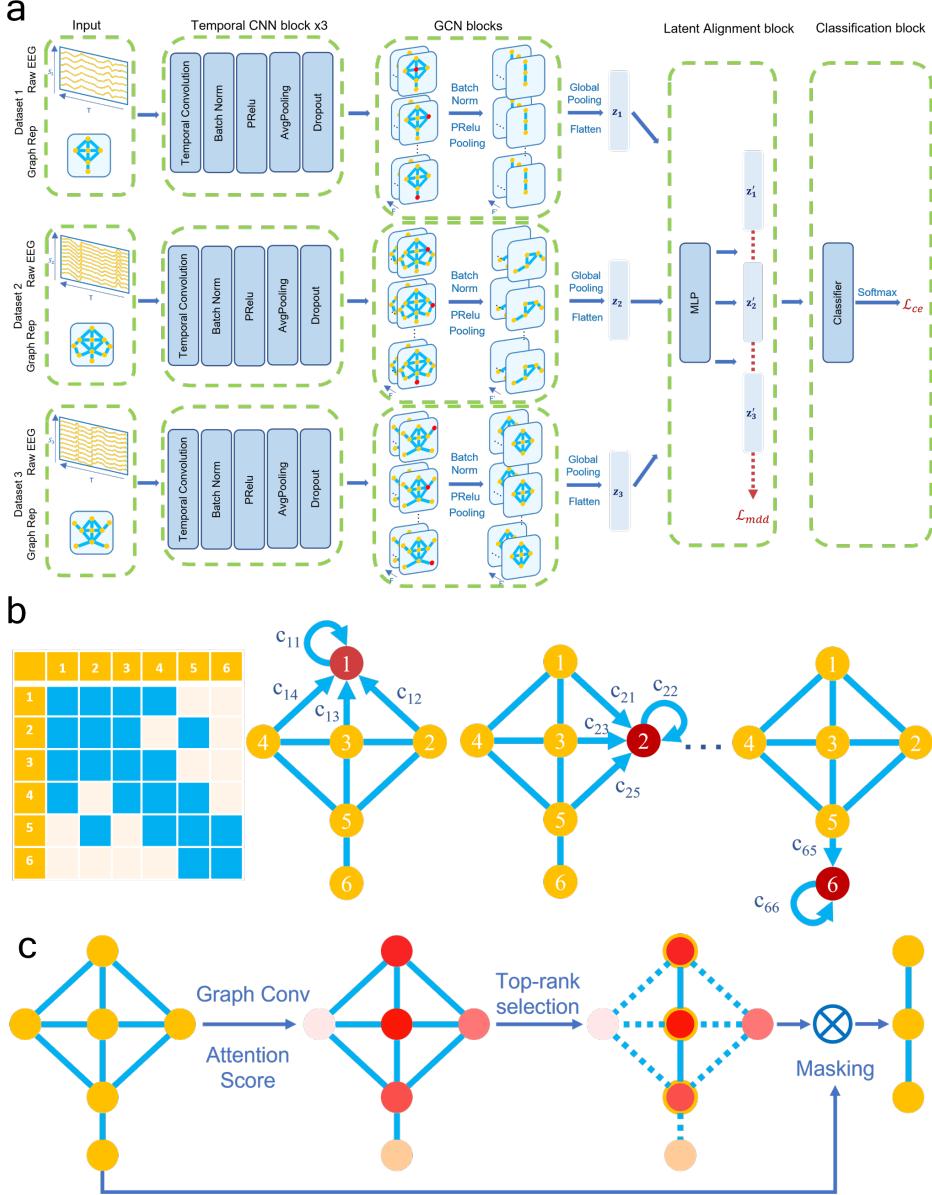


Figure 2: (a) Our proposed frameworks for training with multiple motor imagery EEG datasets. The framework starts with Temporal CNN blocks with similar architecture to EEGNet [15] without spatial filters. After each temporal convolution layer, we apply batch normalization, PReLU activation layer, average pooling and dropout layers to introduce non-linearity and reduce the risk of overfitting. After each GCN block, the number of channels is reduced by approximately a factor of 2. At the final stage of the GCN block, we apply the global mean pooling operator, which averages the remaining node features and produces a graph feature vector. The feature vector passes through an MLP latent alignment block and finally, a Linear classifier was used to perform the classification. (b) Illustration of the adjacency matrix and the local aggregation process in each GCN block. (c) Illustration of the SAGPool process in each GCN block.

number of nodes in the graph, and the entry $\mathbf{A}_{ij} > 0$ if there is an edge between nodes i and j , and 0 otherwise, as shown in Figure 2(b). An edge list is a list of tuples (i, j) representing the edges in the graph, where i and j are the indices of the nodes that are connected by the edge. The node features are typically represented as a feature matrix $\mathbf{X} \in \mathbb{R}^{S \times F}$, where S is the number of nodes in the graph and F is the number of features for each node. Each row of the feature matrix corresponds to a node in the graph and the columns represent the features associated with that node. Once the graph structure and node features are represented, they can be used as input to a GNN. A GNN uses the graph structure to propagate information across the graph and the node features to perform computations at each node. GNN is composed of several layers, each layer having its own set of parameters. The layers are connected in a graph structure, where each node in the graph corresponds to a certain parameter of the layer, and the edges between nodes represent the connections between the parameters. At each layer, the GNN takes in the data from the previous layer and applies a function to it to produce the output for the current layer. This process is called message passing or graph convolution. Graph Convolutional Neural Network (GCN) [44] is the most commonly used variant of GNN in the literature. In GCNs, the adjacency matrix \mathbf{A} is normalized by $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2} \mathbf{A} \tilde{\mathbf{D}}^{-1/2}$, where self-connections are enforced by $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$. $\tilde{\mathbf{D}}_{ij}$ is the diagonal node degree matrix of the graph, which performs a row-wise summation of the adjacency matrix $\tilde{\mathbf{D}}_{ij} = \sum_j \tilde{\mathbf{A}}_{ij}$, which is the degree of each node. The graph convolution operation allows for the incorporation of the structural information of the graph into the convolution operation. The graph convolution operation is defined as follows:

$$\mathbf{X}^{(l+1)} = \sigma(\hat{\mathbf{A}} \mathbf{X}^{(l)} \Theta^{(l)}) \quad (1)$$

Where $\hat{\mathbf{A}}$ is the symmetric normalised adjacency matrix, $\mathbf{X}^{(l)}$ is the node feature matrix at layer l , and $\Theta^{(l)}$ is the trainable weight matrix at layer l , σ is a non-linear activation function, and $\mathbf{X}^{(l+1)}$ is the output feature matrix at layer $l + 1$. From a node-wise local aggregation perspective, graph convolution is performed by aggregating information from the neighbourhood of a node, and then updating the representation of that node based on the aggregated information iteratively:

$$x_i^{(l+1)} = \sigma(\Theta^\top \sum_{k \in \mathcal{N}(i) \cup \{i\}} c_{ij} x_j^{(l)}) \quad (2)$$

Where $x_i^{(l+1)}$ is the representation of node i at layer $l + 1$, $\mathcal{N}(i)$ is the set of neighbours of node i , and c_{ij} is the weight associated with the edge between node i and node j . The process is visualised in Figure 2(b). In this way, the graph convolution operation iteratively updates the feature representation at each node, taking into account the adjacency matrix of the graph in order to propagate information across the graph and effectively learn patterns in the graph structure.

Graph coarsening and pooling layers are important for GNN models to avoid overfitting by reducing the number of parameters [45]. Self-Attention Graph Pooling (SAGPooling) [45, 46] aims to aggregate the information of a graph's nodes and reduce the graph's size for further processing. SAGPooling is a type of graph pooling which is based on self-attention mechanisms. It can be used to reduce the computation and memory requirements as well as improve the performance of GNNs due to its ability to capture the graph's global structure, as well as the relationships between the nodes. The main idea behind SAGPooling is to use a self-attention mechanism to weigh the

importance of each node in the graph, and then pool the nodes by keeping only the most important ones. The self-attention mechanism is used to compute the attention scores for each node, which represents the importance of that node in the graph. The attention scores \mathbf{Z} are computed using a GNN layer by

$$\mathbf{Z}^{(l)} = \text{GNN}(\mathbf{H}^{(l)}, \hat{\mathbf{A}}^{(l)}) \quad (3)$$

The pooling process is then performed by keeping only the top-k nodes with the highest attention scores, $\text{idx} = \text{top}_k(Z_{\text{att}}^{(l)})$, and discarding the rest. Where idx is the index of the new set of selected nodes and top_k is the function that selects the top k nodes based on their attention scores as shown in Figure 2(c). The remaining nodes' feature matrix and adjacency matrix are updated with $\mathbf{X}^{(l+1)} = \mathbf{X}^{(l)} \odot \mathbf{Z}_{\text{idx}}$ and $\hat{\mathbf{A}}^{(l+1)} = \hat{\mathbf{A}}_{\text{idx}, \text{idx}}^{(l)}$, where \odot is the elementwise product [45].

2.3. Representing EEGs as Graphs

To process EEG data using GNN, We first need to recast the EEG sensor grid as a graph, whose node features are the sensor time-series of EEG data (see Figure 3(a)). The process of representing EEGs as graphs involves defining an adjacency matrix that captures the relationships between the different EEG sensors. This matrix is critical to the process of graph representation learning, which is a fundamental part of using GNNs to analyze EEG data. We define an adjacency matrix $\mathbf{A} \in \mathbb{R}^{S_d \times S_d}$ of the EEG sensors and is essential to graph representation learning. In this work, we used two different methods to form the adjacency matrix, namely the neighbourhood method and the correlation method.

The neighbourhood method uses the physical geometry of the EEG sensor grid. We define the graph connections using natural EEG electrode layouts as shown in Figure 3(b), where each node is connected to its direct natural neighbours [22], which are the surrounding electrodes of the node on the scalp. Prior studies has demonstrated the significance of asymmetry in brain activity between the left and right motor cortex for MI prediction [42]. To propagate information across both sides, numerous graph convolution layers are required, as each layer primarily aggregates data from a node's immediate neighbors. However, this approach can result in over-squashing issues in GNNs, where the node representations lose essential information and become indistinguishable due to excessive compression through multiple layers of non-linear transformations when aggregating messages over a long path, ultimately diminishing the network's capacity to represent the graph's intricate structure [47]. Therefore, we added additional global connections between (FC3, FC4), (C3, C4), and (CP3, CP4). These nodes are shared across all three datasets and have large numbers of neighbour nodes within the motor cortex, which capture the EEG asymmetry during motor imagery tasks. In this way, the graph formation method can simultaneously represent the local electrode connectivity and MI-related global functional connectivity of different electrode layouts. The adjacency matrix forms a spatial-temporal graph with the EEG data \mathbf{X} where each node feature is the EEG sensor recording values T .

Our second method is the correlation method, which uses the functional correlation between EEG signals from different sensors to determine the neighbourhood relationships. We used the Absolute Pearson Correlation Coefficient method to capture the dynamic connectivity of different parts of the brain [48]. We initialised a fully connected graph structure for the EEG sensors, where the weight of each edge \mathbf{A}_{ij} is

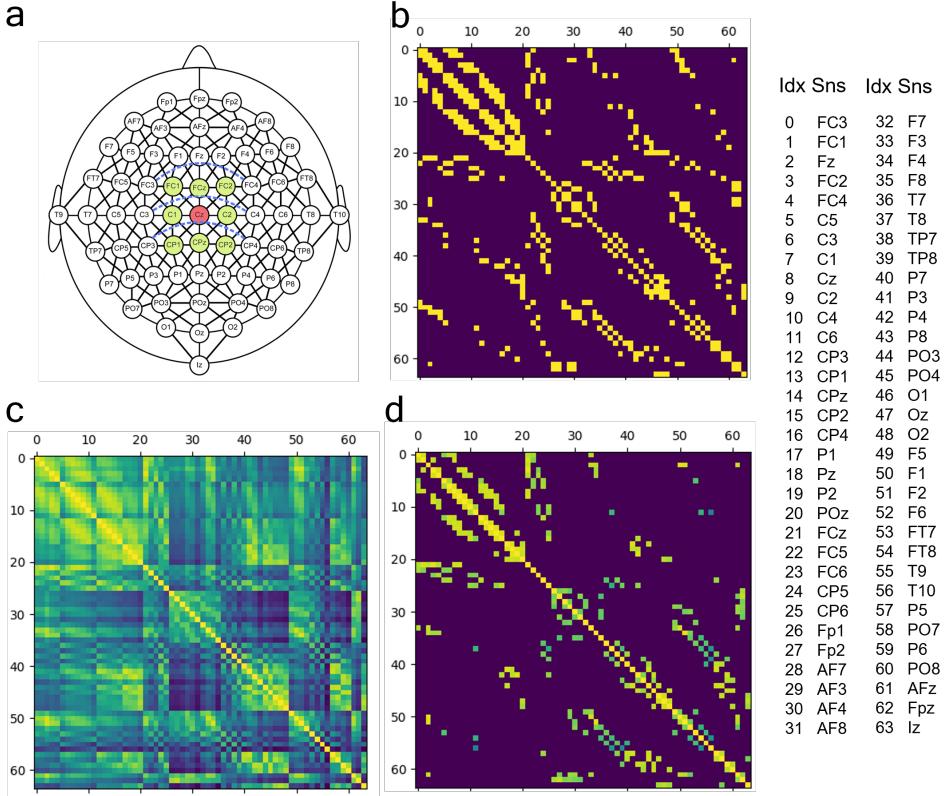


Figure 3: (a) Illustration of the PhysioNet MI sensor grid representation on the EEG 10-20 system and (b) its adjacency matrix A with shape 64×64 , constructed using the neighbourhood method. We draw here the full graph with sensors and their immediate neighbours connected by a black line. As an illustration of the neighbourhood concept, C_z (red node) is connected to its immediate neighbour nodes (green nodes). Additional global connections are shown in blue. Note, that the colour coding is binary in the adjacency matrix (right), as we only consider connections that exist (yellow) and that do not exist (dark blue). (c) The adjacency matrix constructed using the correlation method. (b) The resulting adjacency matrix when only the top 5 connections for each node are kept. The numerical index of each sensor is shown on the right.

the absolute value of the normalised cross-correlation between sensor i and sensor j . We only kept the edges with top k weight values and remove the rest to introduce sparsity and reduce computation cost. The adjacency matrix constructed by the PCC method averaged across all subjects in the PhysioNet dataset is shown in Figure 3(c), where we used the 5 largest correlation values belonging to a given node, to define the neighbourhood as shown in Figure 3(d).

2.4. Temporal/Spatial Feature Learning

We first start our decoding process by extracting the temporal features from the EEG signals. The input into our neural networks from each dataset is a tuple of data $\mathbf{X} = S \times T$ of values from S EEG sensors times T samples and second, a normalised adjacency matrix $\mathbf{A} = S \times S$. Empirical results reported from several EEG studies [39, 43, 49] suggested that GNN models are prone to overfitting as the feature dimension for each node increases. For MI EEG data, the node feature for each electrode would be the recordings in the time domain T , which usually contains hundreds of data points for each trial. Thus it is essential to perform feature extraction on the EEG data prior to the graph convolution. Although there have been studies [41] that used conventional feature extraction methods, such as the frequency band features, we decided to use CNN to extract temporal features in an end-to-end setting. We used the temporal CNN block with large kernel sizes and pooling layers on the time axis to extract temporal EEG features and dramatically lower the node features dimensions to F . Each CNN block consists of a 1-D convolution layer followed by batch normalisation, PReLU activation, average pooling and drop-out layers to prevent overfitting. Since the PReLU activation function achieves better performance than the ReLU in a shallow encoder network [50].

The node feature matrix \mathbf{X} with size $S \times F$ and the normalized adjacency matrix $\tilde{\mathbf{A}}$ are passed to the GNN block G for spatial feature learning. Each GCN layer performs graph convolution and produces the output \mathbf{X}' . After each graph convolution layer, we employ batch normalization and the SAGPooling strategy [45] to learn the top k most important nodes and drop the rest of the nodes. In the final layer of our GNN block, We used global mean pooling which averages the features across the remaining nodes and flattens the graph feature as a latent vector z . The latent vectors of each decoding branch are concatenated in the batch dimension and fed into the latent alignment block. We presented our method to learn the temporal and spatial features from EEG signals using a combination of CNNs and GNNs. By leveraging the advantages of both techniques, we were able to efficiently extract relevant features and reduce overfitting, ultimately resulting in a robust and end-to-end EEG decoding framework.

2.5. Latent Feature Alignment

Building upon the feature extraction and graph neural network framework established in the previous subsection, we now introduce the latent alignment block, which is designed to perform domain adaptation and ensure the generalizability of our model across various datasets. The latent feature alignment block consists of a shared MLP projector P that projects latent vector z of different datasets to z' and aligns their distribution in latent space. We used Maximum Density Divergence (MDD) losses [51] to simultaneously minimize the inter-dataset divergence and maximize the intra-MI class density. The MDD loss can encourage the dataset-specific encoders to extract dataset-invariant, task-related features. Given the set of latent features z_1 from 1 and z_2 from dataset 2, we calculate the pair-wise MDD loss by:

$$\begin{aligned}
\mathcal{L}_{mdd} = & \frac{1}{N} \sum_i^N \|P(z_{1,i}) - P(z_{2,i})\|_2^2 \\
& + \frac{1}{M_1} \sum_{y_{1,i}=y_{1,j}} \|P(z_{1,i}) - P(z_{1,j})\|_2^2 \\
& + \frac{1}{M_2} \sum_{y_{2,i}=y_{2,j}} \|P(z_{2,i}) - P(z_{2,j})\|_2^2
\end{aligned} \tag{4}$$

where, N is the batch size for each dataset, M_1 and M_2 are the numbers of samples with the same labels in a batch. Where the first term is the sum of the mean distance between the target and source features, the second term is the mean distance between features with the same class labels within the target dataset and the third term is the mean distance between features with the same class labels within the source dataset. The total MDD loss, \mathcal{L}_{mdd} is the sum of the pair-wise MDD loss between the feature of two datasets: $\mathcal{L}_{mdd} = \sum_{i,j}^N \text{MDD}(z'_i, z'_j) (i \neq j)$, where N is the number of different datasets and i, j represents the different index of datasets. Since there are different subjects in different datasets, the MDD loss can simultaneously minimise the inter-subject distribution shift of latent features. We finally calculate the cross-entropy loss \mathcal{L}_{ce} from the outputs of a Linear classification layer C and the supervised labels.

Overall all blocks and the classifier are updated based on $w_1 \mathcal{L}_{ce} + w_2 \mathcal{L}_{mdd}$, where w_1, w_2 are the weights for each part of the loss. At the inference stage, the raw EEG sequence passes through the target temporal CNN block, GNN block, the shared latent alignment block and the classifier to predict the imagined actions.

2.6. Experimental Settings

We picked 3 publicly available Motor Imagery (MI) datasets with various numbers of channels and electrode layouts. The size of each dataset was limited to simulate the many small-scale EEG datasets available and balance the total number of trials from each dataset.

- The BCIC IV 2a dataset [10] consists of 22 channels of EEG data from 9 different subjects, with each subject performing 144 trials of imagining left hand, right hand, feet, and tongue movements.
- The PhysioNet MI dataset [11] comprises 64-channel EEG data recorded from 109 different subjects. Each subject was asked to imagine opening and closing their left and right fist in 45 trials. We used the first four seconds of each trial and included only the first 57 subjects in our analysis.
- The OpenBMI dataset features 62-channel EEG signals recorded from 54 different subjects. Each subject was asked to imagine grasping their right or left hand for 4 seconds in 400 trials. We selected only the first 18 subjects for our study since this dataset has a significantly larger number of trials per subject

For our experiment, we included the right-hand and left-hand MI tasks from each dataset and used Inter-Subject Cross-Validation on each target dataset. We treated one of the three datasets as the target dataset while the other two as source datasets in each experiment. The data of some subjects in the target dataset is made available as supporting examples for transfer learning, while the remaining subjects' data were used for testing in an inter-subject cross-validation setting. We included different

numbers of subjects from different datasets for the experiments since the number of trials per subject is different. In each fold, we try to use a similar amount of training and testing data trials for each dataset as shown in Table 1. We used the first 4s of each trial across all the datasets, bandpass filtered between 4 – 40Hz, downsampled to 125Hz, and normalized using the z-score normalization method. We used the average Accuracy and F1 score were used as the performance metrics.

Table 1: Description of Datasets Used for Experiments.

Dataset	#Channel	Train		Test	
		#Subject	#Trial	#Subject	#Trial
BCIC 2a	22	8	2304	1	288
PhysioNet	64	51	2295	6	270
OpenBMI	62	11	2200	1	200

In order to address the trade-off between using more datasets or more channels, we consider three baseline setups that highlight the different approaches. The primary challenge arises from the varying number of channels and channel arrangements present in the three datasets, which can impact the performance of traditional deep learning models. Our baseline setups are as follows:

- Single CNN 1: A CNN trained using only the supporting examples in the target dataset, focusing on maximizing the use of available channels within a single dataset.
- Single CNN 2: A CNN trained using common channels across all three datasets, which aims to leverage the combined information from multiple datasets but at the cost of utilizing fewer channels.
- Triple CNN: A model with three separate CNN encoders, each employing different spatial convolution kernels in place of GNN encoders. This approach incorporates latent alignment using MDD before the shared classification layer, resembling the overall structure of SCSN [28]. By doing so, this baseline setup can leverage multiple datasets while neglecting the different channel arrangements.

For our proposed method, we tested our framework with the neighbourhood adjacency matrix (Proposed + \mathbf{A}_n) and with the correlation adjacency matrix (Proposed + \mathbf{A}_c).

We used CNN encoders with a similar structure to the EEGNet [15] for our baseline methods while splitting it into a temporal feature extraction block and a spatial feature extraction block same as our proposed framework. We implemented all models using open-source software *Python*, *Braindecode*, *PyTorch* and *PyTorch-Geometric*. We used the Adam optimizer (learning rate of 0.005, beta $\in [0.9, 0.999]$, weight decay of 0.0001) and batch size for each dataset of 256 and trained on a single NVIDIA RTX-3090 GPU powered Linux workstation.

3. Results

The cross-subject group level classification results are presented in Table 2 and Figure 4(a). Our method achieved the highest average performance scores compared to the 3 baseline methods. Specifically, the correlation graph method achieved 72.5%,

74.4%, and 72.6% accuracies on the BCIC 2a, PhysioNet MI, and the OpenBMI dataset, respectively. The standard deviation of the accuracy and F1 score across different folds achieved by our methods are smaller compared to other methods, which indicates that our methods are more stable and robust against the subject variations in the test dataset.

The Single CNN 1 model trained using only the target dataset performed poorly on the BCIC 2a and OpenBMI datasets with accuracies of 64.5% and 63.9%, respectively. The poor performance is likely due caused by overfitting due to its limitation on data ingestion, especially for the BCIC dataset, which only includes data from 8 subjects for training in each fold. The Single CNN 1 model performed better for the PhysioNet dataset since it includes more numbers of subjects as examples for training.

The Single CNN 2 that directly trained with data from multiple datasets by using all channels they had in common, without any alignment performed worse. It achieved 5.30% lower accuracy than the Single CNN1 setup on the BCIC dataset with only 21 common channels. This is likely due to using only the subset of common electrodes across datasets leading to the loss of information. Moreover, heterogeneity across datasets caused by different electrode layouts, and negative transfer across different subjects will have decreased the performance [28]. Thus, selecting only the common shared channels between datasets to combine them is truly limiting.

The Triple CNN model, which incorporates separate CNN blocks and a common classifier with latent alignment, significantly outperforms the Single CNN models on both the BCIC 2a (ACC: 71.6%) and OpenBMI datasets (ACC: 70.2%). This improvement in performance confirms the effectiveness of our latent alignment strategies, as the Triple CNN model is able to successfully leverage information from multiple datasets and achieve superior results compared to the traditional Single CNN methods. However, it achieved lower performance than both of our proposed GNN models due to the spatial convolution filtering without taking into account the spatial relationships between the sensors.

Table 2: Average Accuracy and F1 Score of Inter-subject Cross-Validation.

Method	Channel	Train Set	BCIC 2a		PhysioNet MI		OpenBMI	
			Accuracy	F1	Accuracy	F1	Accuracy	F1
Single CNN 1	All	Single	64.5%	62.1%	72.6%	72.0%	63.9%	61.0%
Single CNN 2	Common	All	59.2%	50.5%	68.0%	64.6%	60.2%	54.4%
Triple CNN	All	All	71.6%	68.5%	73.2%	72.5%	70.2%	60.5%
Proposed + \mathbf{A}_n	All	All	72.2%	70.2%	73.9%	71.5%	72.3%	71.5%
Proposed + \mathbf{A}_c	All	All	72.5%	70.0%	74.4%	71.6%	72.6%	72.2%

3.1. Interpreting what the GNN learned

We next investigated the internal representations of the GNN. To this end we use t-SNE to visualise the effect of our latent alignment method in the feature space. Figure 4(b) and (c) show the t-SNE projection of the MI embedding features of test data produced by the Single CNN 2 model, which does not have a latent alignment block. In Figure 4(b), the red, green and blue points represent the latent features extracted from the three different datasets respectively. The two different shapes, circle and square represent the left and right MI classes. The latent features of

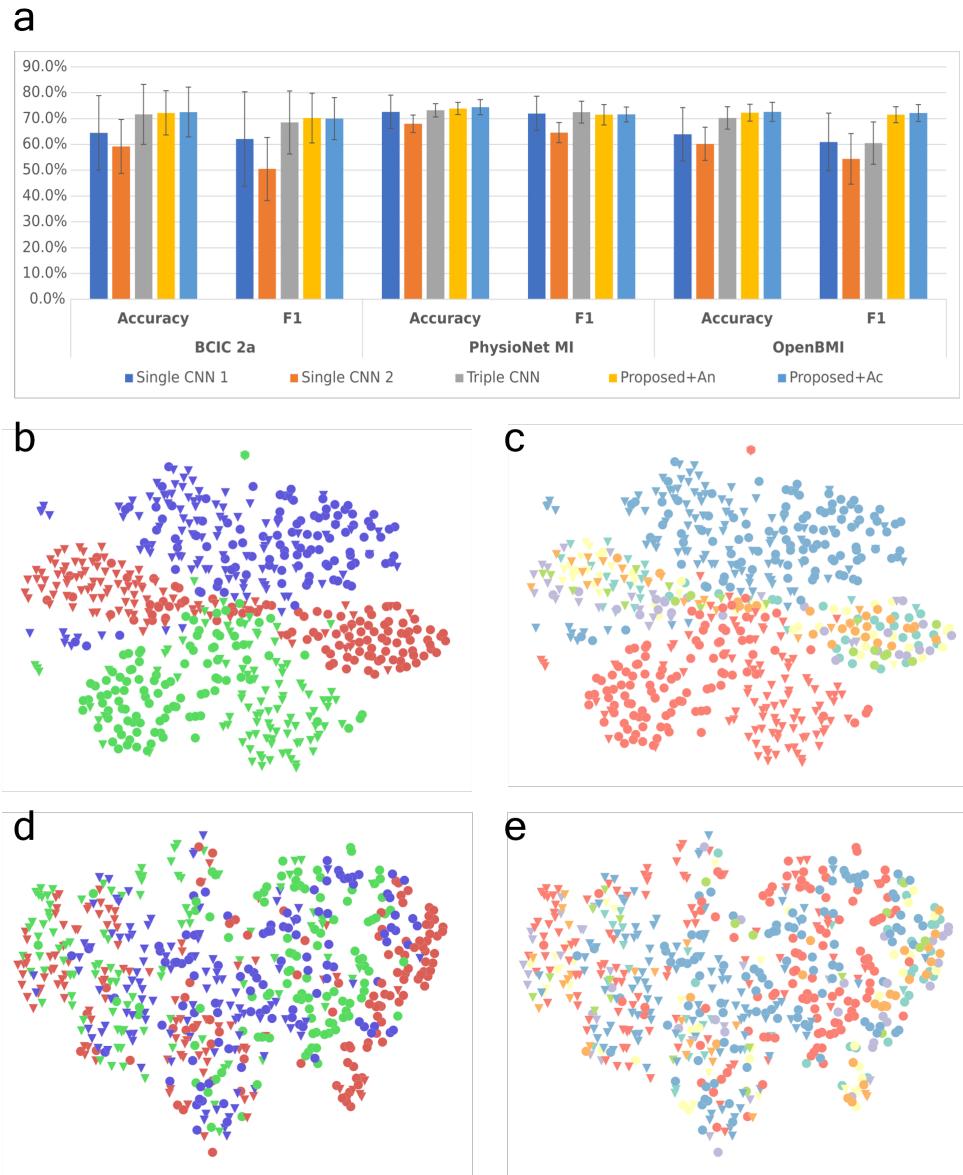


Figure 4: **(a) Average inter-subject cross-validation results.** Single CNN 1: A single CNN model trained with the training samples from a single dataset, Single CNN 2: A single CNN model trained with samples from all datasets by selecting the common channels. **tSNE plots of the latent features of test data from three EEG datasets without latent alignment:** (b) Colours represent different datasets. (c) Colours represent different subjects, Circle and Square represent left and right motor imagery classes respectively. **tSNE plots of the latent features of test data from three EEG datasets with latent alignment:** (d) Colours represent different datasets. (e) Colours represent different subjects, Circle and Square represent left and right motor imagery classes respectively.

Table 3: Average classification Accuracy and F1 score based on different model ablation designs

Ablation		BCIC 2a		PhysioNet MI		OpenBMI	
GNN	L_{md}	Acc (%)	F1	Acc (%)	F1	Acc (%)	F1
✓		66.2%	64.3%	71.9%	70.6%	67.9%	60.3%
		68.9%	69.1%	71.7%	69.8%	69.1%	62.5%
	✓	71.6%	68.5%	73.2%	72.5%	70.2%	68.5%
	✓	72.5%	70.0%	74.4%	71.6%	72.6%	72.2%

the same colour (same dataset) clustered together in the tSNE plot, which indicate that the extracted features contain a lot of dataset-specific information. Moreover, Figure 4(c) uses various colours to represent the latent features extracted from different subjects. The clustering of the same-coloured latent features (from the same subject) indicates the presence of subject-specific information in the extracted latent features. Consequently, without the latent alignment block to address domain shifts across subjects and datasets, a clear decision boundary is not apparent in the learned latent space.

In contrast, Figure 4(d) and (e) present the MI embedding features generated by our proposed method. It is evident that the majority of the latent features cluster based on their shapes (MI classes) rather than their colours (datasets/subjects). Specifically, circles cluster on the right, and triangles cluster on the left. This result demonstrates that our domain adaptation method enabled the MI encoder to effectively learn MI features from all three datasets while suppressing dataset/subject-specific features.

This look into the GNN-black box allow us to inform our interpretation why our proposed methods outperforms the 3 reference approaches.

3.2. Ablation Study

We further conducted ablation studies to validate the effectiveness of both the GNN encoder and the latent alignment module of our proposed method, with the results shown in Table 3. All ablation settings used all the channels from the three datasets. The first ablation setup used a triple CNN encoder with a shared classifier but without latent alignment. The second ablation setup added GNN for spatial feature extraction in addition to the CNN encoder. Both of these methods failed to achieve optimal inter-subject classification performance, as they lacked specific mechanisms to handle the subject variations. The third ablation setup is the same as the Triple CNN setup in the baseline experiment, which is the first ablation setup with the addition of the latent alignment mechanism. Although this setup achieved higher classification performance than the previous two configurations due to the presence of the latent alignment layer, our proposed framework with the GNN encoder and latent alignment proved to be the most effective. These ablation studies demonstrate the importance of both the GNN encoder and the latent alignment module in our proposed method, as they contribute to the optimal inter-subject classification performance across different datasets.

4. Discussion

Our proposed GNN-transfer learning framework achieved the best performance validated using three different datasets and outperforms the popular CNN methods. The results demonstrated that we can use GNNs to aggregate EEG data from multiple very different datasets with heterogeneous electrode configurations. Our GNN framework extracts spatiotemporal features of EEG data. It performs domain adaptation using MDD loss to align the feature distributions in latent space, which enforces the framework to learn features that can generalise across different subjects/datasets. This required us to translate these techniques and map them onto GNNs by first representing EEG data as graph representations using adjacency matrices, then extracting temporal features using CNNs as node features, and using GNN to learn the spatial connection of different electrode configurations.

Compared to other GNN studies in the EEG literature, most of the existing studies used handcrafted features extracted from different EEG bandwidths [4, 41], with a few other end-to-end methods that extract temporal features in the decoding process [39, 43]. However, those existing decoders were only learned on a single dataset during experiments and GNNs are found to be prone to overfitting [39, 52, 53]. In our study, we introduced transfer learning methods, specifically domain adaptation methods to circumvent the overfitting issue of GNN and trained our framework using data from multiple datasets.

Although domain adaptation has been widely used in the literature to address the inter-subject or inter-session variations in MI EEG data [54–56], it is much less explored in the literature when it comes to handling dimensionality changes in input features [40]. Specifically for the EEG data, this refers to scenarios where data was collected using different sensing modalities and spatial-temporal resolutions [57, 58]. Gao *et al.* [57] studied EEG decoding with the signals collected from multiple devices with different numbers of channels. The authors projected the EEG recordings into a common manifold space, and then used the features extracted from the manifold embeddings to perform domain adaptation. Nevertheless, this area of domain adaptation to handle dimensionality changes in EEG is much less explored than domain adaptation between homogeneous dimensionality [40]. Our work opens new ways to address the variation in input dimensionality in biosignals. Our work is one of the few studies that use GNN and domain adaptation to simultaneously address the issues of inter-subject and cross-dataset channel variations. Our work provides new solutions to enable biomedical research to integrate large disparate datasets collected over years and many different sites, which contribute to solving the common data scarcity and heterogeneity issue in medical datasets [59, 60].

5. Conclusion

This work presented an end-to-end framework with GNN and transfer learning for classifying MI EEG signals using multiple datasets with different electrode layouts. We compared the binary classification accuracy of our proposed method with three different baseline setups with different trade-offs.

Our experimental results demonstrate that the proposed network outperforms all baseline methods. This superior performance is attributed to the effective use of additional datasets without compromising non-common channels and the employment of GNNs to better handle the spatial information in the EEG layout.

In summary, our work illustrates the power of combining GNNs and transfer learning to efficiently integrate datasets without sacrificing valuable data that may not be shared across all sources. This approach offers a new way to integrate disparate biosignal databases, paving the way for novel collaboration opportunities and more comprehensive analyses in the field of biomedical research. Moreover, our findings have broader implications in the realm of machine learning in healthcare, as the developed framework can potentially be extended to other modalities and application areas where data heterogeneity and dimensionality pose significant challenges.

Acknowledgments

We acknowledge funding from UKRI Turing AI Fellowship to AAF (EP/V025449/1).

References

- [1] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*. 2015;115:211-52.
- [2] Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets. *arXiv preprint arXiv:190605474*. 2019.
- [3] Aboalayon KAI, Faezipour M, Almuhammadi WS, Moslehpoor S. Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation. *Entropy*. 2016;18(9):272.
- [4] Jia Z, Lin Y, Wang J, Zhou R, Ning X, He Y, et al. GraphSleepNet: Adaptive Spatial-Temporal Graph Convolutional Networks for Sleep Stage Classification. In: IJCAI; 2020. p. 1324-30.
- [5] Rao RP. Brain-computer interfacing: an introduction. Cambridge University Press; 2013.
- [6] Pfurtscheller G. Mapping of event-related desynchronization and type of derivation. *Electroencephalography and clinical neurophysiology*. 1988;70(2):190-3.
- [7] Pfurtscheller G, Guger C, Müller G, Krausz G, Neuper C. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience letters*. 2000;292(3):211-4.
- [8] Carlson T, Millan JdR. Brain-controlled wheelchairs: a robotic architecture. *IEEE Robotics & Automation Magazine*. 2013;20(1):65-73.
- [9] Bensch M, Karim AA, Mellinger J, Hinterberger T, Tangermann M, Bogdan M, et al. Nessi: an EEG-controlled web browser for severely paralyzed patients. *Computational intelligence and neuroscience*. 2007;2007.
- [10] Tangermann M, Müller KR, Aertsen A, Birbaumer N, Braun C, Brunner C, et al. Review of the BCI competition IV. *Frontiers in neuroscience*. 2012:55.
- [11] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on biomedical engineering*. 2004;51(6):1034-43.
- [12] Lee MH, Kwon OY, Kim YJ, Kim HK, Lee YE, Williamson J, et al. EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *GigaScience*. 2019;8(5):giz002.
- [13] Walker I, Deisenroth M, Faisal A. Deep convolutional neural networks for brain computer interface using motor imagery. *Imperial college of science, technology and medicine department of computing*. 2015:68.
- [14] Ortega P, Colas C, Faisal AA. Compact convolutional neural networks for multi-class, personalised, closed-loop EEG-BCI. In: 2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob). IEEE; 2018. p. 136-41.
- [15] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of neural engineering*. 2018;15(5):056013.
- [16] Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping*. 2017;38(11):5391-420.
- [17] Zhang C, Kim YK, Eskandarian A. EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification. *Journal of Neural Engineering*. 2021;18(4):046014.
- [18] Wang P, Jiang A, Liu X, Shang J, Zhang L. LSTM-based EEG classification in motor imagery

- tasks. *IEEE transactions on neural systems and rehabilitation engineering.* 2018;26(11):2086-95.
- [19] Zhang D, Yao L, Zhang X, Wang S, Chen W, Boots R, et al. Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface. In: Proceedings of the aaai conference on artificial intelligence. vol. 32; 2018. .
- [20] Ang KK, Chin ZY, Zhang H, Guan C. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). IEEE; 2008. p. 2390-7.
- [21] Subasi A, Gursoy MI. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert systems with applications.* 2010;37(12):8659-66.
- [22] Zhang D, Chen K, Jian D, Yao L. Motor imagery classification via temporal attention cues of graph embedded EEG signals. *IEEE journal of biomedical and health informatics.* 2020;24(9):2570-9.
- [23] Ortega P, Faisal A. HemCNN: Deep Learning enables decoding of fNIRS cortical signals in hand grip motor tasks. In: 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE; 2021. p. 718-21.
- [24] Ortega P, Faisal AA. Deep learning multimodal fNIRS and EEG signals for bimanual grip force decoding. *Journal of Neural Engineering.* 2021;18(4):0460e6.
- [25] Chen X, Li C, Liu A, McKeown MJ, Qian R, Wang ZJ. Toward open-world electroencephalogram decoding via deep learning: a comprehensive survey. *IEEE Signal Processing Magazine.* 2022;39(2):117-34.
- [26] Ferrante A, Gavriel C, Faisal A. Data-efficient hand motor imagery decoding in EEG-BCI by using Morlet wavelets & common spatial pattern algorithms. In: 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE; 2015. p. 948-51.
- [27] Ponferrada EG, Sylaidi A, Faisal AA. Data-efficient motor imagery decoding in real-time for the cybathlon brain-computer interface race. NEUROTECHNIX; 2018.
- [28] Wei X, Ortega P, Faisal AA. Inter-subject deep transfer learning for motor imagery eeg decoding. In: 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE; 2021. p. 21-4.
- [29] Wu D, Xu Y, Lu BL. Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016. *IEEE Transactions on Cognitive and Developmental Systems.* 2020;14(1):4-19.
- [30] Ortega P, Zhao T, Faisal AA. HYGRIP: full-stack characterization of neurobehavioral signals (fNIRS, EEG, EMG, Force, and Breathing) during a bimanual grip force control task. *Frontiers in Neuroscience.* 2020;14:919.
- [31] Makin TR, de Vignemont F, Faisal AA. Neurocognitive barriers to the embodiment of technology. *Nature Biomedical Engineering.* 2017;1(1):1-3.
- [32] Yu X, Wang J, Hong QQ, Teku R, Wang SH, Zhang YD. Transfer learning for medical images analyses: A survey. *Neurocomputing.* 2022;489:230-54.
- [33] Ruder S, Peters ME, Swayamdipta S, Wolf T. Transfer learning in natural language processing. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials; 2019. p. 15-8.
- [34] Wan Z, Yang R, Huang M, Zeng N, Liu X. A review on transfer learning in EEG signal analysis. *Neurocomputing.* 2021;421:1-14.
- [35] Zhang X, Zhang X, Wu L, Li C, Chen X, Chen X. Domain Adaptation With Self-Guided Adaptive Sampling Strategy: Feature Alignment for Cross-User Myoelectric Pattern Recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* 2022;30:1374-83.
- [36] Li J, Qiu S, Du C, Wang Y, He H. Domain adaptation for EEG emotion recognition based on latent representation similarity. *IEEE Transactions on Cognitive and Developmental Systems.* 2019;12(2):344-53.
- [37] Han M, Özdenizci O, Wang Y, Koike-Akino T, Erdoğmuş D. Disentangled adversarial autoencoder for subject-invariant physiological feature extraction. *IEEE signal processing letters.* 2020;27:1565-9.
- [38] Wei X, Faisal AA, Grosse-Wentrup M, Gramfort A, Chevallier S, Jayaram V, et al. 2021 BEETL Competition: Advancing Transfer Learning for Subject Independence and Heterogenous EEG Data Sets. In: Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track. vol. 176 of Proceedings of Machine Learning Research. PMLR; 2022. p. 205-19.
- [39] Demir A, Koike-Akino T, Wang Y, Haruna M, Erdogmus D. EEG-GNN: Graph Neural Networks for Classification of Electroencephalogram (EEG) Signals. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2021. p.

- 1061-7.
- [40] Gu X, Han J, Yang GZ, Lo B. Generalizable Movement Intention Recognition with Multiple Heterogenous EEG Datasets. In: 2023 IEEE International Conference on Robotics and Automation (ICRA); 2023. .
 - [41] Song T, Zheng W, Song P, Cui Z. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*. 2018;11(3):532-41.
 - [42] Zhong P, Wang D, Miao C. EEG-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*. 2020.
 - [43] Demir A, Koike-Akino T, Wang Y, Erdoğmuş D. EEG-GAT: Graph Attention Networks for Classification of Electroencephalogram (EEG) Signals. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022. p. 30-5.
 - [44] Welling M, Kipf TN. Semi-supervised classification with graph convolutional networks. In: *J. International Conference on Learning Representations (ICLR 2017)*; 2016. .
 - [45] Lee J, Lee I, Kang J. Self-attention graph pooling. In: *International conference on machine learning*. PMLR; 2019. p. 3734-43.
 - [46] Knyazev B, Taylor GW, Amer M. Understanding attention and generalization in graph neural networks. *Advances in neural information processing systems*. 2019;32.
 - [47] Alon U, Yahav E. On the bottleneck of graph neural networks and its practical implications. *arXiv preprint arXiv:200605205*. 2020.
 - [48] Tang S, Dunnmon J, Saab KK, Zhang X, Huang Q, Dubost F, et al. Self-Supervised Graph Neural Networks for Improved Electroencephalographic Seizure Analysis. In: *International Conference on Learning Representations*; 2021. .
 - [49] Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*. 2020;32(1):4-24.
 - [50] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*; 2015. p. 1026-34.
 - [51] Li J, Chen E, Ding Z, Zhu L, Lu K, Shen HT. Maximum density divergence for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*. 2020;43(11):3918-30.
 - [52] Qiu R, Li J, Huang Z, Yin H. Rethinking the item order in session-based recommendation with graph neural networks. In: *Proceedings of the 28th ACM international conference on information and knowledge management*; 2019. p. 579-88.
 - [53] Wu S, Tang Y, Zhu Y, Wang L, Xie X, Tan T. Session-based recommendation with graph neural networks. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 33; 2019. p. 346-53.
 - [54] Hang W, Feng W, Du R, Liang S, Chen Y, Wang Q, et al. Cross-subject EEG signal recognition using deep domain adaptation network. *IEEE Access*. 2019;7:128273-82.
 - [55] Zhao H, Zheng Q, Ma K, Li H, Zheng Y. Deep representation-based domain adaptation for nonstationary EEG classification. *IEEE Transactions on Neural Networks and Learning Systems*. 2020;32(2):535-45.
 - [56] Tang X, Zhang X. Conditional adversarial domain adaptation neural network for motor imagery EEG decoding. *Entropy*. 2020;22(1):96.
 - [57] Gao D, Ju C, Wei X, Liu Y, Chen T, Yang Q. Hhhfl: Hierarchical heterogeneous horizontal federated learning for electroencephalography. *arXiv preprint arXiv:190905784*. 2019.
 - [58] Ju C, Gao D, Mane R, Tan B, Liu Y, Guan C. Federated transfer learning for EEG signal classification. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2020. p. 3040-5.
 - [59] Yan K, Cai J, Zheng Y, Harrison AP, Jin D, Tang Y, et al. Learning from multiple datasets with heterogeneous and partial labels for universal lesion detection in CT. *IEEE Transactions on Medical Imaging*. 2020;40(10):2759-70.
 - [60] Yue L, Tian D, Chen W, Han X, Yin M. Deep learning for heterogeneous medical data analysis. *World Wide Web*. 2020;23:2715-37.

ArEEG_Chars: Dataset for Envisioned Speech Recognition using EEG for Arabic Characters

Hazem Darwish, Abdalrahman Al Malah, Khloud Al Jallad, Nada Ghneim.

201910306@aiu.edu.sy, 201910477@aiu.edu.sy,
k-aljallad@aiu.edu.sy, n-ghneim@aiu.edu.sy

Department of Information and Communication Engineering,
Arab International University, Daraa, Syria.

Abstract

Brain-Computer-Interface (BCI) has been a hot research topic in the last few years that could help paralyzed people in their lives. Several researches were done to classify electroencephalography (EEG) signals automatically into English characters and words. Arabic language is one of the most used languages around the world. However, to the best of our knowledge, there is no dataset for Arabic characters EEG signals. In this paper, we have created an EEG dataset for Arabic characters and named it ArEEG_Chars. Moreover, several experiments were done on ArEEG_Chars using deep learning. Best results were achieved using LSTM and reached an accuracy of 97%. ArEEG_Chars dataset will be public for researchers.

Keywords: EEG, Arabic chars EEG Dataset, Brain-computer-Interface BCI

1. Introduction

Human-Computer-Interface (HCI) is one of the most important fields in computer science that is concerned especially with the relationship between humans and computers.

Brain-Computer-Interface (BCI) is one of the HCI fields that has been a hot research topic in the last few decades. Especially researches that focus on using AI to classify EEG signals automatically where EEG signals are electrical activity in the brain measured by electroencephalography (EEG). Recent advancements in BCI show that brain signals are robust in decoding various mental tasks such as imagined speech, object understanding, etc. However, BCI in envisioned speech recognition using electroencephalogram (EEG) signals still have few researches.

Arabic language is a rich language with a long history and cultural significance. It is spoken by over 400 million people worldwide, and it is the official language in many countries. However, Arabic Language is still one of the most challenging low-resource languages.

To the best of our knowledge, there is no dataset for Arabic characters EEG signals. In this paper, we have created an EEG dataset for signals of Arabic characters and named it ArEEG_Chars. Moreover, several deep learning models were evaluated. The contributions of this paper are:

- The first dataset for EEG signals of all Arabic chars has been created.
- Baseline models to classify Arabic chars from EEG signals automatically using deep learning.

The rest of the paper is organized as follows. Section 1 is an introduction. Related works are presented in section 2. It discussed the proposed solution in section 3. Finally, we have results, discussion and conclusion in sections 4 and 5.

2. Related Works

The first paper about EEG classification was published in 1924 by Hans Berger [1] where Berger described a method for classifying EEG signals into different categories based on their frequencies. This method was later used to develop EEG-based brain-computer interface (BCI).

We investigated existing datasets and used models in detail. Some papers do not present their data collection process, nor other related information. we organized this section into “Benchmark datasets”, and “Previous models” subsections.

2.1. Benchmark Datasets

There are several datasets for different languages that were collected using multiple EEG devices. The recognized units can be digits, chars, directions, audio, images, and words. In this paper, we will focus on Chars EEG datasets. Each study has a number of participants with different ages and various collecting methods. The number of participants varied from 3 to 29, the majority of the them were adults with good health conditions. As for the data collection method, there were many different scenarios of seeing the object for a duration of seconds then imagining the object, hearing the word they imagined in their minds, or even combining these methods. In some experiments, authors mentioned that the participants have to be free from any effects on their nervous system, such as coffee, alcohol, cigarettes, and so on.

In [2] Pradeep Kumar et al. developed envisioned speech recognition using EEG sensors, they collected data by EPOC+ device on 23 participants. A dataset of EEG signals has been recorded using 30 text and non-text class objects letters, images and digits being imagined by multiple participants. As for the collection methodology, they recorded EEG signals by viewing an object on a screen then asked participant to imagine it with their eyes closed for 10 seconds. Before viewing the second object, the participants rested for 20 seconds. They collected 230 EEG samples per category for each participant. This produced 230*30 *23 samples.

Nicolás Nieto in [3] Thinking out Loud, developed an open-access EEG-based BCI dataset for inner speech recognition using EEG 16, 17, 18, 19, 20, and 21 sensors on 10 participants (6 male, 4 female), The participant first heard one of 4 directions presented through a loudspeaker (average length: $800\text{ms} \pm 20$). A first cross was then displayed on the screen (1500ms after trial onset) for 1000ms, indicating that the participant had to imagine hearing the word. Finally, a second cross was displayed on the screen (3000ms after trial onset) for 1500ms, indicating that the participant had to repeat out loud the word. The words were chosen to maximize the variability of acoustic representations, semantic categories, and the number of syllables while minimizing the variability of acoustic duration.

As for [4] researchers worked on imagined speech classification for directions (up, down, left, right) using EEG and deep learning. They used Unicorn Hybrid Black + headset on 4 participants. The EEG dataset was acquired from eight EEG sensors and contained different frequency bands with different amplitude ranges. When the recording began, the question was announced after 10s to 12s as audio caused by one of the other three participants. After 10 seconds, the participant started executing his response for 60 seconds by continuously imagining and saying the required command, and the recording was stopped after 10 seconds. In each recording, the participant responded by imagining saying the specified command, which was one of the four commands. Since we have four commands, the total recorded EEG dataset for all was 400 recordings.

Table 1 shows comparisons between EEG datasets.

Type	Study	# of Samples	# of Participants	Age	Device	Frequencies	Collecting Methods
Digits and chars	[2]	230 EEG sample for each per category for participant	23	15-40	14 Channels EMOTIV EPOC	2048Hz down-sampled 128Hz	They record EEG signals by viewing an object on a screen then as participant to imagine it with their eyes closed for 10 seconds. In rest state before viewing the second object, the participants rest for 20 seconds.
Directions	[3]	from 18 to 24 trials for each	10	31 average	16, 17, 18 19. 20, and 21 sensors	1024Hz	The participant first heard one of 4 words presented through a loudspeaker (average length: $800\text{ms} \pm 20$), then imagined hearing it again after a short delay, and finally repeated the word aloud. The researchers designed the experiment to control for the length of the words and the sounds they contained but focused on varying the meaning and number of syllables in the words.
	[4]	/	4	32	Unicorn Hybrid Black+	250Hz	When the recording began, the question was announced after 10s to 12s as audio caused by one of the other three participants. After 10s, the participant started executing his response for 60 s by continuously imagining and saying the required command, and the recording was stopped after 10 seconds. In each recording, the participant responded by imagining saying the specified command, which was one of the four commands.
Vowels	[5]	30 randomly selected epochs	3	26-29	128 Channels	1024Hz	/
	[6]	/	5	21-24	19 Channels	/	/
Directions and Select	[7]	subject has 165 epochs	27	/	14 Channels EMOTIV EPOC	128Hz	/
	[8]	/	27	/	14 Channels EMOTIV EPOC	128Hz	/

Table 1 Comparison Between EEG Datasets

2.2. Previous Models:

In the last few years, several researches were conducted to classify EEG signals automatically using machine learning and deep learning. In this section, we review them and make comparisons between the different approaches and results.

Figure 1 shows our brief summarization of EEG signals classification state-of-the-art.

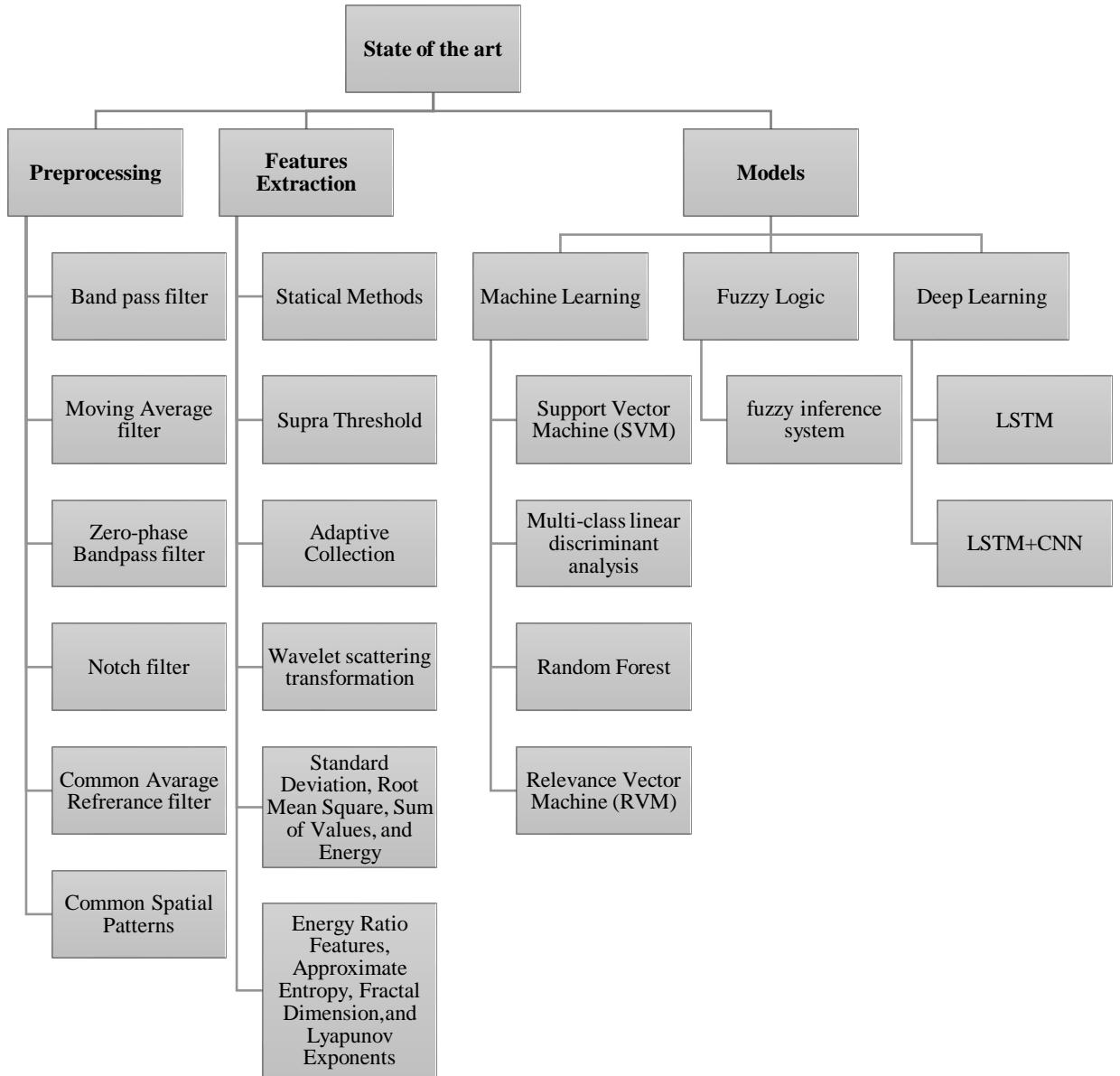


Figure 1 EEG Classification State of the art

Pradeep Kumar et al. [2] used moving average filter, standard deviation root mean square, root mean square, sum of value and ENERGY feature extractions to pre-processes raw EEG data and random forest classifier and achieved an accuracy of 82.20% on a dataset that consists of three categories images of words, letters and numbers.

Nicolás Nieto et al. [3] provided the scientific community with an open-access multiclass electroencephalography database of inner speech commands. Authors have

used zero-phase bandpass filter using the corresponding MNE [9] function 0.5 and 100 Hz with no feature extraction.

In [4] researchers used combination of Band pass Filter between 10 and 100 HZ and wavelet scattering transformation as feature extraction to preprocess the EEG data and achieved an accuracy of 92.74%, 92.50%, and 92.62% for precision, recall, and F1-score, respectively using LSTM model on dataset of 4 inners speech words about directions.

DaSalla et al [5] have proposed a BCI system to recognize English vowels using EEG signals of three subjects. The study has been conducted for three classes including two vowels and no-action state. A zero-phase bandpass filter has been applied within a frequency band of 1–45 Hz to remove lower frequencies and electronic noises. Common spatial patterns (CSP) has been applied to the EEG signals to generate new time series. An average accuracy of 71% has been recorded in all three classes using support vector machine (SVM) classifier.

In [6], authors have applied CSP and relevance vector machine (RVM) with Gaussian kernel for the classification of imagined Japanese vowels. They compared three methods, support vector machine with Gaussian kernel (SVM-G), relevance vector machine with Gaussian kernel (RVM-G), and linear relevance vector machine (RVM-L), results show that using RVM-G instead of SVM-G reduced the ratio of the number of efficient vectors to the number of training data from 97% to 55%.

In [7], Torres-García et al. have proposed a methodology for channel selection and classification of imagined speech using EEG signals. The selection of channels has been performed using fuzzy inference system (FIS) whose objectives are to minimize the number of channels and the error rate. Discrete Wavelet Transform (DWT) analysis has been performed on EEG data and five features including four statistical values and the relative wavelet energy (RWE) have been extracted for classification. The SVM has been used extensively for classification of EEG signals, they reached 70.33% accuracy.

The authors in [8] provide biometric security system. The main contribution of this paper is two methods for feature extraction. First, the Common Average Reference was applied to improve the signal-to-noise ratio. The first method was based on discrete wavelet transform, and the second method was based on statistical features directly from raw signals. The proposed methods were tested on a dataset of 27 Subjects who did 33 repetitions of 5 imagined words. As for classification, random forest classifier was used to classify inner speech. The accuracy obtained was different in each test case because it depends on number of sensors of the headset, and features extraction methods ranging from 29% up to 96%.

In [10] a second-order blind identification (SOBI) algorithm has been used to remove noises and to decompose the raw EEG signals into mutually orthogonal components. Also, they have utilized the Hilbert-Huang transformation (HHT) to extract joint temporal and spectral features. Bayesian classifier based on multi-class linear

discriminant analysis (LDA) has been used for the classification purpose, where the average accuracy of 72.67% has been recorded.

In [11], the authors have proposed an imagined speech classification system for two Chinese characters using EEG signals Down sampling, band-pass filtering, and windowing. In the spatial domain Common Spatial Patterns has been applied to preprocess the EEG Signals, In the frequency domain, methods like Mel Frequency Cepstral Coefficients (MFCC), Short-Time Fourier Transform (STFT), Fast Fourier Transform (FFT), Wavelet Transform (WT), Discrete Wavelet Transform (DWT), and Continuous Wavelet Transform (CWT) are commonly used, through statistical analysis, obtaining statistical features such as Standard Deviation (SD), Root Mean Square (RMS), Mean, Variance, Sum, Maximum, Minimum, Hjorth parameters, sample Entropy, autoregressive (AR) coefficients used for feature extraction from the acquired signals and are directly fed to the SVM classifier for recognition purposes. An average accuracy of 66.87% has been recorded on EEG data of eight participants.

Table 2 shows comparisons between EEG state of the art models.

Ref	Preprocessing		Model	Dataset	Accuracy
	Filter	Feature extraction			
[2]	Moving Average	Standard Deviation, Root Mean Square, Sum of Value and ENERGY	Random Forest	Consists of 3 categories Photos of words, letters, numbers	82.20%
[3]	Zero-phase bandpass Using the corresponding MNE function 0.5 and 100 Hz,	/	/	/	/
[4]	Bandpass filter between 10 and 100 Hz	Wavelet scattering transformation	LSTM	4 Directions	92.74%, 92.50%, and 92.62% for precision, recall, and F1score
[5]	A zero-phase bandpass filter Common Spatial Patterns	/	SVM		
[6]	Special pattern filters (CSPS) filtering	Adaptive collection (AC)	RVM	Japanese vowels	79%
[7]	/	Energy ratio features, Approximate entropy, Fractal dimension, Lyapunov exponents.	FIS3×3	Directions and select	70.33%
[8]	Common average reference	The first one is based on the discrete wavelet transform and the second one using statistical values Stander	Random Forest		29% to 96%
[11]	Down sampling, band-pass filtering ,and windowing. In the spatial domain Common Spatial Patterns	In the frequency domain, methods like Mel Frequency Cepstral Coefficients, Short-Time Fourier Transform, Fast Fourier Transform, Wavelet Transform, Discrete Wavelet Transform, and Continuous Wavelet Transform, were used for feature extraction. Then, through statistical analysis, obtaining statistical features such as standard deviation (SD), root mean square (RMS), mean, variance, sum, maximum, minimum, Hjorth parameters, sample Entropy, autoregressive (AR) coefficients	SVM	Two Chinese characters	66.87%

Table 2 Comparison between EEG state of the art models

a. Our Dataset

i. Dataset Collection

EEG recordings were collected using the Emotiv EPOC X. Thirty participants have been enrolled to collect data, their ages ranged between 15 up to 62 years. All participants are educated and have been requested to remain calm during the whole process with clear thoughts. Moreover, all of them have been requested not to consume caffeine or alcohol and not to smoke before the recording process to avoid any effects of these substances on the nervous system. Collecting EEG signals took about an hour for each participant to record his thoughts about all Arabic chars. The participant set in a room with a comfortable chair and adjusted the headset on his/her head. A presentation that contained 31 Arabic letters was shown to participants, with a slide for each character. A pictorial representation of letters is shown in Figure 2. Letters slides in the presentation were randomly ordered to prevent the participant from thinking of the next letter. Each slide was shown to every participant for 10 seconds. Next, the participant was asked to envision the shown item for 10 seconds in the eyes-closed resting state. Between two successive recordings, a gap of 20 seconds has been introduced to clear the previous imaginary thoughts of the participant. To ensure that the participant imagined the letter as per the mentioned protocol, we asked the participant during the rest phase if he/she lost his/her focus while imagining the letter, this was only at the beginning, middle, and end of the experiment, the question helped us to keep the participants focused and having more responsibility about what they were Imagining.

Using this protocol, 930 EEG recordings have been collected. To analyze the EEG signals, each recording has been split into multiple signals with a time duration of 250ms, respectively. Therefore, a total of 39857 recordings of EEG signals have been analyzed in this study.



Figure 2 Letters Used in the Study

ii. Statistics about our dataset

Thirty Arabic-native-speakers contributed in our experiments. Table 3 provides information about all participants. The average age is 26.

Participant No.	Age	Gender
Par.1	16	Male
Par.2	16	Male
Par.3	23	Male
Par.4	21	Male
Par.5	22	Female
Par.6	22	Male
Par.7	24	Male
Par.8	23	Male
Par.9	24	Male
Par.10	22	Male
Par.11	28	Female
Par.12	28	Female
Par.13	24	Female
Par.14	29	Male
Par.15	24	Male
Par.16	23	Male
Par.17	37	Male
Par.18	31	Female
Par.19	40	Female
Par.20	38	Female
par.21	54	Male
Par.22	21	Female
Par.23	25	Male
Par.24	57	Male
Par.25	22	Male
Par.26	62	Male
Par.27	37	Male
Par.28	23	Male
Par.29	24	Male
Par.30	25	Female

Table 3 Participants information

The majority of participants are male (70%) with ages between 16 and 64 as shown in Table 4.

Age Range	1-19	20-39	30-39	40-49	50-59	60-69
# of Participants	2	20	4	1	2	1

Table 4 Participants Age Ranges

The majority of participants are male and the majority are high school students, As shown in Figure 3.

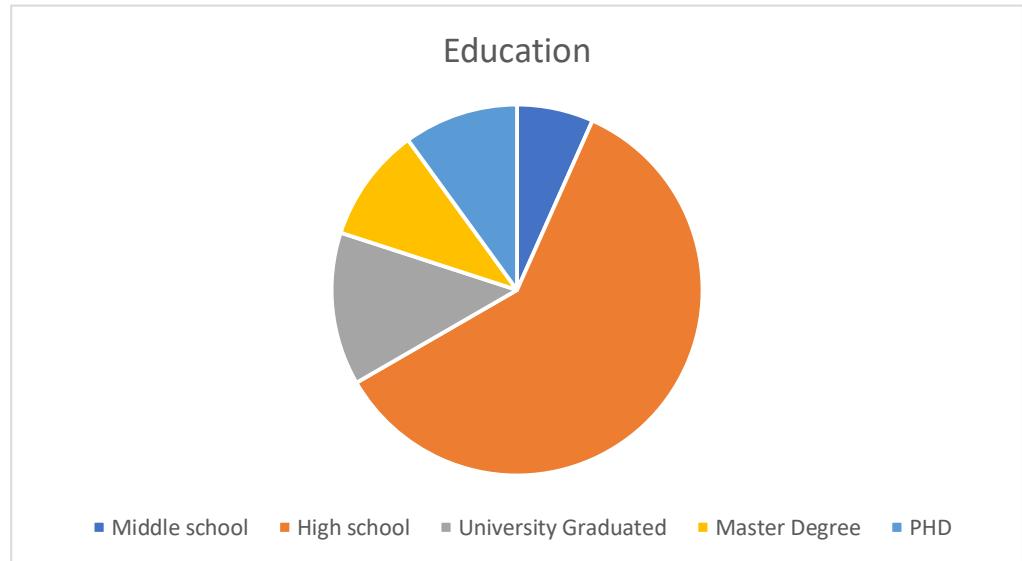


Figure 3 Participants Education Ratio

iii. Headset:

In our framework, a wireless neuro-headset known as Emotiv EPOC X has been used for the acquisition of the envisioned brain signals. For recording brain signals, this device incorporates 14 channels namely AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, which are placed over the scalp of the user according to the International 10-20 system as shown in Figure [4], where two reference electrodes, i.e., CMS and DRL, are positioned above the ears. It captures the brain waves at a frequency of 2048 Hz and later down samples them to 128 Hz per channel. The captured brain waves are sent to the computing device through Bluetooth technology.

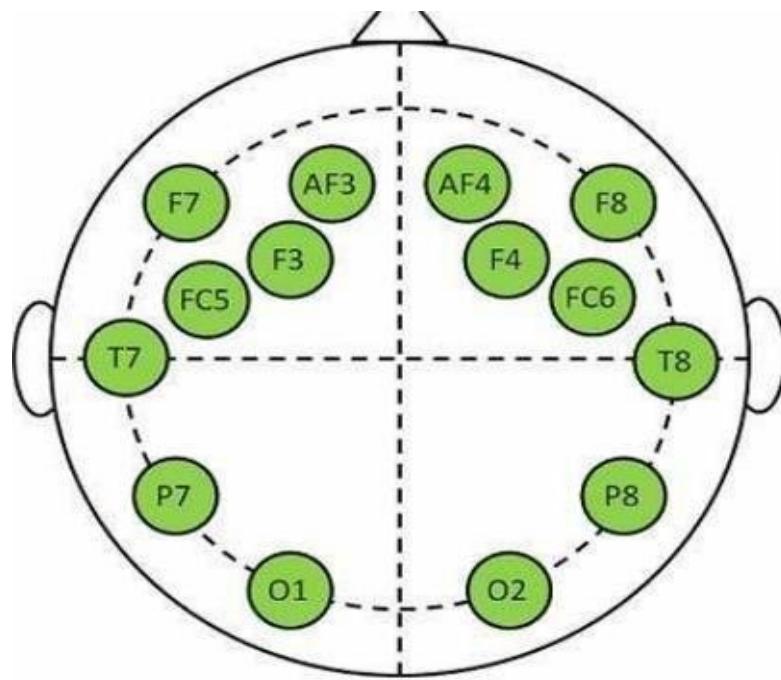


Figure 4 International 10-20 system

3. Proposed Solution

Our proposed solution, as shown in Figure 5, consists of three stages. First stage is preprocessing using moving average. Second stage is feature extraction using several algorithms to compare between their results. Third stage is deep learning model where we used CNN, LSTM, and CNN-LSTM models to compare between their results.

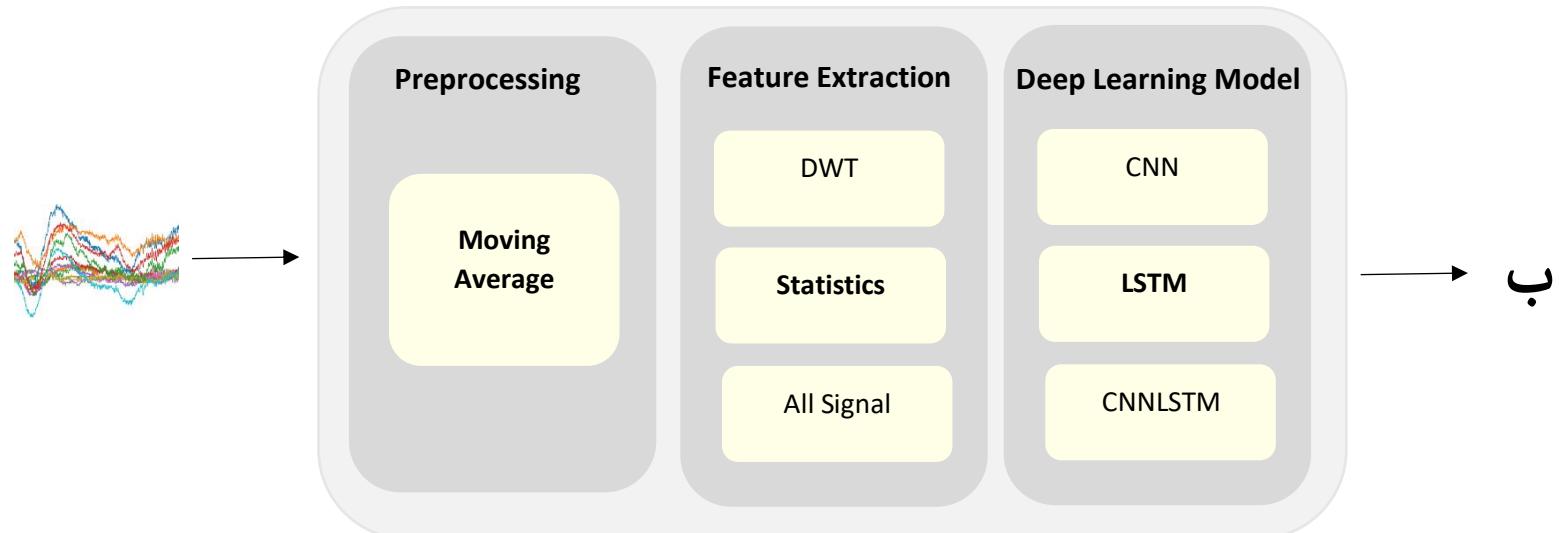


Figure 5 Our Proposed Solution

3.1. Preprocessing

The preprocessing phase is the first stage in obtaining significant characteristics from electroencephalogram (EEG) signals that are usually corrupted with various types of noises, trends, and artifacts that occur due to eye blinks, muscular activities, or various electrical noise. A preprocessing stage was done to improve the quality of signal by removing noise before extracting features, as a result improving the Arabic letter classification for EEG signals.

It was inferred that the moving average filter is widely applied and efficacious based on our exhaustive literature review and empirics. Our preprocessing strategy focused on a moving average filter which is well known to alleviate the effects of the baseline drift and noise in a high-density EEG signal since it has been recognized. Figure 6 and Figure 7, shows the effect of Filtering on the EEG signals.

Moving Average (MA) approach smooths data by averaging nearby points within a given window, lowering high-frequency noise and baseline drift. The calculation of the mean over consecutive data points in this technique enhances signal clarity [2].

$$y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i+j]$$

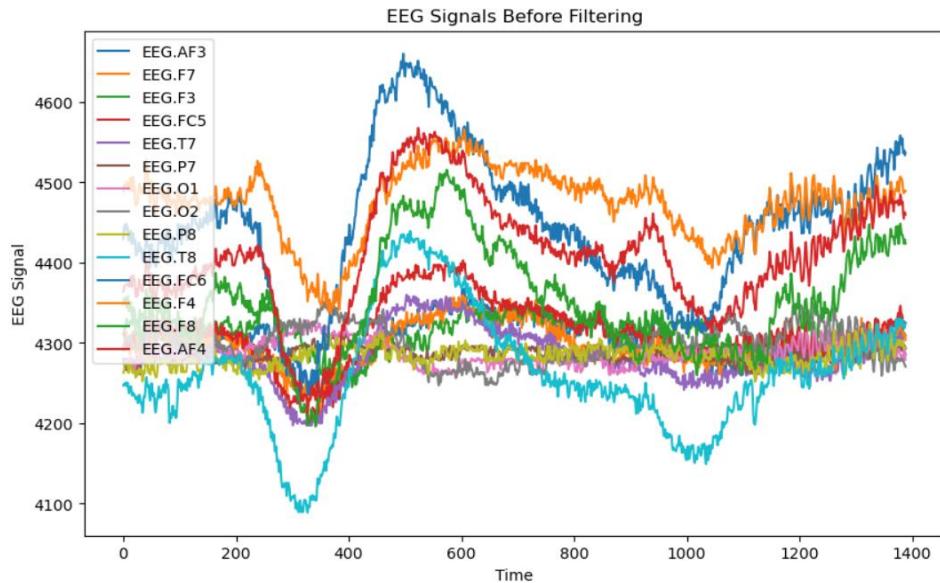


Figure 6 EEG Signals before using MA filter

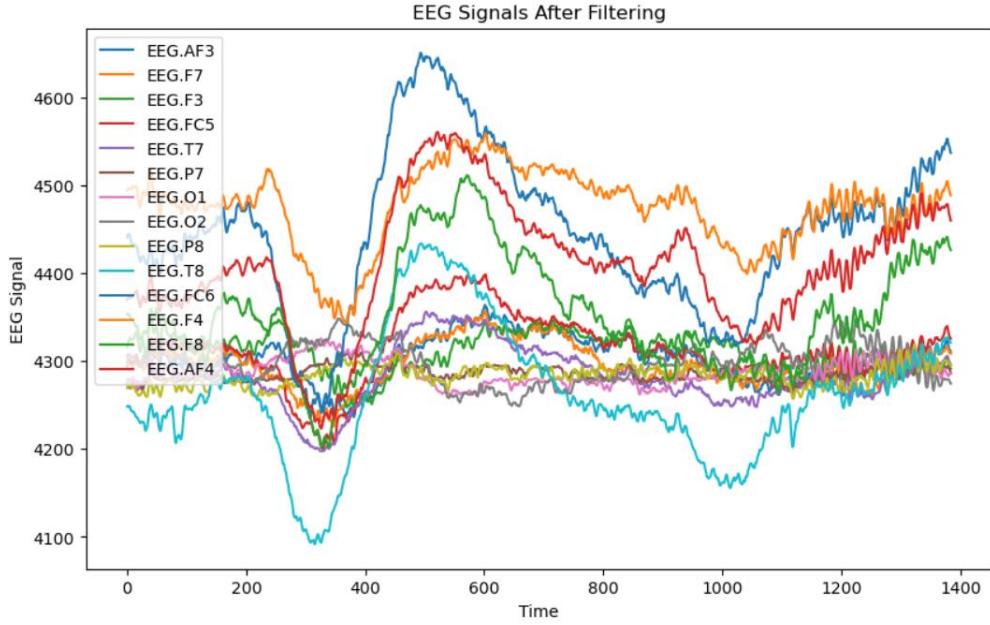


Figure 7 EEG Signals after using MA filter

3.2. Feature Extraction

Feature extraction is essential in the field of EEG analysis. Because EEG signals are inherently complex and multidimensional, they contain nuanced patterns that reflect cognitive processes. Effective feature extraction is critical in transforming these complex patterns into meaningful and distinguishing characteristics. We reveal the subtle nuances and variations within EEG signals that correlate with the representation of specific Arabic letters by isolating and quantifying these discriminative features, so we tried multiple feature extraction methods to enhance the accuracy of our model

3.2.1. Discrete Wavelet Transform (DWT)

It is based on proven approaches [8], DWT provides a multi-resolution analysis that allows time-frequency representations of EEG data to be extracted. This strategy aids in capturing both localized and global properties, therefore enhancing the discriminative potential of the retrieved features for future classification.

3.2.2. Mixture of statistical measures applied to the EEG data

A feature vector labeled as $S(F)$ with 56 dimensions was generated. Standard Deviation (SD), Root Mean Square (RMS), Sum of Values (SUM), and ENERGY (E) are all included in this feature set $S(F) = \{SD, RMS, SUM, E\}$. Previous research [2] has established these statistical characteristics as excellent markers for distinguishing various patterns within EEG data related with Arabic letter representations.

a. *Standard deviation (SD):*

It has proven to be effective in EEG signal analysis by estimating signal variation. Incorporating the raw signal (x) (\bar{x}), mean, and the signal's element count (n). [2] Standard deviation is equivalent to

$$SD = ((\frac{1}{n} - 1) \sum_{i=1}^n (x_i - \bar{x})^2)^{\frac{1}{2}}$$

b. *Root Mean Square (RMS):*

RMS is a unique feature vector assigned to each electrode and is defined as the square root of the mean of squared values within a certain interval [2] This equation simplifies its computation by expressing the sample count as 'n' and the amplitude vector as 'x' for the i^{th} input.

$$RMS = \frac{1}{n} \sum_{i=1}^n (|x_i|^2)^{\frac{1}{2}}$$

c. *Sum of Values (SUM):*

SUM [2] is calculated by adding all of the values in a signal. In this case, ' x_i ' represents the signal's i^{th} element, and 'n' indicates the signal's length.

$$SUM = \sum_{i=1}^n x_i$$

d. *ENERGY (E):*

Energy [2] is used to discrete temporal signal x , where ' x_i ' refers to the signal's i^{th} element and 'n' denotes the signal's length.

$$E = \sum_{i=1}^n x_i^2$$

3.3 Models:

i. CNN

Convolution Neural Network is one of the most popular deep learning algorithms that are used for recognizing and classifying objects [12]. CNN is used in complex problems because of its efficiency in analyzing and finding the most important elements from the input data. CNN is built using a specific architecture that have number of layers. Those layers are Convolutional layer, Pooling Layer, Activation Function (such as sigmoid, Tanh, ReLU, etc...), Connecting Layer, and Loss Function [13]. This architecture gives CNN the advantages over other networks because it lowers the number of trainable network parameters, helps in extraction of the features of the output, and can be used for large scale networks [13].

ii. LSTM

LSTM stands for Long Short-Term Memory is a type of RNN model and it is also used in tasks that have time-series and sequential data, but the key difference between LSTM and RNN is that LSTM fixes the vanishing gradient problem that is presented in traditional RNN [14], The Architecture of the LSTM consists of three main components which are an input gate, an output

gate, and a forget gate [14]. This architecture enables the network to retain relevant information over long sequences.

3.4. Experiment Setup:

In our pursuit to build an effective model for EEG-based Arabic letter recognition, we experimented with a number of models and feature extraction approaches.

The Moving Average technique has been found to be the best preprocessing method for EEG data based on our study findings [2]. The dataset was divided to training and testing sets using an 80-20 ratio split. Adam [12] optimizer was used with the specified learning rate. All models were categorical, cross entropy is chosen as the loss function, and the accuracy as evaluation metric

All the experiments in the tables [8, 9, 10] below has the same preprocessing filter but different feature extraction methods.

The highest accuracies were obtained by taking each 10 seconds record and splitting it into 250ms blocks. The 10 seconds are approximately 40 blocks of data, which means taking the whole 10 seconds to predict the letter. Therefore, if any miss imagining of a letter happens, the structure of the data after processing is cumulative and will fill this gap to prevent the failure of predicting the right letter.

As for LSTM model, the following structure was used for all LSTM experiments. Table 5 shows LSTM hyper parameters

Layer #	Type	# of units	Activation
Layer 1	Batch Normalization		
Layer 1	LSTM	512	Tanh
Layer 2	LSTM	360	Relu
Layer 3	Dropout (0.5)		
Layer 4	LSTM	120	Tanh
Layer 5	Flatten		
Layer 6	Dense	31	Softmax

Table 5 LSTM hyper parameters

As for CNN models, the following structure was used for all CNN experiments, with learning rate 0.00001. Table 6 shows CNN hyper parameters

Layer #	Type	# of units	Activation	Conv Hyper parameters
Layer 1	1D Conv.	256	Relu	kernel size= 3, stride= 1
Layer 2	Batch Normal.			
Layer 3	MaxPooling1D			size = 2
Layer 4	1D Conv.	128	Relu	kernel size= 3, stride= 1
Layer 5	Batch Normal.			
Layer 6	MaxPooling1D			size = 2
Layer 7	1D Conv.	64	Relu	kernel size= 3, stride= 1
Layer 8	Batch Normal.			
Layer 9	MaxPooling1D			size = 2
Layer 10	Flatten			
Layer 11	Dense	256	Relu	
Layer 12	Dropout (0.5)			
Layer 13	Dense	31	Softmax	

Table 6 CNN hyper parameters

As for CNN+LSTM model, the following structure was used after several experiments to determine the optimal structure with learning rate 0.0001. Table 7 shows CNN+LSTM hyper parameters

Layer #	Type	# of units	Activation	Conv Hyper parameters
Layer 1	Batch Normal.			
Layer 2	1D Conv.	128	Relu	kernel size= 10, stride= 1, padding = same
Layer 3	LSTM	256	Tanh	
Layer 4	Batch Normal.			
Layer 5	Dense	128	Relu	
Layer 6	Dropout (0.5)			
Layer 7	Flatten			
Layer 8	Dense	31	Softmax	

Table 7 CNN+LSTM hyper parameters

4. Results & Discussions:

Our preprocessing strategy focused on moving average filter which is well known to alleviate the effects of the baseline drift and noise in a high-density EEG signal, so it was used for all experiments. Then, our experiments were designed by changing the feature extraction method and the learning model.

As for CNN results, the results of using Discrete Wavelet Transform showed underfit. When using extracted statistical features (mean, maximum, minimum, standard deviation, variance, kurtosis, skewness, sum, median or standard deviation, root mean square, sum of values and energy) the accuracy was 70%, and when feeding the pre-processed input directly to deep learning model the accuracy increased to reach 88%. Table 8 shows the CNN approach results.

Experiment #	Model	Filter	Feature Extraction	Accuracy
1	CNN	Moving Average	Discrete Wavelet Transform	41.72
2			standard deviation, root mean square, sum, energy	70.11
3			—	88.33

Table 8 CNN Experiments

As for LSTM results, we noticed that the methodology of features extraction played the biggest role in enhancing the accuracy. The best results achieved using moving average filter as a preprocessing with statistical features (mean, maximum, minimum, standard deviation, variance, kurtosis, skewness, sum, median or Standard Deviation, Root Mean Square, Sum of values and energy) where the accuracy was 97%. On the other hand, using moving average filter alone or with Discrete Wavelet Transform showed underfit results. The best results between all paper experiments were achieved using moving average filter as a preprocessing with statistical features and LSTM model with an accuracy of 97%. Table 9 shows LSTM results.

Experiment #	Model	Filter	Feature Extraction	Accuracy
1	LSTM	Moving Average	Discrete Wavelet Transform	3.41
2			standard deviation, root mean square, sum, energy	97.25
3			—	27.22

Table 9 LSTM Experiments

As for CNN_LSTM approach, the Discrete Wavelet Transform also showed underfit results. The best results were achieved using moving average filter as a preprocessing then feeding it as a raw data without any feature extraction to deep learning model, or extracting statistical features (mean, maximum, minimum, standard deviation, variance, kurtosis, skewness, sum, median or standard deviation, root mean square, sum of values and energy) where the accuracy was approximately 97%. Table 10 shows CNN_LSTM results.

Experiment #	Model	Filter	Feature Extraction	Accuracy
1	CNN + LSTM	Moving Average	Discrete Wavelet Transform	6.57
2			standard deviation, root mean square, sum, energy	96.83
3			-	96.95

Table 10 CNN+LSTM Experiments

The best results in all proposed models were achieved using Moving average as a pre-processing step using LSTM model on the statistical values features (standard deviation, root, mean square, sum of value and energy).

5. Conclusion

In this paper, we have created an EEG dataset for all 31 Arabic characters and named it ArEEG_Chars that we will make it public for researchers. Moreover, several experiments were done on ArEEG_Chars using deep learning models. Best results were achieved using LSTM deep learning model and reached an accuracy of **97%**.

Our future works include enriching our dataset to include more samples, doing more experiments using different network's structure, considering shorter imagining time in order to determine which milliseconds threshold is enough to do correct predictions.

References

- [1] Lorainne Tudor Car; Mario Tudor; Katarina Ivana Tudor, "Hans berger The history of electroencephalography," *Acta medica Croatica: časopis Hrvatske akademije medicinskih znanosti* , p. 8, 2005.
- [2] Pradeep Kumar; Rajkumar Saini; Partha Pratim Roy; Pawan Kumar Sahu; Debi Prosad Dogra, "Envisioned speech recognition using EEG sensors," *Pers Ubiquit Comput*, p. 15, 2017.
- [3] Nicolás Nieto; Victoria Peterson. Hugo Leonardo Rufner; Juan Esteban Kamienkowski; Ruben Spies, "Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition," *Scientific data*, p. 17, 2022.
- [4] Mokhles M. Abdulghani; Wilbur L. Walters; Khalid H. Abed, "Imagined Speech Classification Using EEG and Deep Learning," *bioengineering*, p. 15, 2023.
- [5] C.S. DaSalla; H. Kambara; M.Sato; Y. Koike, "Single-trial classification of vowel," p. 70, 2009.
- [6] Mariko Matsumoto; Junichi Hori , "Classification of silent speech using support vector machine and relevance vector machine," 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1568494613003578>.

- [7] Alejandro A Torres-García; Carlos A. Reyes-García; Luis Villasenor-Pineda; Gregorio García-Aguilar, "Implementing a Fuzzy Inference System in a Multi-Objective EEG Channel Selection Model for Imagined Speech Classification," *Expert Systems With Applications*, p. 41, 2016.
- [8] Luis Alfredo Moctezuma; Alejandro A. Torres-García; Luis Villasenor-Pineda; Maya Carrillo-Ruiz., "Subjects Identification using EEG-recorded Imagined Speech," *Expert Systems With Applications*, p. 26, 2018.
- [9] MNE Developers, [Online]. Available: <https://mne.tools/stable/index.html#>.
- [10] Siyi Deng; Ramesh Srinivasan, "EEG classification of imagined syllable rhythm using Hilbert spectrum methods," *Journal of Neural Engineering*, p. 25, 2010.
- [11] Diego Lopez-Bernal; David Balderas; Pedro Ponce; Arturo Molina, "A State-of-the-Art Review of EEG-Based Imagined Speech Decoding," *Frontiers in Human Neuroscience*, p. 14, 2022.
- [12] Diederik P. Kingma; Jimmy Lei Ba, "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION," *a conference paper*, p. 15, 2017.
- [13] Carlos Mosquera; Greg Van Houdt; Gonzalo Nápoles, "A Review on the Long Short-Term Memory Model," *researchgate*, p. 34, 2020.
- [14] Laith Alzubaidi; Jinglan Zhang; Amjad J. Humaidi; Ayad Al-Dujaili; Ye Duan; Omran Al-Shamma; J. Santamaría; Mohammed A. Fadhel; Muthana Al-Amidie; Laith Farhan , "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, p. 74, 2021.
- [15] Prabhakar Agarwal; Sandeep Kumar, "Electroencephalography-based imagined speech recognition using deep long short-term memory network," *ETRI*, p. 14, 2022.
- [16] Taye; Mohammad Mustafa, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," *computaion*, p. 23, 2023.

Understanding data analysis aspects of TMS-EEG in clinical study: a mini review and a case study with open dataset

Hua Cheng^①

Understanding Data Analysis Aspects of TMS-EEG In Clinical Study: A Mini Review	2
ABSTRACT	2
INTRODUCTION	3
<i>TMS-evoked EEG potentials (TEPs)</i>	<i>4</i>
<i>TMS-induced EEG oscillations</i>	<i>6</i>
<i>Time-frequency approach</i>	<i>7</i>
TMS-EEG METHODOLOGY AND ARTIFACT CORRECTION	8
<i>Artifacts in TMS-EEG signals</i>	<i>8</i>
TMS-EEG DATA ANALYSIS	11
<i>Linear TMS-EEG analysis</i>	<i>12</i>
<i>Non-linear TMS-EEG analysis</i>	<i>12</i>
<i>Toolboxes for TMS-EEG data analysis</i>	<i>13</i>
TMS-EEG PROTOCOLS IN CLINICAL STUDY	14
<i>TMS protocols for assessment of cortical inhibition and excitation</i>	<i>14</i>
<i>TMS-EEG brain stimulation protocols</i>	<i>18</i>
CONCLUSION	22
Case Study of Analyzing Theta-Burst Stimulation to Frontal Cortex Based on Open Neuroscientific Data	23
ABSTRACT	23
DATA DESCRIPTION	23
<i>Resting-state EEG (RS-EEG) data</i>	<i>26</i>
<i>TMS-evoked potentials (TEPs) data</i>	<i>27</i>
<i>MRI scan data</i>	<i>28</i>
DATA PROCESS	30
<i>Independent component analysis (ICA)</i>	<i>30</i>
<i>Signal-space projection (SSP)/Principal components analysis (PCA)</i>	<i>31</i>
<i>TMS-evoked EEG potentials (TEPs)</i>	<i>33</i>
<i>TMS-induced oscillations</i>	<i>33</i>
<i>Resting state EEG power, LMFP, TRSP and ITPC</i>	<i>34</i>
STATISTICAL ANALYSIS	35
<i>Paired t tests</i>	<i>35</i>
<i>Cluster-based permutation tests</i>	<i>36</i>
CONCLUSION	37
REFERENCES	37

^① trernghwhuare@aliyun.com; trernghwhuare@hotmail.com. (extra: leoloverory@outlook.com)

Understanding Data Analysis Aspects of TMS-EEG In Clinical Study: A Mini Review

Abstract

Concurrency of transcranial magnetic stimulation with electroencephalography (TMS-EEG) technique is a powerful and challenging methodology for basic research and clinical applications. Aspects considered in experiments for effective TMS-EEG recordings and analysis, including artifact management, data analysis and interpretation and protocols. This review offers an extensive insight of TMS-EEG methodology in experimental and computational procedures.

Introduction

Transcranial magnetic stimulation (TMS) produces non-invasive brain stimulation to probe neurophysiological processes within the brain¹. TMS pulse initiated by flowing an intense current through the TMS coil windings. The current inducing an E-field is a time-varying magnetic field that penetrates the scalp and skull unimpeded. And the eddy currents induced in the brain can depolarize neurons. E-field along neurites changes rapidly at cortical neurons that have axonal bends or other geometrical inhomogeneities or endings. With short pulse duration of 1-3 T in strength and a rise time of about 50-100 μ s, TMS has temporal resolution of sub-milliseconds which allows for real-time modulation of the brain. Superficial cortical layers simulated more strongly than deeper layers as the results of magnetic fields attenuate rapidly with distance and the induced E-field approaches zero at the center of the head. But when applying adequate stimulation intensity (SI) by TMS, action potentials evoked locally may propagate along anatomical connections across cortical layers within the same cortical column and to other cortical and subcortical regions, and may result in the activation of an entire network².

Electroencephalogram (EEG) studies the electrophysiological dynamics in brain non-invasively with millisecond temporal resolution and centimeters spatial resolution *via* measuring differences in electrical potential of postsynaptic potentials synchrony rather than action potentials between electrodes placed on the scalp².

Compared to other neuroimaging techniques like fMRI, near-infrared spectroscopy (NIRS) and PET that can record TMS evoked neuro activity, EEG is the most successful and commonly used combination has been with EEG due to its inexpensive and simplicity to combine online with TMS². TMS-EEG data derived from EEG responses to TMS can be used as a neurophysiological marker of excitability or connectivity in cortex. TMS-EEG is capable of manipulating and investigating brain rhythms by measuring the impact of a TMS pulse on EEG and associated behavioral effects further investigated in the frequency domain.

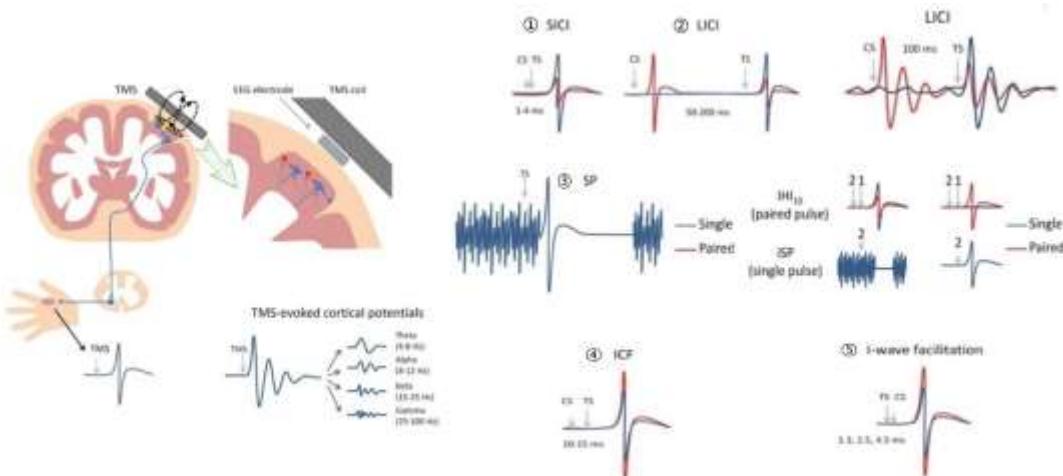


Fig. 1 TMS over cortex and TMS paradigms assessing various inhibitory and excitatory neuronal populations. ① Short-interval intracortical inhibition (SICI) involves comparing MEP amplitude of a single, suprathreshold test stimulus (TS) to a paired-pulse condition with a subthreshold conditioning stimulus (CS) and suprathreshold TS at 1-4-ms intervals. ② Long-interval cortical inhibition (LICI) involves comparing a suprathreshold with a paired-pulse suprathreshold CS and TS at 50-200-ms intervals.

intervals. ③Silent period (SP) involves measuring the duration of absent muscle activity following a single, suprathreshold TS given during a muscle contraction. ④Intracortical facilitation (ICF) involves comparing a suprathreshold TS with a paired-pulse subthreshold CS and suprathreshold TS at 10-15-ms intervals. ⑤I-wave facilitation involves comparing a suprathreshold TS with a paired-pulse condition in which a subthreshold CS follows the TS at specific intervals of 1.3, 2.5, and 4.5ms.

Adapted and modified from *N. C. Rogasch, et al., 2014*³.

TMS-evoked EEG potentials (TEPs)

TMS-evoked EEG potentials (TEPs) are the EEG responses to TMS averaged in the time domain⁴. Cortical excitatory (glutamatergic) and inhibitory (GABAergic) neurotransmitter systems activations create separate components or peaks that construct TEP. TMS of both the motor and the frontal cortices of healthy adults generally elicits a sequence of TEP components (see Fig.2-A) or positive (P) and negative (N) peaks at around milliseconds of 30 (P30), 45 (N45), 60 (P60), 100 (N100), 180(P180), and 280(N280)⁵. It is believed that peaks within the first 30 ms (P30) reflect excitatory neurotransmission⁶⁻⁸. N45 and N100 peaks (see Fig.2-B, C) are associated with GABA_A and GABA_B receptor-mediated neurotransmission, respectively⁹. Later peaks have been linked to the balance between glutamatergic excitatory and GABAergic inhibitory neurotransmission^{6,7,10,11}.

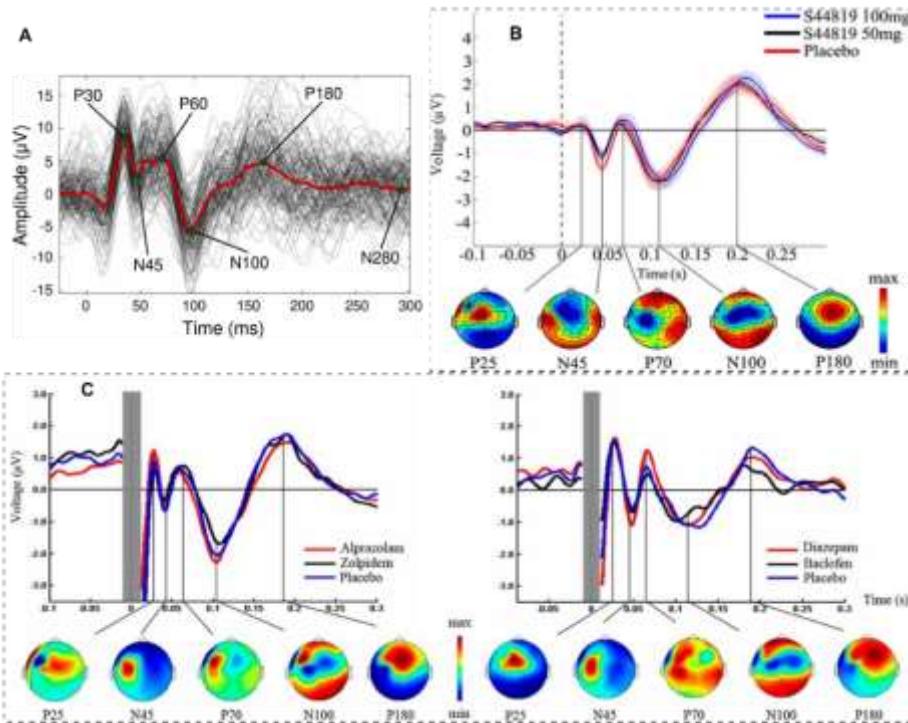


Fig. 2 TMS-evoked EEG potentials. (A)Typical TEPs(P30-N45-P60-N100-P180-N280) in a healthy young adult. Adapted from *E. Kallionniemi, et al., 2022*⁵. (B) Grand average TEPs after single-pulse TMS of M1 at baseline in the three different drug conditions of selective α 5-GABA_AR antagonist. Adapted from *G. Darmani, et al., 2016*⁹. (C) grand average TEPs after single-pulse TMS of M1 at baseline in conditions of subunit-containing GABA_ARs (left) and specific GABA_BR agonist(right).

Adapted from *I. Premoli, et al., 2014*⁷.

Many aspects of TEPs remain to be better characterized such as their morphology and physiology in different non-motor regions, test-retest reliability of peaks measured in several non-motor regions, the specific origin of each of the TEPs peaks, the exact involvement of somatosensory and auditory evoked responses in specific cortical regions and the involvement of subcortical structures in their generation.

Electrodes of interest (EOI) quantify TEPs amplitude and latency over subsets of electrodes, is one of the most common approaches. Electrodes selection in EOI analyses can be approached in priori or using data-driven methods. Data are presented as a waveform of varying amplitude as a function of time. Scalp voltage distributions visualize and quantify the spread of activity at selected time points across the cortex in ROI. EOI method is particularly relevant when there is a clear a priori hypothesis on the location of the expected brain response evoked by TMS. This method is not optimal when there is a big TMS artifact as it can mask the TEP peaks.

Local mean field power (LMFP) / cortical evoked activity (CEA) measures the area under the curve (*i.e.*, the integral) of the rectified signal or standard deviation (root mean square) across specific EOI at a given point in time corresponding to TEP peaks. LMFP/CEA is an alternative approach in measuring TEP peak amplitudes and latencies. LMFP/CEA does not present an obvious main peak and ignores the polarity of the signal, but it takes into account the width as well as the peak of the evoked activity. The LMFP method is relevant when there is an a priori hypothesis related to an expected change in brain activity that is localized and not related to a specific TEP peak (see Fig.3).

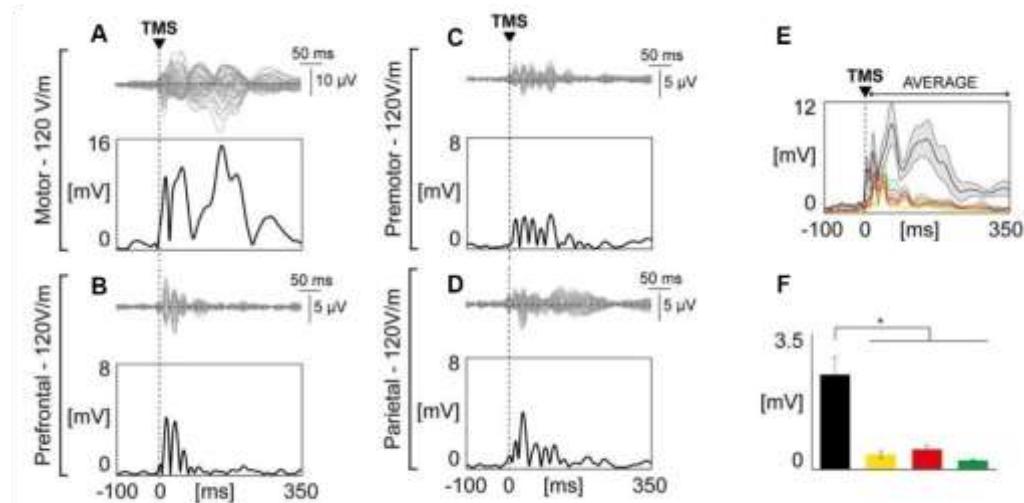


Fig. 3 different cortical areas employing LMFP. Butterfly plots of all channels with corresponding LMFP in area of (A) motor; (B) prefrontal; (C) premotor; (D) parietal. (E) Grand-average of LMFP for each stimulated area. Thick traces indicate the grand-average LMFP across subjects. Responses recorded after the stimulation of different cortical areas are color coded as follows: motor in black, prefrontal in yellow, premotor in red, parietal in green. (F) the LMFP values averaged between 8 and 350 ms post-TMS for each stimulated area. Adapted and modified from *M. Fecchio, et al., 2017*¹².

The global mean field power/amplitude (GMFP/GMFA) measures the impact of the TMS pulse on activity evoked across all electrodes, which is the averaged signal of TMS activity

over the entire surface of the head, or the standard deviation (root mean square) across electrodes at a given point in time.

$$GMFP(t) = \sqrt{\frac{[\sum_i^k (V_i(t) - V_{mean}(t))^2]}{k}} \quad (1)$$

Where t is time, k is the number of channels, V_i is the voltage in channel i averaged across subjects and V_{mean} is the mean of the voltage in all channels.

GMFP analysis is the method of choice when there is no a priori hypothesis with regards to local activity, but rather when the goal is to explore global brain activity following the TMS pulse¹³ (Fig.4).

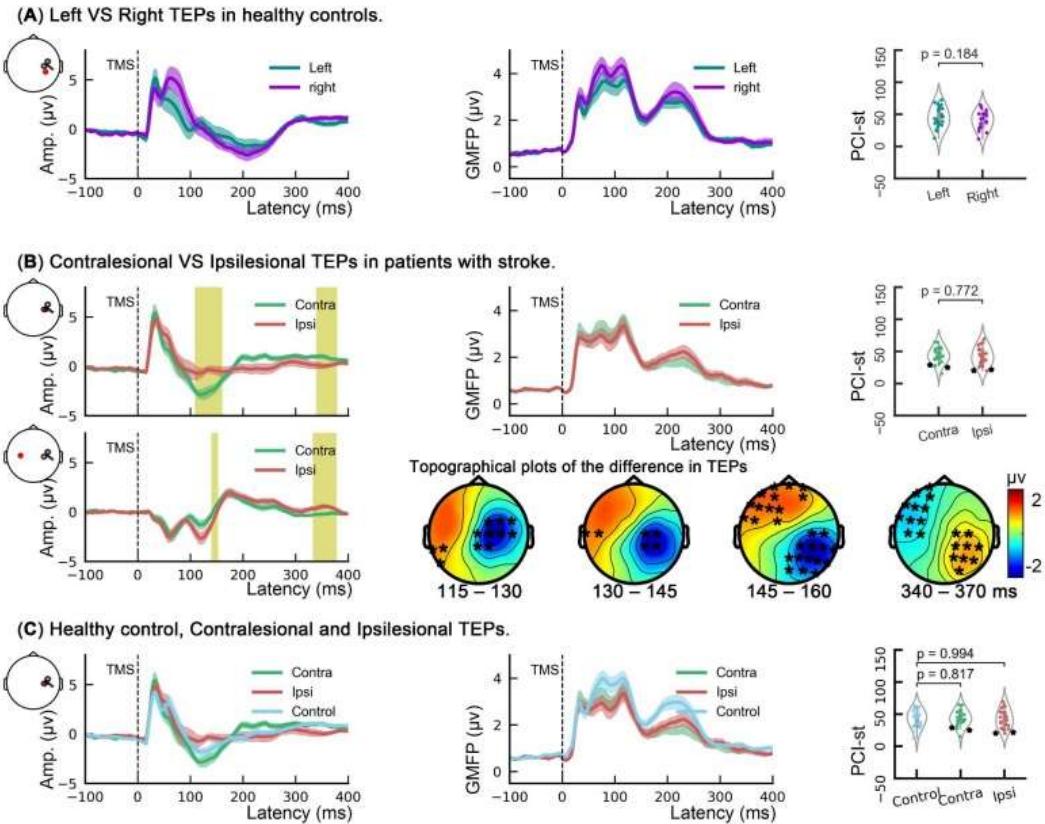


Fig. 4 TEPs Analyses. (A) TEPs comparison of the left and right in healthy controls: the TEP amplitudes of a representative channel, the GMFP of the TEPs and the comparison of left and right PCI-st. (B) TEPs comparison of the contralesional and ipsilesional in patients with stroke: the TEP amplitudes of representative channel located at the right and left hemispheres. Yellow rectangles indicate time windows in which significant differences between the ipsilesional and contralesional hemispheres were found. These differences are shown in four topographies with the black asterisk (*) represents significant clusters. And plots of The GMFP and PCI-st plots. (C) TEPs comparison in the patients with stroke and healthy controls. Adapted and modified from Z. Bai, et al., 2023¹⁴.

TMS-induced EEG oscillations

TMS induces oscillations (TIOs) specific to the brain area that can be quantified with EEG frequency domain analyses⁵. Resting-state EEG power was classified TMS-related oscillation for discrete frequency bands, *i.e.*, δ (2-4 Hz), θ (4-7 Hz), α (8-12 Hz), β (13-30 Hz) and

γ (30-45 Hz) frequency bands. Time-frequency decomposition of the TMS-EEG signal reveals TIOs' typical profile following M1 stimulation is characterized by an early increase of δ , θ , α and β band power up to 200 ms, followed by α and β suppression (often termed de-synchronization) with a final increase in β power. The occipital cortex TMS evokes α oscillations, parietal cortex TMS evokes β oscillations, and frontal cortex TMS evokes fast β/γ oscillations (see Fig. 5)¹⁵.

TMS mostly synchronizes pre-existing and ongoing oscillations instead of eliciting new neural responses¹⁶. The induced time and phase-locked oscillations create a TEP, while the non-phase-locked responses induced that average out in the TEP can be seen with specific signal-processing methods¹⁷.

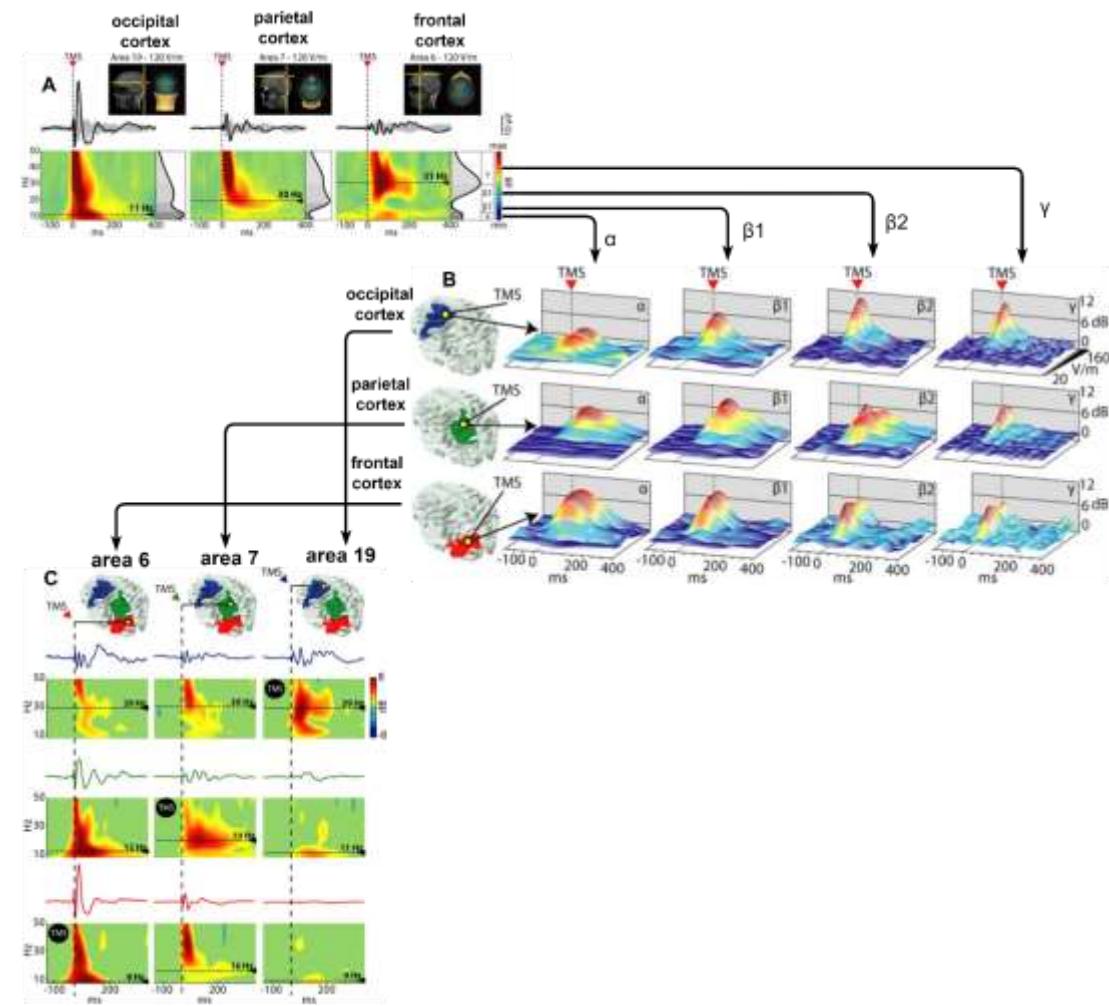


Fig. 5 TMS induces oscillations. (A) TMS elicited early γ components immediately followed by prominent α -band oscillations after occipital stimulation, β -band oscillations after parietal stimulation, and fast β/γ oscillations after perturbation of frontal cortex. (B) Different natural frequencies in different cortical areas are not attributable to different stimulation intensities. EEG frequency bands (α : 8-12; β_1 : 13-20; β_2 : 21-29; γ : 30-50) (C) The natural frequency is a local property of individual corticothalamic modules. Adapted and modified with the permissions of *M. Rosanova, et al., 2009*¹⁵.

Time-frequency approach

Time-frequency representation (TFR) approaches, such as, wavelet transforms (WT), short-time Fourier transform (STFT), extract the frequency and amplitude of cortical oscillations over time. The one focused on evoked oscillatory response (EOR), while the one accounts for the so-called induced oscillatory response (IOR) but actually better characterized by the definition of total oscillatory response (TOR) is the two approaches are generally used¹⁸:

EOR involves applying the time-frequency decomposition to the data averaged across trials (*e.g.*, the TEP) and returns information only on phase-locked oscillations following TMS (*i.e.*, evoked oscillations). **TOR** involves applying the time-frequency decomposition to individual trials, and therefore captures both the phase-locked and non-phase locked oscillations following TMS (*i.e.*, evoked and induced oscillations)¹³.

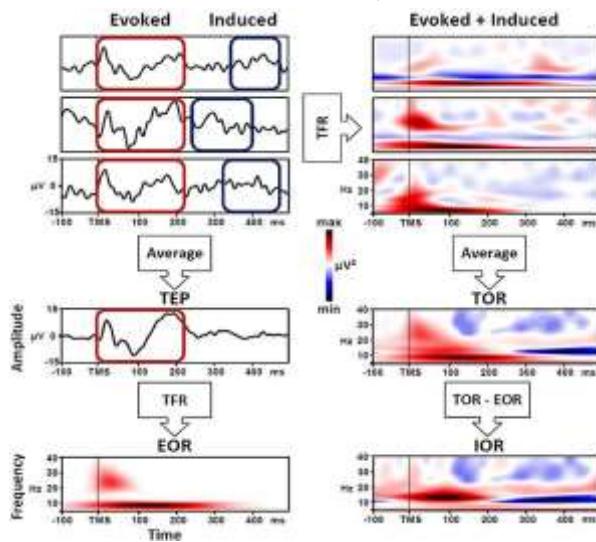


Fig. 6 Evoked oscillatory response (EOR), induced oscillatory response (IOR) and total oscillatory response (TOR) triggered by TMS pulse. **Left panel:** EOR is the time-frequency representation (TFR) of the average across all single cortical responses to TMS pulse (TEP). **Right panel:** TOR is the average of the TFR of each single response to TMS pulse, and includes EOR and IOR. To isolate pure IOR, the EOR must be removed from the TOR. Adapted from *M. C. Pellicciari, et al, 2017*¹⁸.

TMS-EEG methodology and artifact correction

A basic TMS-EEG system consists of a TMS stimulator and coil, a TMS-compatible EEG amplifier and TMS-compatible EEG electrodes⁵.

Signal-to-noise ratio (SNR) The SNR depends on the square root of the number of trials, provided that the meaningful signal and noise remain similar from trial to trial. The total SNR of averaged responses, such as TEPs, is,

$$SNR = \left(\frac{S}{N} \right) \times \sqrt{T} \quad (2)$$

Where S is the size of the signal, N is the size of the noise on a single trial, and T the number of trials. The SNR on a single trial is defined as S/N (the signal divided by the noise)².

TMS threshold determination Thresholds can be determined by measuring motor threshold (MT), phosphene threshold (PT), TEP amplitude, or induced E-field².

Artifacts in TMS-EEG signals

TMS-EEG artifacts

The major hindering the identification of TMS-EEG responses are physiological and non-physiological artefacts that interfere with the measured EEG signal. TMS coil generating electromagnetic field produces artifact in concurrent application of TMS and EEG are several orders of magnitude larger than electrophysiological activity of the brain recorded by the EEG, which resulted in saturation of EEG amplifiers. Common artifacts in TMS-EEG recordings include

- 1) common EEG artifacts. EEG artifacts arise from environmental noise (*e.g.*, power line) and physiological noise (*e.g.*, Eye blinks, cranial muscle twitch, auditory responses to the coil click, and SEPs, are all physiological but unwanted signals that can be induced by the TMS pulse).
- 2) TMS-related artifacts. The coil inevitably contacting with the electrodes, movements of EEG sensors, the pressure of the coil on the electrodes, the magnetic field applied on the electrode, the electrode-skin interface, as well as the capacitor recharge in TMS stimulators will also contribute to the production of TMS-related artifacts in the signal. TMS-induced decay artifact Offline procedures for artifact removal is a large positive shift in the signal that linearly recovers within up to 50 ms. Using TMS-compatible recordings and off-line artifacts correction, the decay artifact can recover within 10-12 ms allowing to measure early latency TMS-evoked potential. For artifacts reduction, using EEG electrodes designed for TMS applications, appropriate skin preparation to lower signal impedance under the coil for reducing direct contact with electrodes and the electrode wires re-orientation perpendicular to the stimulating coil can also help minimize the TMS-decay artifacts.

Some confounding factors secondary to the TMS pulse that should be reduced by adopting specific strategies. They are,

- 1) the TMS pulse inducing loud clicking noise (100-120 dB) can cause an auditory-evoked potential.
Wearing sound protective headphones and/or playing white noise in earphones is typically used to maximally reduce this artifact.
- 2) the TMS pulse activating sensory afferents results in a tapping sensation on the scalp that can induce a somatosensory-evoked potential.
Using a thin layer of foam under the coil may help attenuating this effect.
- 3) the TMS pulse also produce facial muscle activation and time-locked blinks.
Making sure that TMS elicits strong initial cortical responses at the stimulation site. Unlike SEPs, TEPs are specific for the stimulation parameters, like site, intensity, orientation and are characterized by even larger responses upon loss of consciousness. Most important, intracranial electrical stimulation can replicate specific changes across states in the time-frequency features and overall complexity of the responses to TMS without eliciting any sensory percept both at early and late latency. To the extent and morphology of the impacted cortex, eliciting prominent cortical responses to TMS depends on stimulation intensity, coil orientation and design. Thus, develop and apply real-time standardized data visualization tools during the experimental procedures is important to ascertain the amplitude of early TEP components. It's useful that using a sham condition to control sensory-related confounding factors in experiments where

TMS is aimed at exploring subtle changes occurring in specific, local circuits¹³.

Offline procedures for artifact removal

Blind source separation (BSS) method unmixes original source signals from their intermixed observations without prior knowledge of the mixing algorithm or source signals.

$$S = WX; S \in [m, t]; W \in [m, n]; X \in [n, t] \quad (3)$$

Where n is the number of channels and m the number of independent components; t represents the time course.

BSS removing TMS related artifacts help contributing to TMS-EEG development, includes,

- 1) Independent component analysis (ICA), which is assumed that EEG signals originate from temporally and spatially independent sources and can be modeled as a linear combination of cortical and non-cortical sources with independent time courses. But if the assumption of independence is not valid, then ICA may not separate the artifacts correctly. ICA-based artifacts correction optimization follows analysis pipeline (*Fig. 7*).

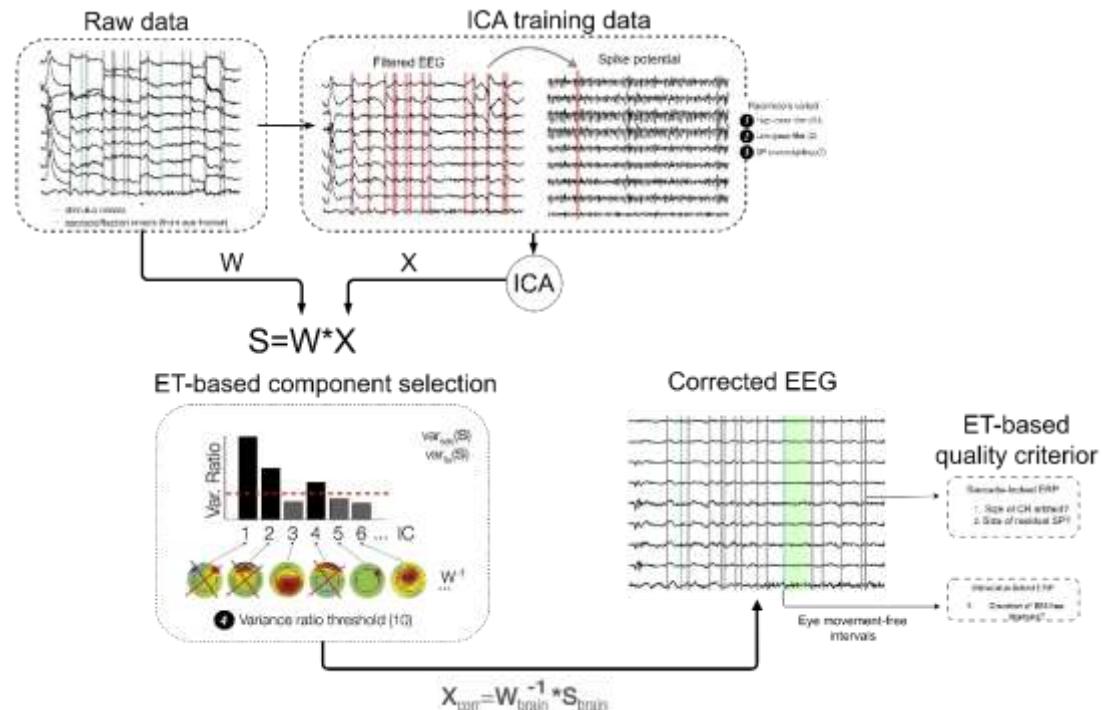


Fig. 7 Schematic workflow to determine parameters for optimized ICA training and component classification. Adapted from *O. Dimigen*, 2020¹⁹.

- 2) Principal components analysis (PCA), applied to remove eye blink artifacts in EEG signals initially, is based on a linear combination of orthogonal principal components. PCA linearly transforms a set of input data channels into an equal number of linearly-uncorrelated variables and reduce the dimensionality by orthogonal rotation, a preprocessing step of ICA.

The artifact corrected with ICA and PCA using data from all the electrodes to smooth the signals, while per electrode artifact correction may be effective when the intensity of the TMS related artifacts is locally concentrated.

TMS-EEG data analysis

When analyzing TMS-EEG data, parameters and protocols should be carefully chosen and controlled for when designing TMS-EEG experiments. Three types of parameters that can be selected in TMS-EEG experiments are parameters of input as TMS parameters, output as EEG parameters and brain state parameters (*Fig. 8*).

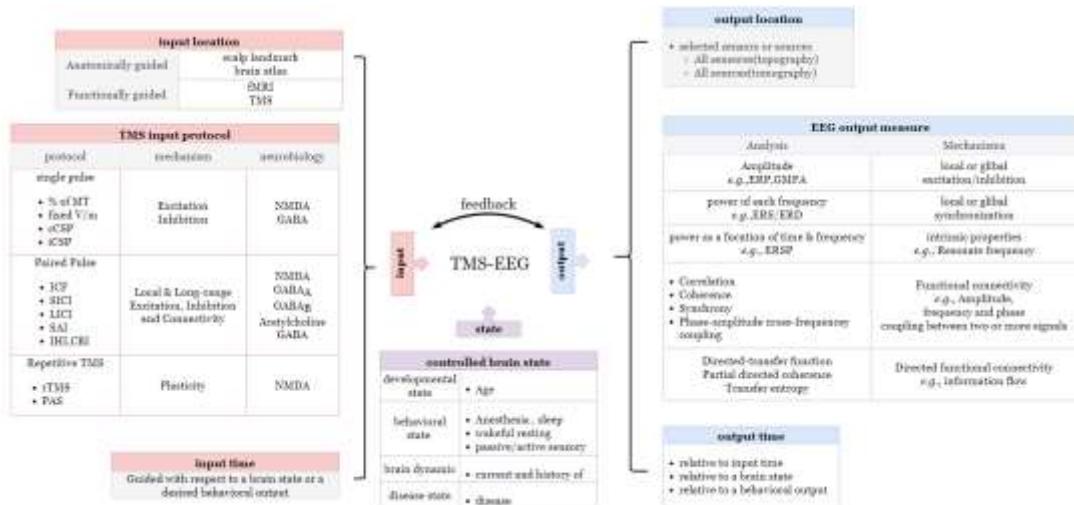


Fig. 8 A system diagram guiding the design of TMS-EEG studies. Adapted and modified from *F. Farzan, et al., 2016*²⁰.

While many steps for standard EEG preparation can be applied to TMS-EEG²¹, additional steps are required to minimize the impact of confounding factors and artifacts introduced by TMS. specific requirements for TMS-EEG preparation, *e.g.*, very low impedances (<5 kΩ), positioning of reference and ground electrodes far from the stimulation target, proper selection of the EEG amplifier settings (hardware filtering bandwidth, sampling rate, amplitude resolution).

The TMS-EEG data construction resulted in a three-dimensional tensor representing a time varying spectrum of all channels: channel (or space) × frequency × time²².

Assuming a 3D tensor W , each component comprises three matrices (A , B and C), as,

$$w_{ijk} \approx \sum_{r=1}^n a_{ir} \cdot b_{jr} \cdot c_{kr} \quad (4)$$

Where w_{ijk} is an element in the tensor W , which is approximated by the summation of N rank-1 components which are the outer product of a_r, b_r, c_r , where a_{ir} is an element in the matrix A which contains the profiles of the extracted components along the first dimension (channel or space) in its columns a_r . Likewise, B and C contains the estimated components along the second (frequency) and third (time), respectively²².

EEG analysis involves quantification of EEG signals in terms of amplitude, frequency, phase, the interaction between these attributes, the direction of information flow, and the dynamics of EEG topography, chronometry or tomography. Extracted from one or more sensors or sources, the EEG features can be described relative to the time of TMS application or change

in the brain state.

Linear TMS-EEG analysis

EEG analysis is often based on the assumption that the EEG signal represents a linear dynamical system. As the equation,

$$Y = B + A + N = LS + L_A S_A + L_N S_N \quad (5)$$

Where Y represents the EEG recorded signals, B represents the brain signals of interest, A represents the sum of the artifacts, and N represents the noises that contaminate the recorded data. When $B = LS$, B equals a product of two matrices L and S . Where L is the lead field or mixing matrix whose entry $L_{i,j}$ determines the sensitivity of channel i to the source j , and S is the source matrix whose entry $S_{j,t}$ denotes the amplitude of the source j at a time t . Meaning, the elements $A_{i,t}$ and $N_{i,t}$ of matrices A and N add artifacts and noise to the recorded signal $Y_{i,t}$. L_A , S_A , L_N , and S_N represents the artifact-mixing-, artifact-signal-, noise-mixing-, and noise-signal matrices, respectively².

When considered a linear dynamical system, the EEG signal can be decomposed into Fourier series, *i.e.*, Sine waves described by amplitude, frequency, and phase. In this model, amplitude represents the maximum vertical peak of the Sine wave (unit of μV), frequency is the number of complete cycles per second (unit of Hz), and phase describes the time point position with respect to the beginning of the Sine wave (unit of radian or degrees, ranging from -180° to 180°). To obtain the frequency and phase component, the EEG time series is multiplied by a transfer function, such as Fast Fourier Transform (FFT) or discrete wavelet transforms. In this procedure, a complex number is identified that can be used to compute the instantaneous power (proportional to the square of the maximum amplitude that the signal could reach) and phase of the signal.

Non-linear TMS-EEG analysis

The EEG signal can also be considered as a non-linear, stochastic or deterministic, and dissipative dynamical system. In a non-linear dynamical system, described by non-linear equations, a small change in initial conditions may cause a large effect. In non-linear EEG analysis, chaos theory may be applied to reconstruct an attractor from the EEG time-series. The attractor is described by its dimension, Lyapunov exponents, and entropy. The non-linear EEG analyses were employed to describe non-linear synchronization between brain regions and network nodes. EEG analyses can be grouped into general categories of reactivity and connectivity analysis.

- 1) The aim of reactivity analysis is to characterize the regional or global brain response to an event or change in brain state. In these analyses, EEG signals are often characterized by

Temporal analysis Identifying time domain features including latency and amplitude of event-related potentials (ERPs) or evoked potentials (EPs) and Global Mean Field Amplitude (GMFA).

Frequency analysis Decomposing the time domain signals into frequency sub-bands including δ ($\sim 1-3$ Hz), θ ($\sim 4-7$ Hz), α ($\sim 8-12$ Hz), β ($\sim 13-28$ Hz), and γ ($\sim >30$ Hz) oscillations, and identifying outcome measures such as evoked and induced power, relative and absolute power, or event-related synchronization (ERS) or desynchronization (ERD).

Time-frequency analysis Performing spectral decomposition using a sliding time window to calculate the change in power of each frequency as a function of time, thereby, revealing time and frequency domain information and identifying outcome measures including event-related spectral perturbation (ERSP).

Phase analysis Identifying the phase of the EEG signal at a specific time point or relative to an event.

- 2) The aim of connectivity analysis is to describe how two or more functional units, such as two or more brain regions, network nodes/hubs, or brain dynamics (*e.g.*, oscillatory activity) interact, such as function in “synchrony,” to form a larger-scale functional unit that underlies a specific brain-state. Connectivity techniques fall within two broad classes.

Non-directed connectivity analysis Measuring without quantification of the direction of information flow, including correlation, coherence or synchrony. These describe the relationship between signals recorded across the sensors (or sources), and/or across trials, by quantifying the interaction between signal attributes such as amplitude, frequency, and phase. Numerous connectivity and network dynamic metrics can be realized by quantifying the interaction between EEG features across brain regions.

Directed connectivity analysis Measures includes directed transfer function and partial directed coherence based on the Granger causality principle. For example, Directed Transfer Function (DTF) allows to determine the sources localization and the EEG activity propagation direction. defined by the Akaike information criterion (AIC) as follows,

$$AOC(d) = 2 \times \log(\det(V)) + \frac{2kd}{N} \quad (6)$$

Where V is the noise variance matrix, N is the window size, d is the model order, and k is the number of EEG channels. Model quality might fit if the conditions are satisfied: $k \times d < 0.1 \times N$ ²³.

These measures can capture the direction of information flow, rather be complex computationally, and had been applied to EEG data recently. Notably, the validity and reliability of EEG markers of functional connectivity should be examined against simulated data. Studies suggest that some connectivity analyses are confounded by the effects of volume conduction and are sensitive to the methods of temporal filtering and source reconstruction²⁰.

Toolboxes for TMS–EEG data analysis

Common EEG analysis software toolboxes used *Fieldtrip* and *EEGLAB*, combine their use with custom-written scripts on the *Matlab* platform. This complexity and the lack of a common “gold standard” analysis approach currently limit the implementation of TMS-EEG laboratories in clinical settings that do not have a strong expertise in scripting/coding. Moreover, this also contributes to the current heterogeneity in techniques employed and restricts generalization of results between studies. As such, the very recent publication of two open-source analysis approaches for TMS-EEG pre-processing, *i.e.*, *TESA* software and *TMSEEG* toolbox, as well as functionality within the *FieldTrip* toolbox, is an important step towards a standardization of TMS-EEG analysis procedures and will definitively facilitate the development of the field in the upcoming years¹³.

TMS-EEG protocols in clinical study

TMS can be used to measure various parameters in motor cortex and, allowing us to evaluate different aspects of cortical excitability. The threshold for producing an MEP in resting muscle reflects the excitability of a central core of neurons, which arises from the excitability of individual neurons and their local density. As it can be influenced by drugs that affect sodium and calcium channels, threshold must indicate membrane excitability.

TMS protocols for assessment of cortical inhibition and excitation

Various single-pulse and paired-pulse TMS-EMG techniques empower the evaluation of both inhibition and excitation in M1. These methodologies have also been integrated into TMS-EEG experiments and extended to areas beyond M1, thereby broadening the scope of research for investigating excitatory and inhibitory processes across the cortex. The principle underlining in such TMS-EEG studies is the intracortical inhibitory or excitatory processes indexed by a change in EMG amplitude related MEPs could be quantified *via* TEPs and TMS-evoked measures such as evoked cortical oscillations.

The most frequently used protocols and metrics include evaluation of motor threshold, ipsilateral cortical silent period (iCSP), contralateral cortical silent period (cCSP). Paired-pulse measures intracortical facilitation (ICF), short interval intracortical inhibition (SICI), long interval intracortical inhibition (LICI), interhemispheric inhibition (IHI), cerebellocortical inhibition (CBI), short-latency afferent inhibition (SAI).

These measures investigate the integrity of a cascade of fast- and slow-acting excitatory and inhibitory processes, occurring either within local cortical circuitry or involving long-range cortico-subcortical feedback loops²⁰.

Long-interval intracortical inhibition (LICI) and Cortical silent period (CSP)

LICI and CSP are thought to be as indices of GABA_B receptor inhibition. **LICI** is obtained when two suprathreshold stimuli are applied at intervals between 50 and 200 ms, and is thought to reflect GABA_B receptor (GABA_BR) mediated neurotransmission. Application of LICI to the motor cortex results in attenuation of δ , θ , and α oscillations, whereas β and γ oscillations were significantly inhibited in DLPFC²⁴(Fig.9).

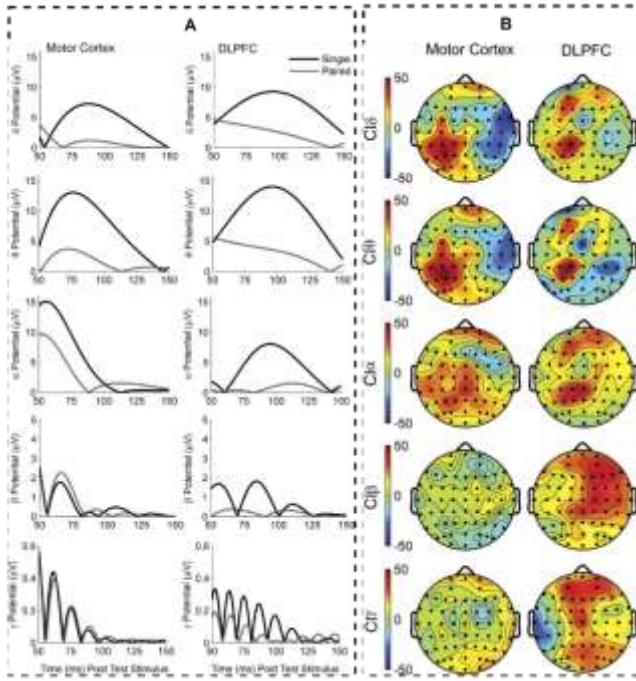


Fig. 9 Cortical inhibition (CI) measured by TMS-EEG through LICI and CSP paradigms. (A) TMS evoked cortical oscillations following the application of LICI to the motor cortex and DLPFC. (B) Topographic illustration of modulation of cortical oscillations following application of long-interval cortical inhibition (LICI) to the left motor cortex and dorsolateral prefrontal cortex (DLPFC). Adapted and modified from F. Farzan, et al., 2010²⁴.

The CSP is obtained when a TMS pulse is administered during target muscle's tonic contraction and serves as an indicator of cortical activity inhibition, most likely reflecting GABA_BR-mediated neurotransmission. GABA_Bergic-mediated inhibition in M1 was determined by the duration of CSP²⁵.

Short interval intracortical inhibition (SICI) and intracortical facilitation (ICF)

In SICI, a subthreshold conditioning stimulus (CS) inhibits a suprathreshold test stimulus (TS)-elicited MEP at inter-stimulus intervals (ISIs) of 2-3 ms while longer ISIs (7-30 ms) produce facilitation of MEPs, like ICF, for instance. SICI protocol is associated to GABA_A receptor (GABA_{AR}) activity, while the ICF produce excitation associated to both GABA_{AR} and NMDAR¹³. SICI and ICF led to inhibition and facilitation modulation of P30 and P60 TEP amplitude with TMS at M1, while P60 was bidirectionally modulated by SICI and ICF in the same manner whereas P30 was absent when DLPFC stimulation²⁶ (see Fig. 10). An increase of amplitude of N100 by the SICI paradigm, N45 amplitude increased and N100 amplitude decreased by ICF indicated age-related alterations of excitatory and inhibitory functions in the prefrontal cortex in healthy adults²⁷ (see Fig. 11). Modulation of P60 by SICI and N100 by ICF may be associated with prefrontal GABA_A and glutamatergic dysfunctions, in the expression of symptoms of schizophrenia²⁸ (see Fig. 12).

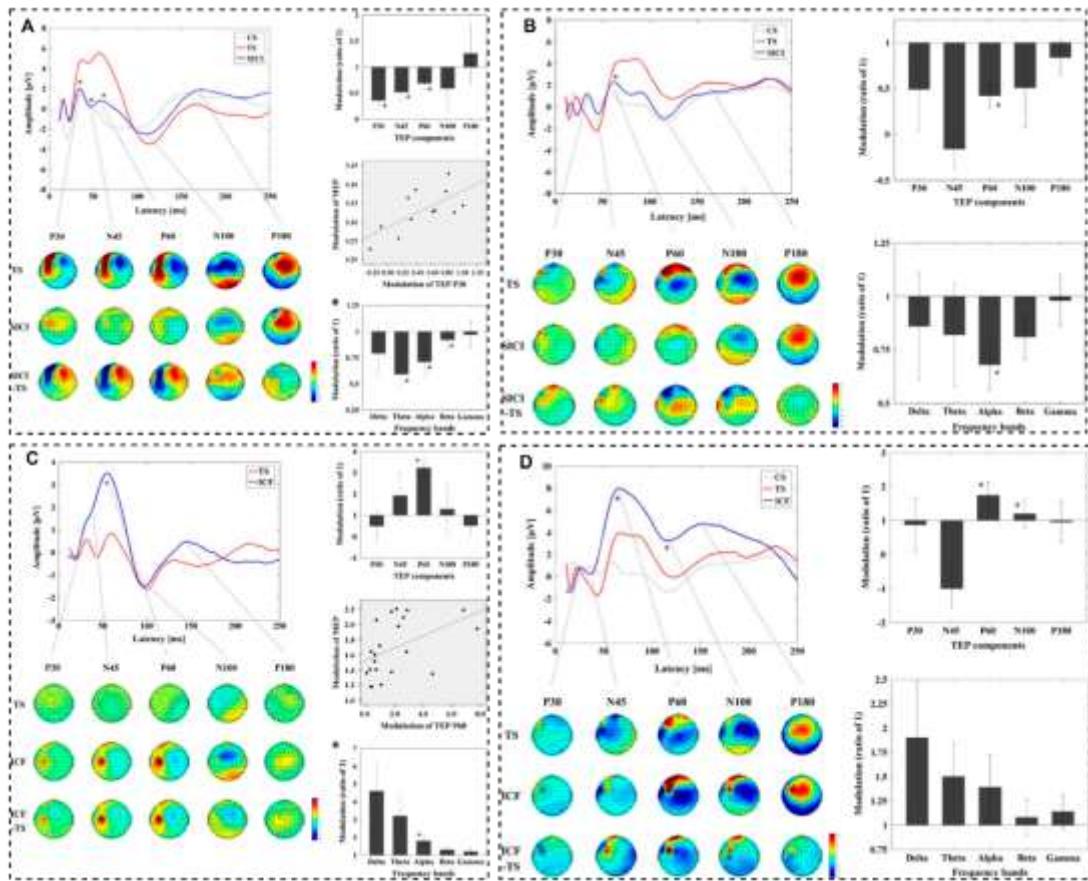


Fig. 10 Inhibitory/facilitatory influence of SICI and ICF on TEPs with TMS over M1 and DLPFC. (A) Inhibitory influence of SICI on TEPs with TMS over M1. (B) Inhibitory influence of SICI on TEPs with TMS over DLPFC. (C) Facilitatory influence of ICF on TEPs with TMS over M1. (D) Facilitatory influence of ICF on TEPs with TMS over DLPFC. Adapted from *R. F. H. Cash, et al., 2017*²⁶.

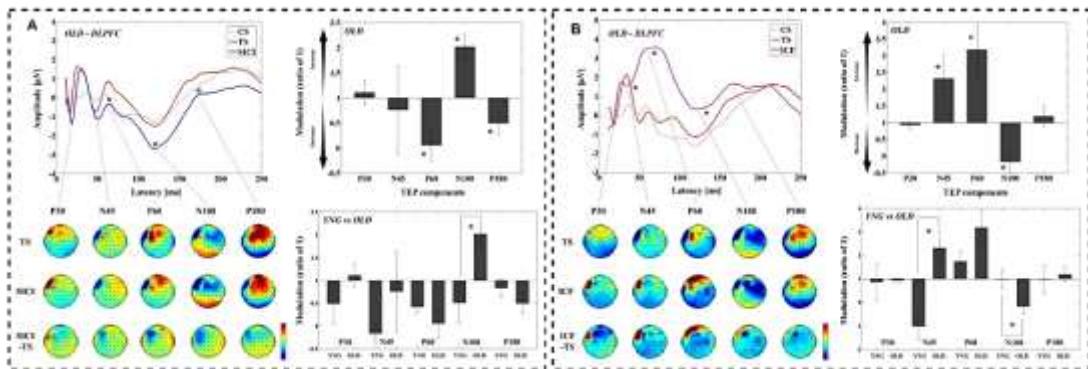


Fig. 11 Modulation of TEPs by the DLPFC-SICI & ICF paradigm following condition stimulus (CS), test stimulus (TS) in older (OLD) vs. young adults (YNG). (A) Modulation of TEPs by the DLPFC-SICI paradigm in older adults. (B) Modulation of TEPs by the DLPFC-ICF paradigm in older adults. Adapted from *Y. Noda, et al., 2017*²⁷.

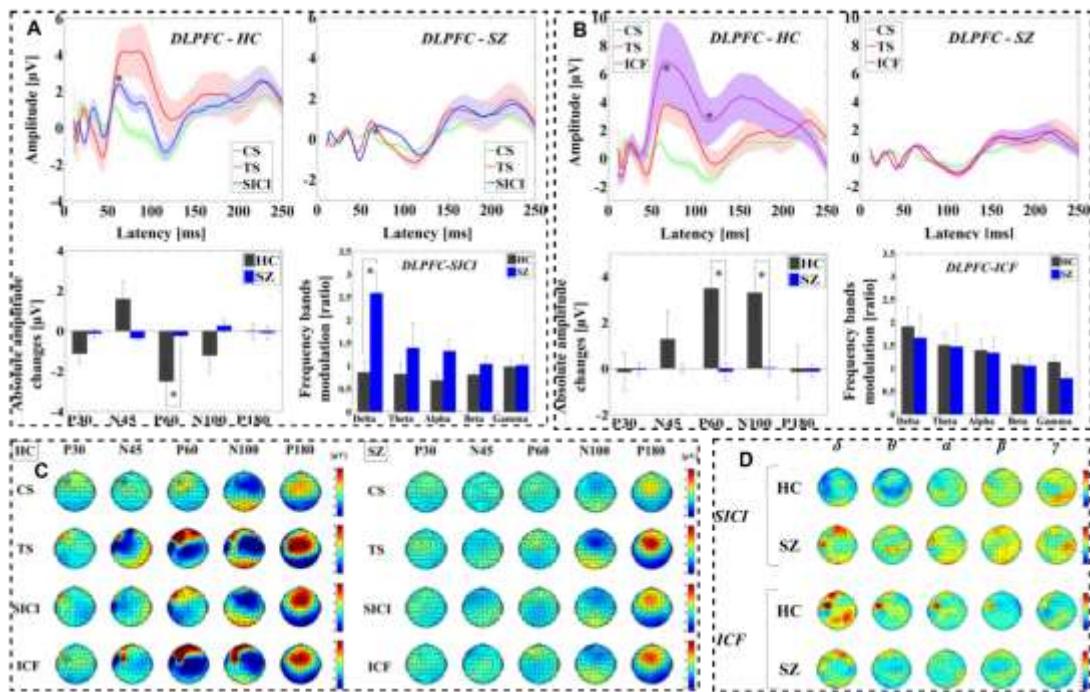


Fig. 12 SICI and ICF from the DLPFC in healthy control (HC) and schizophrenia (SC) following CS, test TS. (A) Modulation of TEPs by SICI paradigm administered TMS to DLPFC. (B) Modulation of TEPs by ICF paradigm administered TMS to DLPFC. (C) Topographical plots of paired pulse SICI and ICF paradigms. (D) Topographical distributions of frequency band modulations by SICI and ICF paradigms. Adapted from Y. Noda, et al., 2017²⁸.

Short latency afferent inhibition (SAI)

SAI is obtained when applied to M1 with the combination of median nerve electrical stimulation and TMS precedes a TS at ISIs of 20-25 m leads to MEP suppression. SAI over M1 has been predominantly associated with cholinergic and GABA_Aergic circuits. Assess correlates of SAI using TMS-EEG has indicated a decrease in the N100 component. If apply SAI protocol over M1, MEP amplitude reduce. A similar reduction of N100 amplitude accompanied by a P60 attenuation and a ERSP decrease in the beta band was found. Modulation of SAI associated with N100 component with MEP suppression, is an increase rather than a decrease was also observed (Fig.13-A).

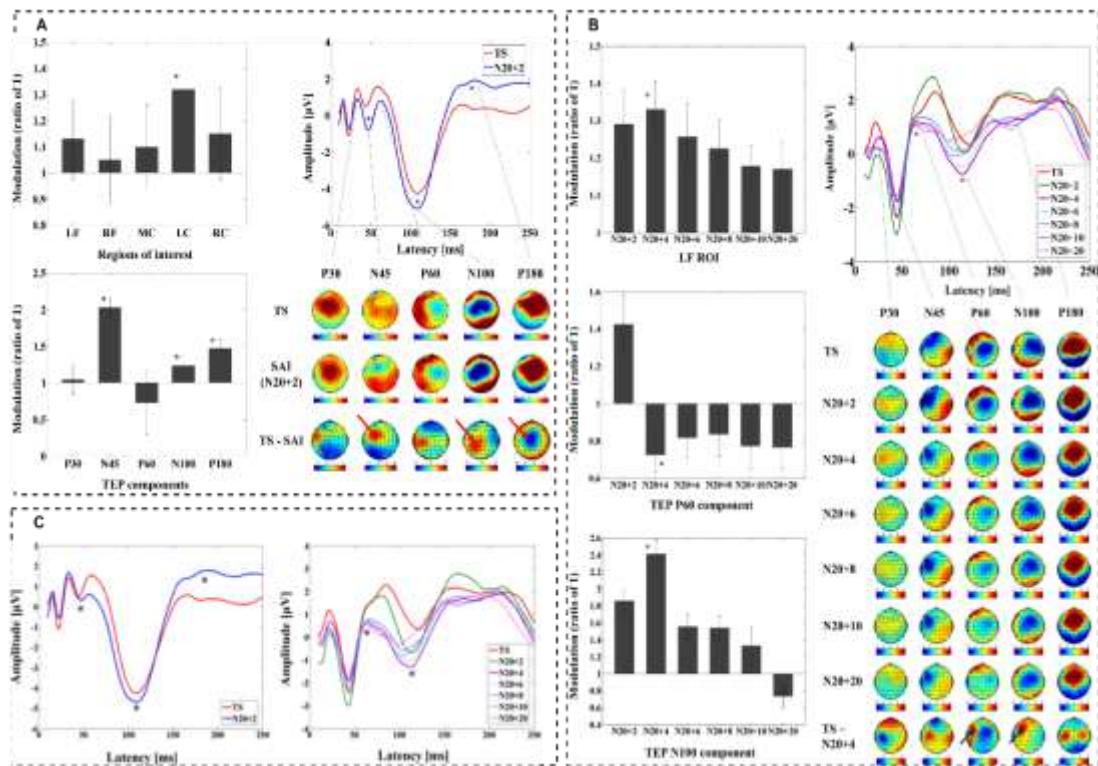


Fig. 13 TMS-EEG study of SAI in the M1 and DLPFC. (A) Modulation of cortical activity by SAI with TMS delivered to M1 (M1-SAI). (B) Modulation of cortical activity by SAI with TMS delivered to DLPFC (DLPFC-SAI). (C) TEP traces in the SAI paradigm for stimulation of M1 (left) and DLPFC (right) stimulation without SSEP subtraction. Adapted and modified from *Y. Noda, R, et al., 2016²⁹*.

TMS-EEG brain stimulation protocols

Different non-invasive transcranial brain stimulation (NTBS) protocols that modulate cortical circuits through plasticity-like effects include regular repetitive transcranial magnetic stimulation (rTMS), theta-burst stimulation (TBS), paired-associative stimulation (PAS) and transcranial direct current stimulation (tDCS).

Repetitive application of TMS pulses (rTMS) can assess neuroplasticity. Applying the repetitive pairing of TMS pulses to two brain regions, or to a sensory cortex with an appropriately timed peripheral sensory stimulus (paired associative stimulation, PAS) would induce spike-timing dependent plasticity. rTMS and PAS protocols of different stimulation frequency, pattern, location can enhance or suppress neural activity beyond the stimulation duration. Following active rTMS to the motor cortex, increases / decreases in MEP amplitudes in response to fixed intensity single-pulse TMS are thought to provide an index of long-term potentiation-like (LTP-like) /long-term depression-like (LTD-like) plasticity. Plasticity-inducing protocols can have behavioral effects and might be leveraged for therapeutic applications²⁰.

Regular repetitive transcranial magnetic stimulation(rTMS)-EEG protocol

rTMS entails the delivery of sequences or trains of magnetic pulses at diverse frequencies. When administered with an appropriate temporal pattern, duration, and intensity, these

magnetic pulses in rTMS are anticipated to induce lasting changes in synaptic efficiency that persist beyond the stimulation period. Conventional rTMS paradigms typically employ suprathreshold pulses and extended sequences of stimuli lasting 10 to 25 min. When applied over motor regions, it is widely accepted low frequencies (≤ 1 Hz) predominantly results in suppression, while higher frequencies (5-20 Hz) tend to facilitate MEPs. One central hypothesis about the modulatory impact of rTMS in human cortex bear resemblance to long-term potentiation (LTP) and depression (LTD) observed in animal trials, which can be quantified *via* the response amplitude alterations to electrical extracellular electrodes stimulation (Fig. 14).

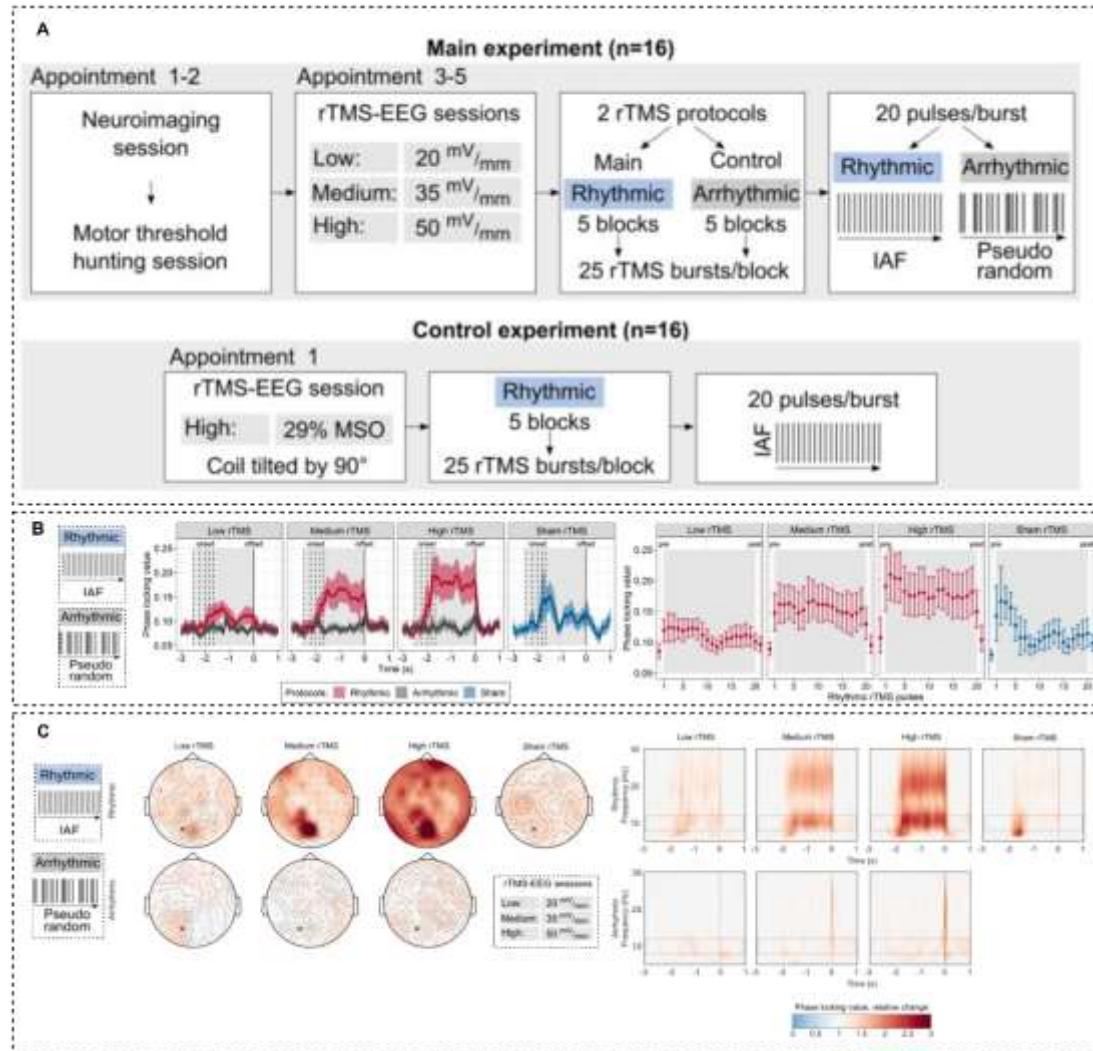


Fig. 14 rTMS-EEG experiment. (A) experiments setting. (B) Increased and sustained neural synchronization during rhythmic but not during arrhythmic or sham rTMS. (C) Rhythmic rTMS synchronized ongoing posterior α rhythms indicated by increased phase locking values. Adapted and modified from E. Zmeykina, et al., 2020³⁰.

Theta-burst stimulation (TBS)-EEG protocol

TBS composed of three or more subthreshold stimuli at a high frequency within the theta range (30-50 Hz), which are then repeated at a lower carrier frequency (usually 5 Hz). θ -

frequency patterns have been shown to mimic the natural rhythms associated with synaptic plasticity mechanisms such as long-term potentiation (LTP) and long-term depression (LTD) according to animal studies. When applied intermittent TBS (iTBS), increases in corticospinal excitability are expected over motor regions. Inversely, when applied continuous TBS (cTBS) inhibition is observed¹³. To assess modulation of brain activity in response to TBS, investigating effects of active iTBS and cTBS protocols on resting state EEG in the same group of participants and compare to sham TBS could be explored (Fig.15).

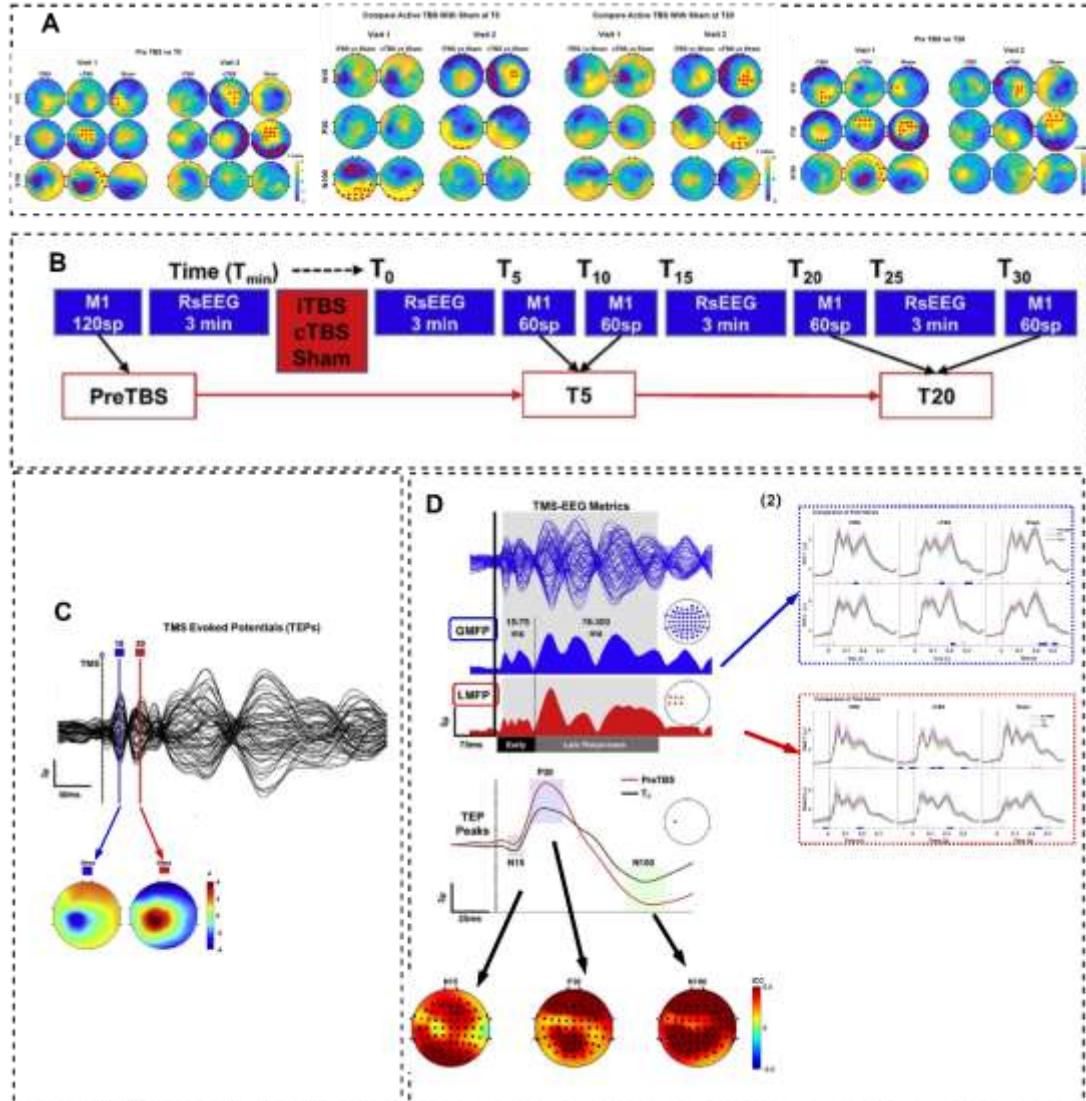


Fig. 15 TSB-EEG experiment. (A) TEP responses to different TBS at T5(left), T20(right) and sham control(middle). (B) TMS-EEG session protocol. (C) TMS evoked potentials from 63 channels (upper panel) with selected peaks (colored vertical lines) and topographical distribution of selected peaks of N15 & P30 (D) (1) Computation of GMFP & LMFP in left motor cortex within time-windows (early - late responses following TMS pulse), (2) Statistical comparison GMFP & LMFP responses at the millisecond level for both visit1 and visit2. (3) selected TEP peaks extracted from C3 electrode before (red line) and after (black line) iTBS at T20 in a representative subject (4) related topographical distribution of intra-class correlation coefficients of each TEP peaks across all visits at the electrode level. Adapted and modified from R. A. Ozdemir, et al., 2021³¹.

Paired-associative stimulation (PAS)-EEG protocol

PAS involves pairing a suprathreshold electrical stimulus applied to a peripheral nerve, usually the median nerve, with a suprathreshold TMS pulse applied to contralateral M1 using brief intervals known as ISIs. Modifying the ISI in PAS protocols effectively adjusts its impact, reflecting the principles observed in animal models of spike-timing-dependent plasticity (STDP). If a presynaptic input precedes postsynaptic excitation, synaptic transmission is facilitated, whereas, if postsynaptic excitation precedes a presynaptic input, transmission is inhibited. There is a general consensus that ISIs of 21.5-25 ms are facilitatory, aligning with a scenario where a pre-synaptic input preceding post-synaptic excitation, whilst shorter ISIs about 10 ms or so, are inhibitory.

PAS at an ISI of 25ms led to an enhancement of cortical excitability, as evidenced by increased GMFP and MEP amplitudes, not only in the hemisphere ipsilateral but also contralateral to the stimulation site. Conversely, PAS at an ISI of 10ms resulted in a decrease of GMFP and MEP amplitudes at the direct site of stimulation. Changes detected through TMS-EEG and TMS-EMG were characterized by a large inter-subject variability¹³. Combining TMS with EEG we aimed at investigating PAS effects and the connectivity modulation induced in humans (*Fig.16*).

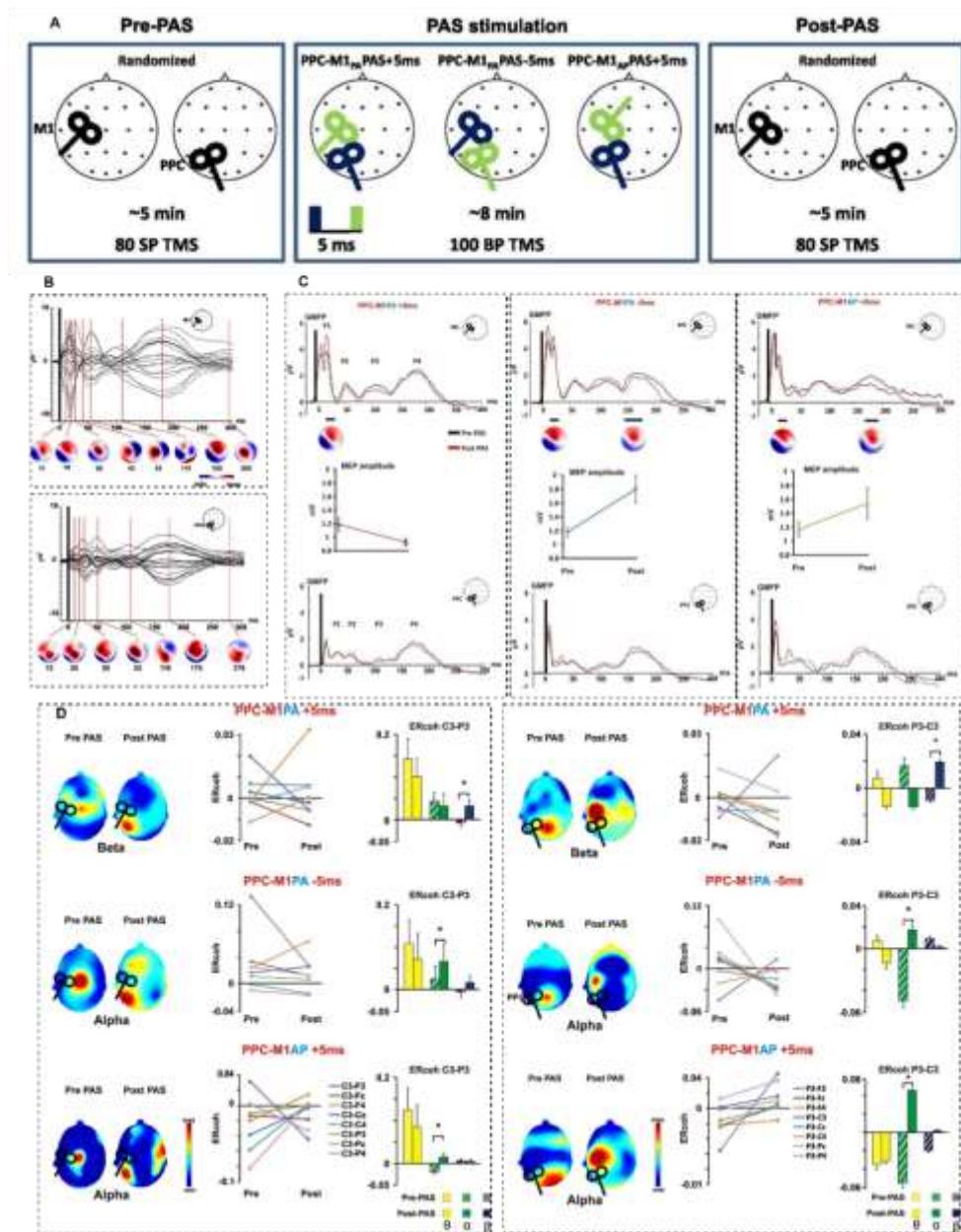


Fig. 16 PAS-EEG experiment. (A) Schematic representation of the experimental setting: eighty single pulse TMS were delivered before and after the administration of three different PAS protocols in a randomized order over M1 or PPC of the left dominant hemisphere. posterior-anterior (PA) to anterior-posterior (AP). (B) Grand-average of TEPs in a butterfly plot as a result of M1 (upper) and PPC (lower) stimulation with related distribution tomographs in specific time-domain. (C) Global cortical reactivity changes induced by PAS protocols of PPC-M1PA+5ms, PPC-M1PA-5ms and PPC-M1AP +5ms. (D) M1 (left) & PPC (right) event-related coherence changes induced by distinct PAS protocols.

Adapted and modified from *D. Veniero, et al., 2013*³².

Conclusion

In summary, TMS-EEG methodology is widely applied in neuroscience fundamental research and clinical implement due to the electrophysiological properties for studying neuron

excitatory/inhibitory and neuronal population plasticity. Aspects of artifact correction, data analysis and clinical protocols of TMS-EEG are discussed for references considering to avoid pitfalls, though pros and cons. The future expectations would be combining more modern pharmaceutical and biological methods in exploring neuron conditions of all kinds in humans.

Case Study of Analyzing Theta-Burst Stimulation to Frontal Cortex Based on Open Neuroscientific Data

Hua Cheng

Abstract

This study aims to leverage an openly available, high-quality EEG dataset to delve into the alterations in cortical activity. By applying Intermittent theta-burst stimulation (iTBS) and continuous theta-burst stimulation (cTBS) to the left dorsolateral prefrontal cortex (DLPFC) in healthy individuals, we observe changes in oscillatory patterns within the EEG data.

The dataset includes meticulously extracted resting-state EEG recordings, TMS-evoked potential data, and MRI scans. To process these data, we utilized Brainstorm, an open-source Matlab application, which facilitated noise reduction through independent component analysis and signal-space projection techniques. It allowed us to identify, visualize, and analyze TMS-evoked potentials (TEPs) and TMS-induced oscillations (TIOs). In addition, the study presents detailed plots of resting-state EEG power, local mean field power (LMFP), TMS-related spectral perturbation (TSRP), and inter-trial phase clustering (ITPC). Paired t-tests and cluster-based permutation tests have been performed for statistical analysis.

The wealth and quality of this dataset make it ideal for examining the neuromodulatory impact of TBS on the prefrontal cortex. Brainstorm's extensive feature set greatly supports the exploration of such neurological data. Future research directions could concentrate on conducting source localization analyses and comparative group studies.

Data description

A comprehensive EEG dataset, openly accessible, offers data from both resting-state measurements and simultaneous single-pulse TMS-EEG sessions. This dataset facilitates an investigation into the alterations of cortical activity resulting from the application of Intermittent theta-burst stimulation (iTBS) and continuous theta-burst stimulation (cTBS) to the left dorsolateral prefrontal cortex (DLPFC) in individuals with no health issues. TBS intervention modifies oscillatory patterns within the specific frequency bands inherent to this type of intervention (*i.e.*, 5 Hz and 50 Hz).

The EEG data were acquired using a *Refa* 2048 Hz EEG system and an appropriately sized 64-channel 10–20 EEG cap as determined by head circumference, with sintered, interrupted disk, Ag-AgCl TMS-compatible electrodes. The position of the EEG cap was confirmed by matching the **Cz** electrode with the intersection of the participants' nasion-ion and tragus-tragus axes.

Electrodes were grounded to **Fpz**, and EEG signals were measured against a common average reference. To reduce scalp impedance, participants were instructed to wash their hair before attending each experiment session. Secondly, prior to cap placement, the participant's scalp was cleaned with alcohol swabs. Lastly, an electro-conductive gel and blunted needles were used to

lightly abrade the scalp to limit impedances to less than $50\text{ k}\Omega$, which is well below 1% of the input impedance ($100M\Omega$) of the EEG amplifier³³.

TBS and single-pulse TMS were delivered using a *MagPro® X100*. with a 65 mm diameter *Cool-B65 figure-8* stimulation coil. The coil was positioned tangentially to the scalp over the **F3** electrode in order to target the DLPFC. A 5-mm customised 3D-printed spacer was placed between the coil and the scalp at all times to avoid contact with electrodes to minimise post-pulse artefacts, electrode movement, and bone-conducted auditory input. The coil was oriented at a 45-degree angle relative to the parasagittal plane, and the TMS pulse was delivered using a biphasic waveform. A hard foam headrest connected to a mechanical arm was positioned on the contralateral temporal region of the stimulation site to ensure minimal participant movement³³. Open dataset² files are stored in *.mat* format and contain raw EEG data. The data in each file includes 68 labelled signals (64 EEG channels, ECG, HEOG, VEOG, and a Trigger channel to mark events). Experiment blocks are labelled using the naming structure, as below *Tab.1*.

Tab. 1 Naming structure of EEG task blocks for each session³³.

Steps	Experiment Block Sequence	Filename Format	Blocks type	Time courses
1	Eyes-open resting-state EEG	pre-rest_run-01	RS-EEG	5
2	Single-pulse TMS-EEG	pre-tep_run-01	TMS-EEG (pre)	10
3	Eyes-open resting-state EEG	pre-rest_run-02	RS-EEG	
4	cTBS/iTBS/Sham	tbs	Condition intervention	5
5	Single-pulse TMS-EEG	post-tep_run-01	TMS-EEG (T2)	15
6	Eyes-open resting-state EEG	post-rest_run-01	RS-EEG	
7	Single-pulse TMS-EEG	post-tep_run-02	TMS-EEG (T15)	15
8	Eyes-open resting-state EEG	post-rest_run-02	RS-EEG	
9	Single-pulse TMS-EEG	post-tep_run-03	TMS-EEG (T30)	10
10	Eyes-open resting-state EEG	post-rest_run-03	RS-EEG	

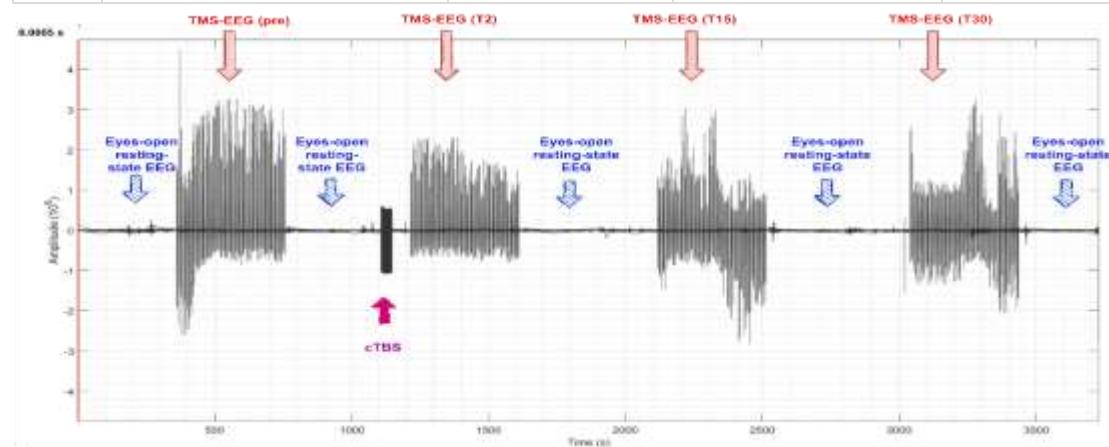


Fig. 17 cTBS protocol.

² The raw data files and code can be accessed via the FigShare open access repository service (<https://doi.org/10.25452/figshare.plus.c.5910329>). Url: <https://doi.org/10.25452/figshare.plus.c.5910329.v1>.

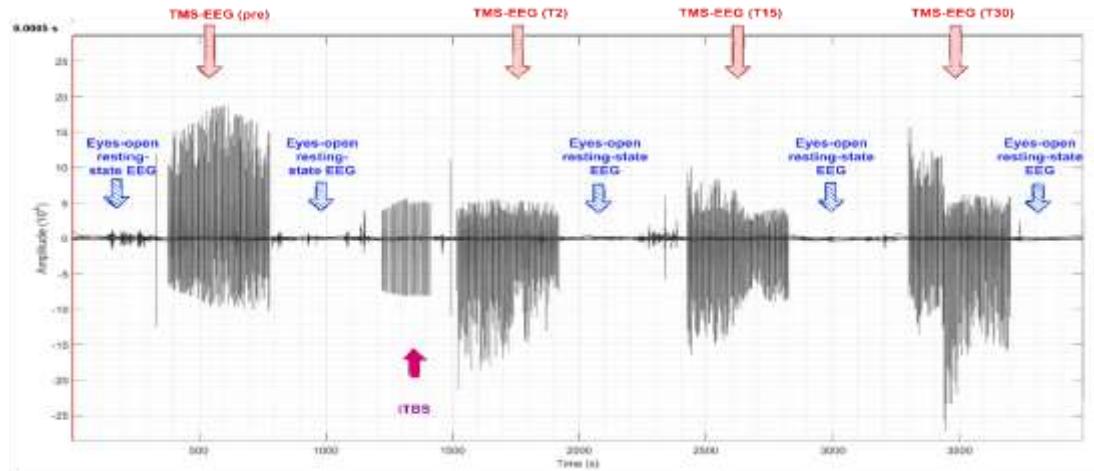


Fig. 18 iTBS protocol.

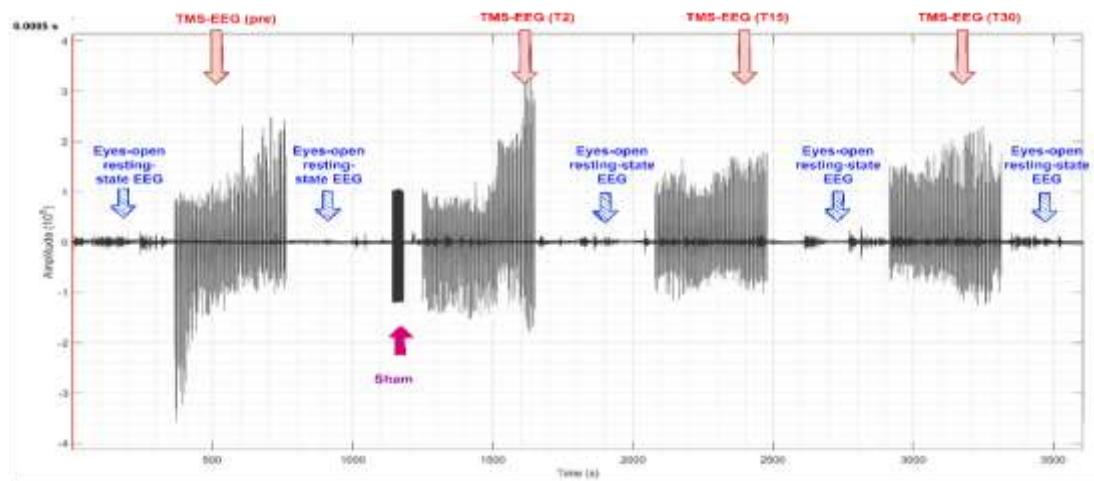


Fig. 19 Sham stimulation protocol.

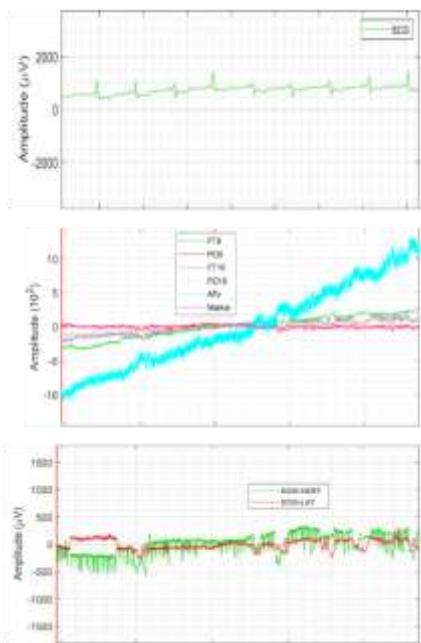


Fig. 20 ECG channel (top), EEG_NO_LOC channels (middle) and EOG EOG-VERT&EOG-LAT

channels(bottom).

Additional files include: a generic description of the metadata (*dataset_description.csv*), participant demographic and stimulation parameter details (*participants.csv*), metadata of the experiment tasks and EEG recording system (*eeg.csv*), a list of all EEG channels (*channels.csv*), neuronavigated coordinates (*electrodes.csv*), source data is provided in the MATLAB file format (.mat).

Resting-state EEG (RS-EEG) data

- 1) Eyes-open RS-EEG data were down-sampled to 512 Hz, baseline-corrected (demeaned) and detrended.
- 2) Then remove electrical line noise by using a second-order bandpass filter (0.1–70 Hz) and a notch filter at 50 Hz.
- 3) RS-EEG data were epoched in 1-s intervals. Reject epochs by using an automated algorithm in which epochs with data ranges greater than 3 standard deviations (SD) or absolute maxima greater than 12 SD of other epochs.
- 4) Reject remaining noisy epochs with a visual inspection.
- 5) Remove components containing eye blinks and muscle artefacts with a single round of independent component analysis (ICA).
- 6) Re-referenced EEG data to the common average reference.
- 7) Power spectral densities (PSD) were calculated using 180 s of data. Log-normalised power spectral density values ($\mu V^2/Hz$) were estimated for each EEG electrode over a range of 1–70 Hz using the fast Fourier transform (FFT) with 2-s sliding Hamming windows with 50% overlap.

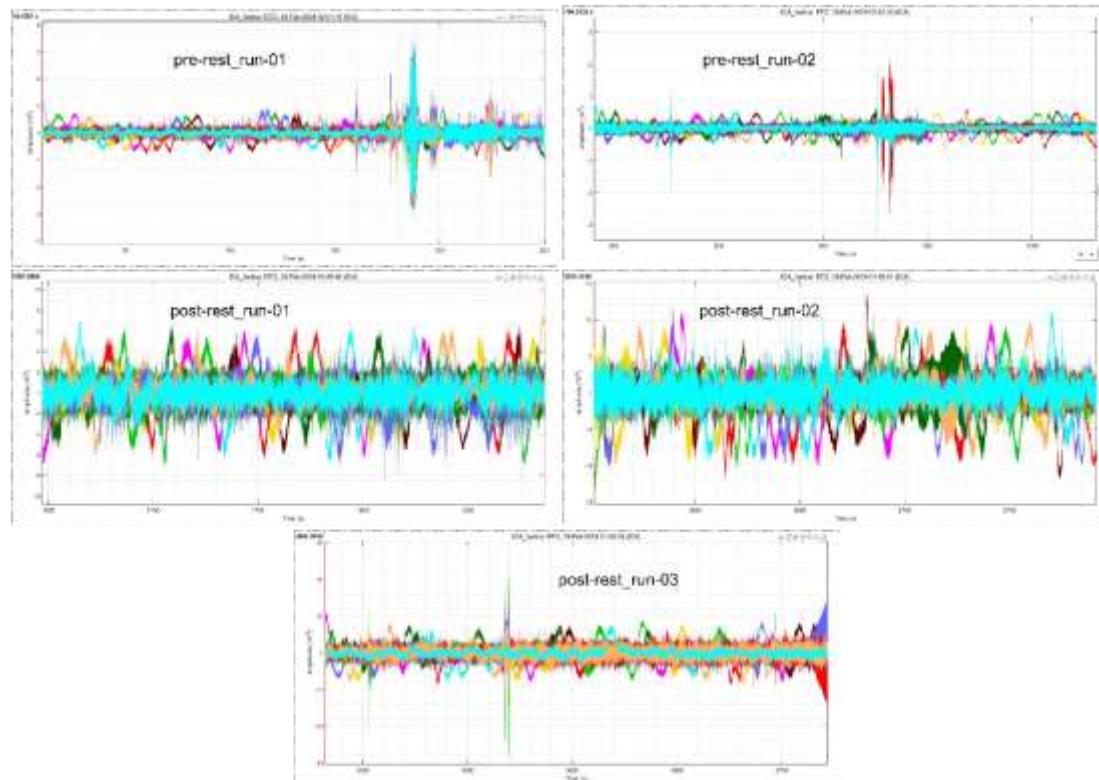


Fig. 21 Eyes-open resting-state EEG.

TMS-evoked potentials (TEPs) data

A double-blinded crossover design using five different occasions, includes sessions of 2 iTBS, 2 cTBS and 1 sham session. A burst of 3 pulses at 50Hz with 200 ms between bursts in a total of 600 pulses. The iTBS protocol involved a 2 s train of TBS repeated every 10 s for a total of 190 s (600 pulses), and cTBS consisted of a 40 s train of uninterrupted TBS (600 pulses). Sham TBS consisted of an inactive coil positioned on the head (at the same position as the active conditions) and a second active coil positioned 20 cm from the back of the head, facing away from it, with an increased stimulation output of 20% to compensate for the attenuation of the sound due to the additional distance from the ear.

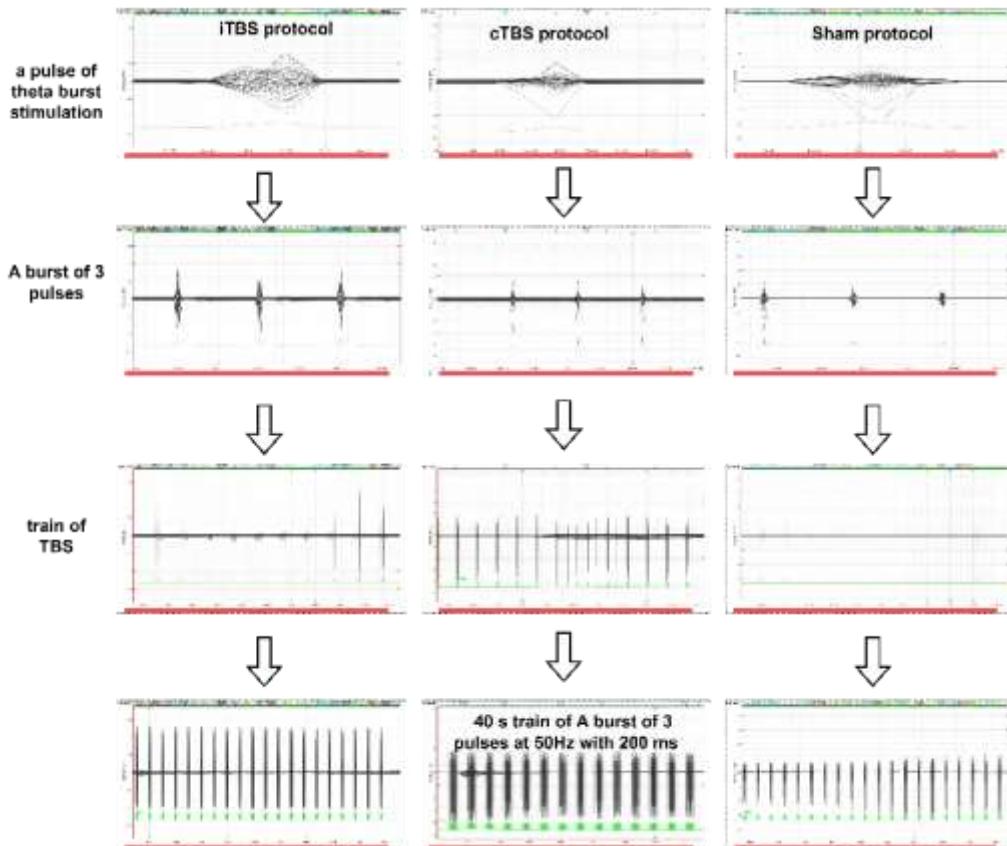


Fig. 22 TBS Experimental protocols.

Record the EEG responses to 100 single TMS pulses(sTMS) before each TBS condition (baseline block) and at 2-, 15- and 30-min post-TBS (T2, T15 and T30, respectively)³⁴.

- 1) TEPs were epoched around the TMS pulse (-1000 to 1000 ms).
- 2) Electrodes in which the TMS artefact exceeded the maximum absolute value of the range of the amplifier ($10^7 \mu V$) were removed and linearly interpolated from neighboring channels.
- 3) EEG traces were detrended and baseline-corrected relative to pre-TMS data (-500 to -50 ms).
- 4) Line noise (50Hz) was removed using linear regression by fitting and subtracting a sine wave from the EEG.
- 5) Data between -5 and 10 ms around the TMS pulse were removed.
- 6) An initial round of ICA was performed using the TESA *compselect* function to eliminate components containing eye blinks.

- 7) TMS-muscle and decay artefacts were removed by fitting a power law to the most negative and positive EEG signal deflections caused by the TMS artefact to obtain regression fit parameters, and then removing the artefact from the data by subtraction.
- 8) EEG data from before and after the TMS stimulus were filtered separately using a bandpass filter (1–90 Hz).
- 9) A second round of ICA was performed to remove this as well as components associated with blinks, eye movement, persistent muscle activity, decay artefacts and electrode noise.

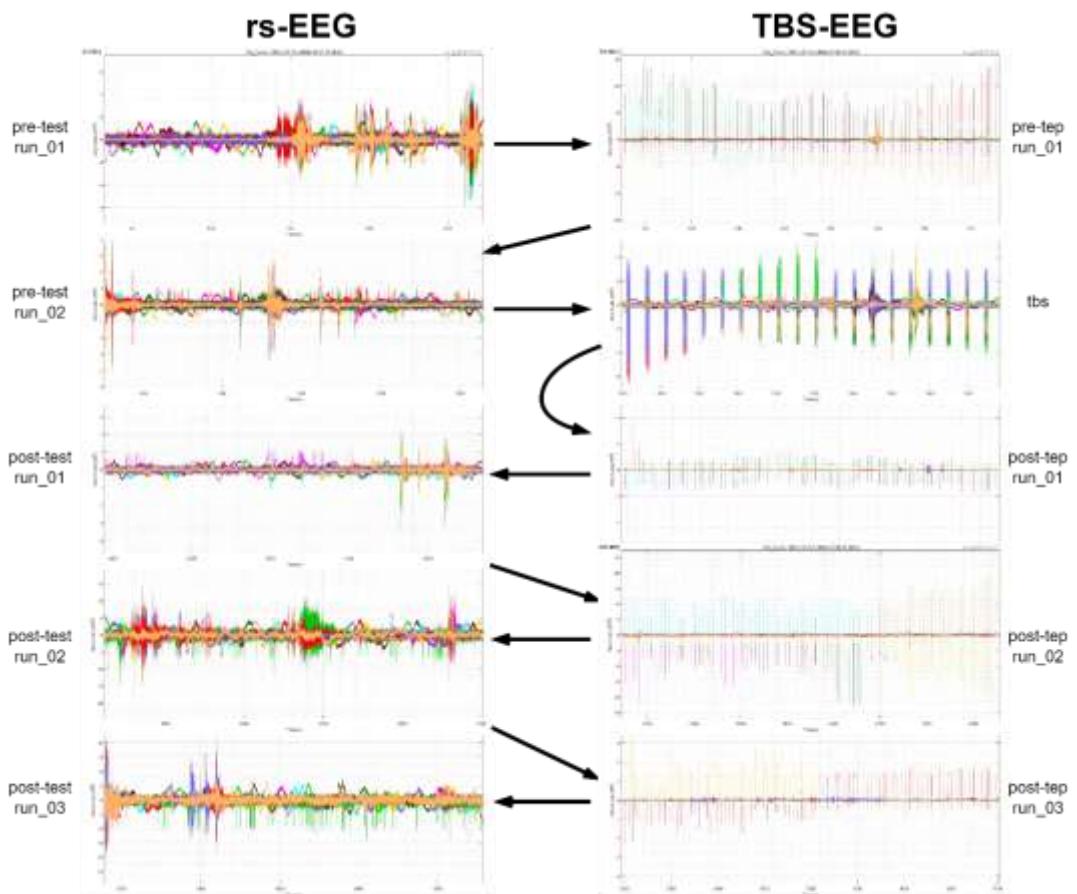


Fig. 23 Resting state EEG (left column) and Transcranial Magnetic Stimulation evoked EEG (right column) in different epochs. Black arrows represent the time flow.

MRI scan data

The MRI dataset can be used to assess how individual differences in brain morphology, including white matter fibre bundle size, grey matter volume, and whole brain volume, may affect the TMS-evoked potential. The MRI scans were obtained using a *Philips Achieva 3 T (TX) - DS MRI* scanner based at *Neuroscience Research Australia (NeuRA), Sydney, Australia*. A single scan was obtained for each participant at baseline prior to the start of the TMS sessions³³.

- 1) All MRI scans have been de-identified and anonymised using the **Fieldtrip ft_defacevolume** and **ft_anonymizedata** functions.
- 2) For all participants, *T1-weighted sequences* (TR=5.7ms, TE=2.6ms, FOV=250 ×250×190 mm, voxel size=1×1×1mm, matrix 250×250, Flip angle 8°, 190 sagittal plane

slices) were used to acquire structural MR images covering the whole brain.

- 3) In the same session, *high-resolution DTI* (TR=13737ms, TE=59ms, FOV=240×240×120 mm, voxel size=2×2×2mm, matrix 120×120, Flip angle 90°, 30 transverse plane slices) was also acquired.

Tab. 2 Naming structure of MRI scans for each session³³.

Experiment Block Sequence	Filename Format	type
3D Ultrashort Echo Time sequence	3DUTESkull2mmiso	MRI scan
Susceptibility weighted imaging	sWIP3DUTESkull2mmiso	MRI scan
Diffusion Tensor Imaging	DTI	DTI scan
T1-weighted image	T1075TFESag	MRI scan

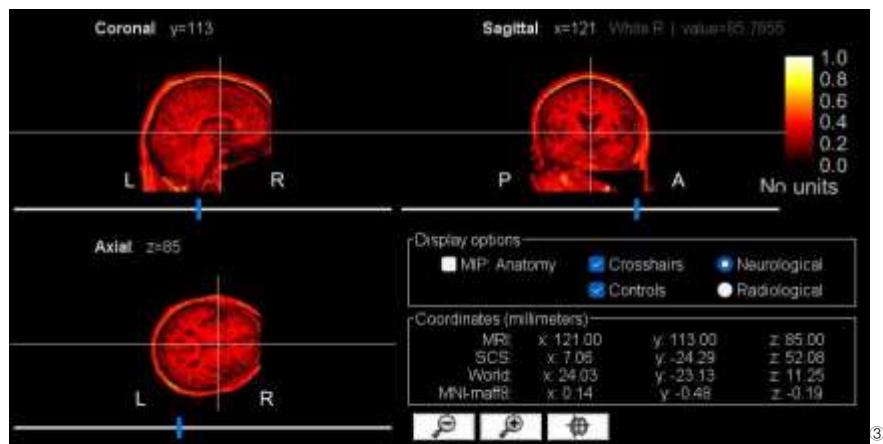


Fig. 24 MRI scan (from subject01).

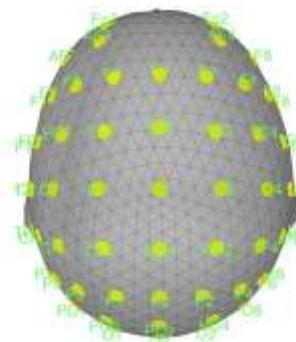


Fig. 25 EEG head sensors display.

Data process

Initially, upon importing the data, the ICBM152 2023b^④ template was incorporated as the default anatomical reference after warping. Subsequently, the process involved identifying physiological events such as heartbeats and eye blinks. Following this step, a series of filtering procedures were implemented for preprocessing purposes, aimed at cleaning and excluding unwanted signals from the ECG (electrocardiogram) and EOG (electrooculogram) channels.

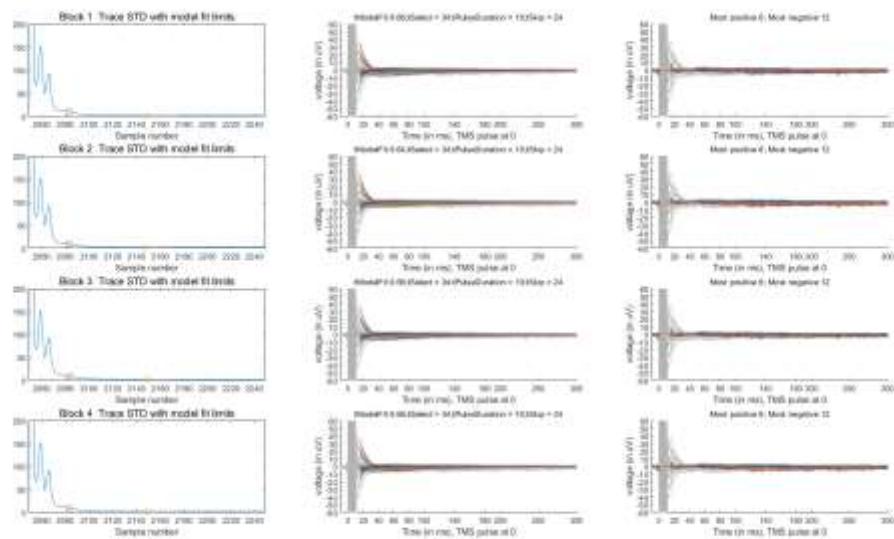


Fig. 26 Regression removal of TMS artefact.

Independent component analysis (ICA)

When conventional frequency filters fail to eliminate transient artifacts or those that spectrally overlap with the targeted brain signals, Independent Component Analysis (ICA) can be employed. This method discerns unique spatial patterns associated with artifacts and subsequently separates them from the EEG recordings. The key aspect of ICA is that it isolates components that are temporally independent. Alternatively, Signal-Space Projection (SSP) is another technique that can be used to correct for such artifacts, providing additional strategies to address these challenges in EEG data preprocessing.

^④ [Download - Brainstorm \(usc.edu\).](#)

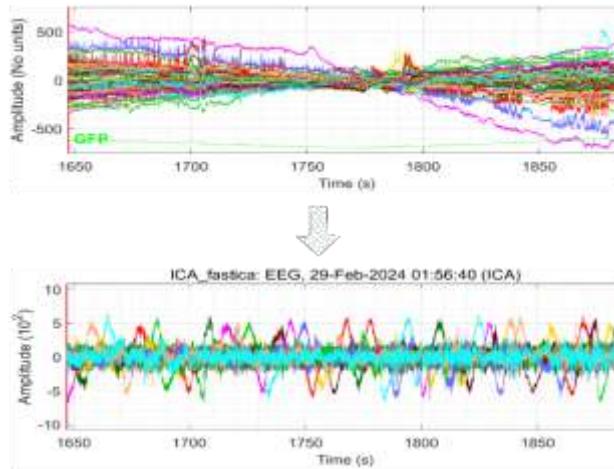


Fig. 27 ICA decomposition.

Signal-space projection (SSP)/Principal components analysis (PCA)

Signal-Space Projection (SSP) serves to pinpoint the characteristic sensor distributions linked to particular artifacts and generates spatial filters to effectively subtract the influence of these patterns from the recorded data. Implementing a Principal Components Analysis (PCA) on a concatenated set of artifacts aids in decomposing the distinct spatial elements present. In the case of removing heartbeats and eye blinks, SSP provides a streamlined solution that represents a subset of a broader, more generalized artifact correction procedure. The percentage (%) indicates the amount of signal (S_i) that

was captured by the component during the decomposition: $(\%) = \frac{S_i}{\sum S_i}$.

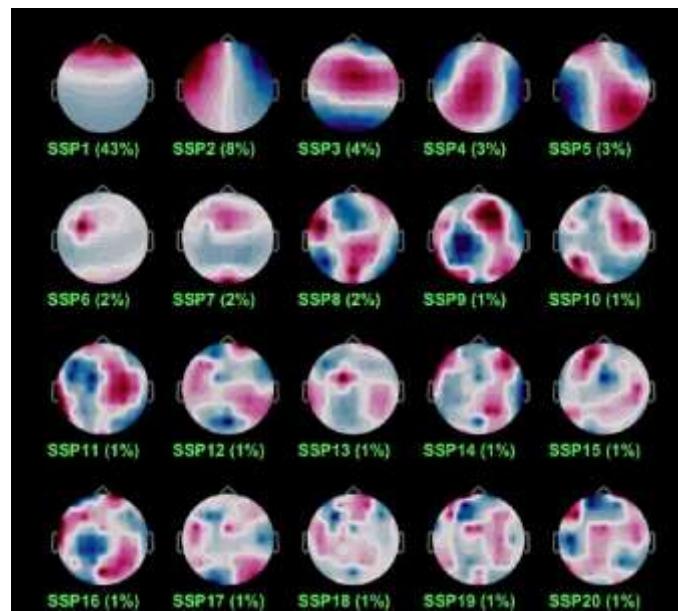


Fig. 28 Distribution topology plotting change with SSP/PCA decomposition percentage.

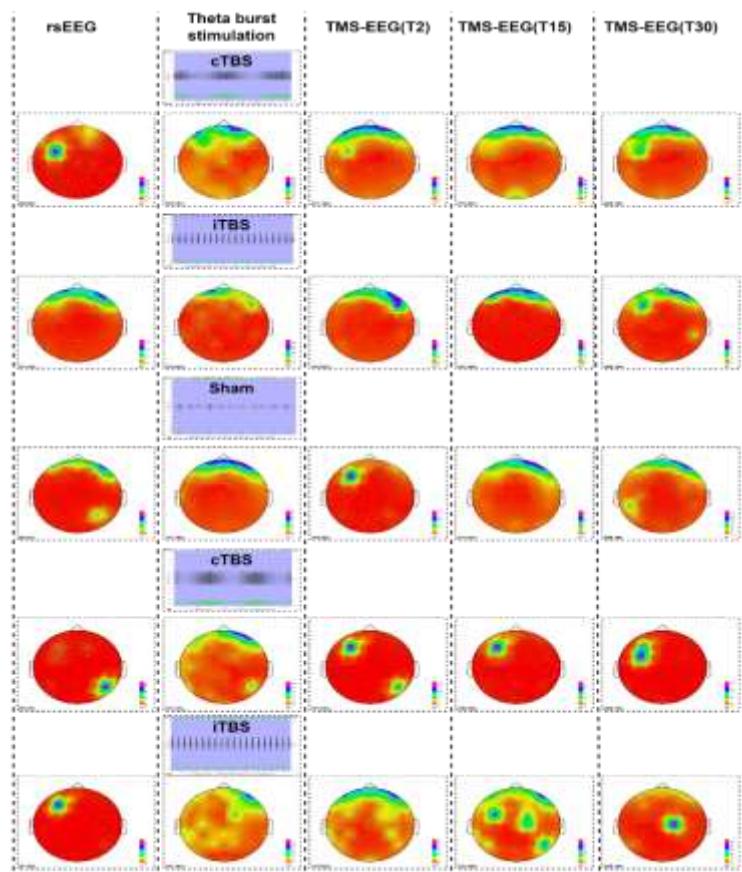


Fig. 29 Average distributions of subject 01 in difference condition.

TMS-evoked EEG potentials (TEPs)

TMS-evoked EEG potentials (TEPs) refer to the EEG responses elicited by Transcranial Magnetic Stimulation (TMS), which are obtained through the temporal averaging of EEG signals in the time domain³⁵.

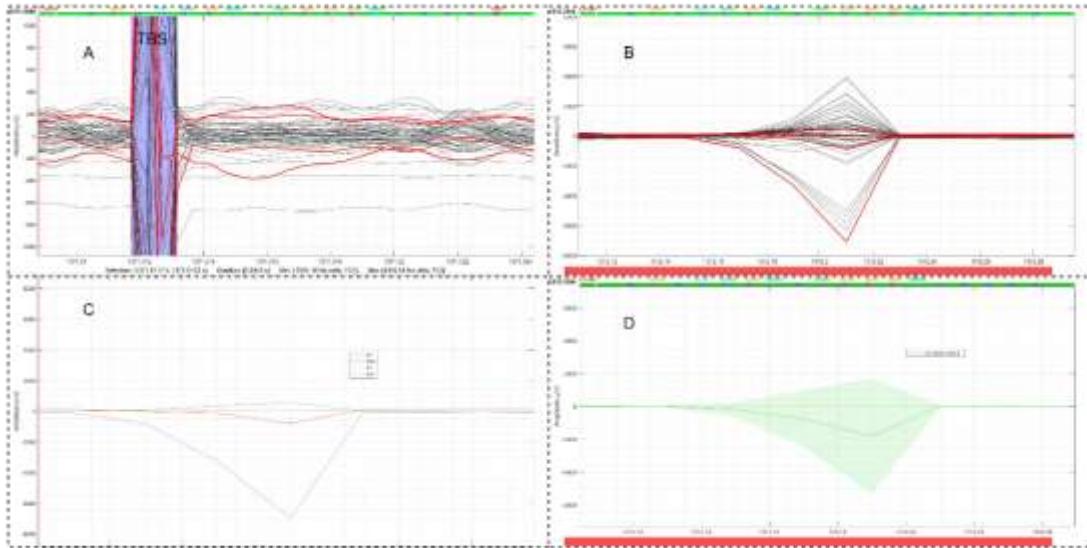


Fig. 30 TEPs plotting in time series. (A) Butterfly plot from all electrodes of one pulse of TBS and evoked potentials. (B)one pulse of a TBS (red lines represent channel of F1, F3, FC1, FC3). (C)one pulse of TBS in channels of interested (F1, F3, FC1, FC3). (D)Mean+Sdt of four electrodes (F3, FC3, F1, FC1).

TMS-induced oscillations

The time-frequency analysis of TMS-EEG data unveils TMS-induced oscillations that encode stimulus-phase-locked information, initially marked by a surge in δ (delta), θ (theta), α (alpha), and β (beta) band power, succeeded by suppression or de-synchronization in α and β bands, and ultimately culminating in an augmentation of β band power.

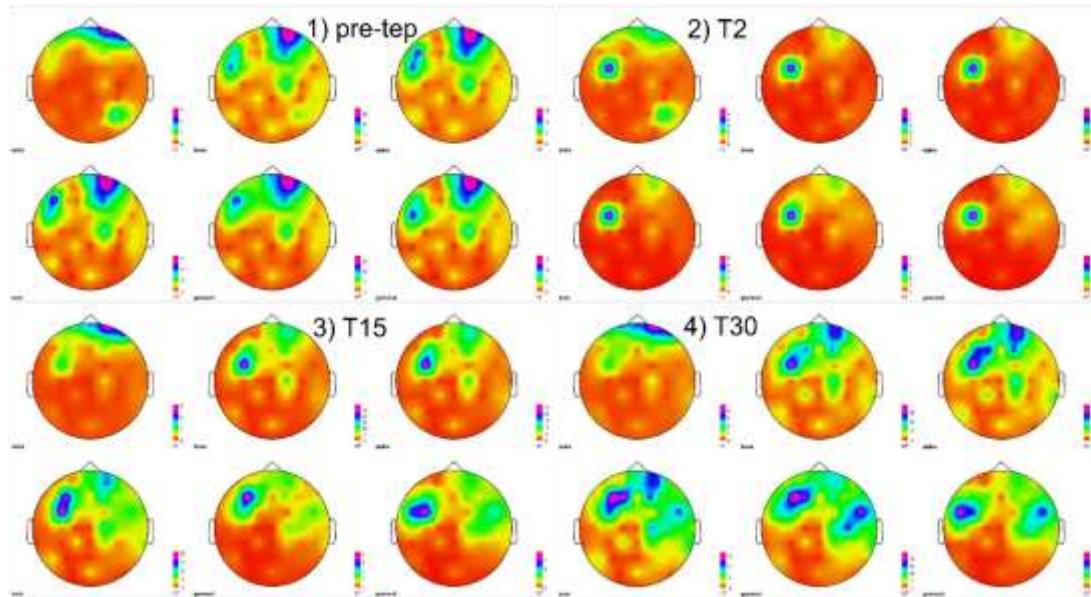


Fig. 31 Topology distribution plots of cortical status of δ (2-4Hz), θ (5-7Hz), α (8-12Hz), β (15-29Hz), γ 1(30-59Hz), γ 2(60-90Hz) in the time point of (1) pre-TMS-EEG, (3)T2, (4) T15, (5)T30 in Sham condition.

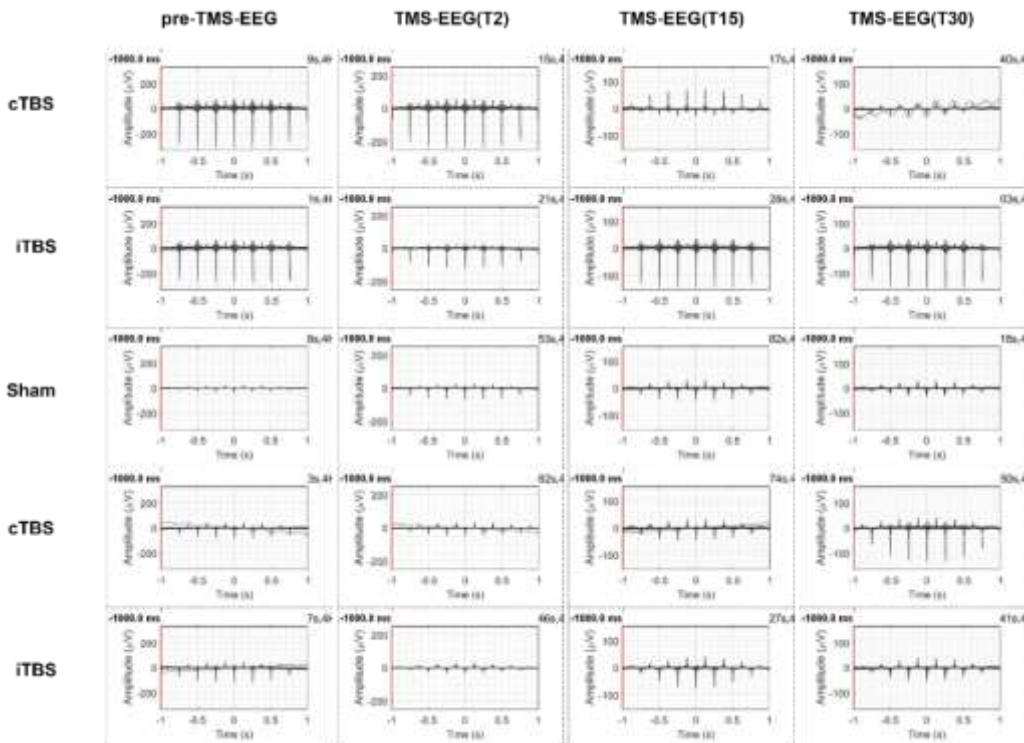


Fig. 32 Time-frequency decomposition visualizations of subject01.Assess the phase-amplitude coupling in different conditions. The nesting frequency (low) is set to 4Hz.

Resting state EEG power, LMFP, TRSP and ITPC

Resting EEG measures combining TMS-EEG might represent a more thorough reflection of

cortical excitability. The local mean field power (LMFP) as the square root of squared TEPs averaged across the four channels of interest. TMS-related spectral perturbation (TRSP) was evaluated locally by averaging the values obtained by the electrodes surrounding the stimulation site (FC3, FC1, F3, F1). The amount of TMS-related spectral perturbation (TRSP) was computed as³⁶:

$$TRSP(f, t) = \frac{1}{n} \sum_{k=1}^n |F_k(f, t)|^2 \quad (1)$$

Inter-trial phase clustering (ITPC) was computed according to³⁶:

$$ITPC(f, t) = \frac{1}{n} \sum_{k=1}^n \frac{F_k(f, t)}{|F_k(f, t)|} \quad (2)$$

In Equ (1) & (2), according to Delorme and Makeig³⁷, for n trials, the spectral power or amplitude estimates P and F were computed at trial k , at frequency f and time t .

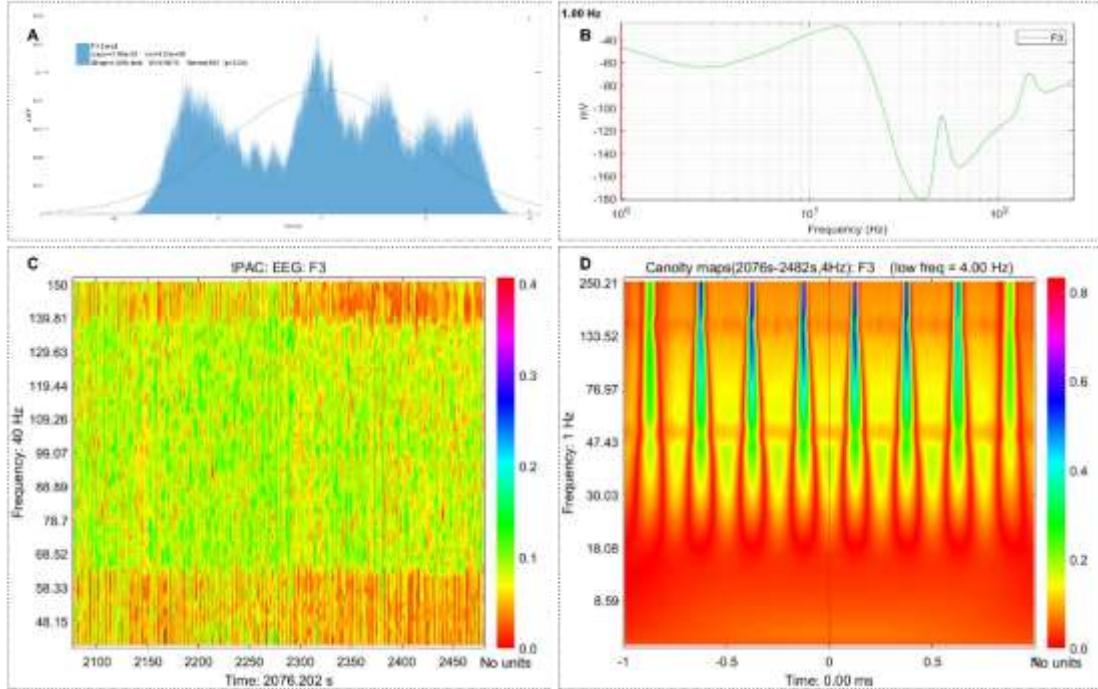


Fig. 33 Example of F3 from subject-01_session-03_post-tep_run-02. (A) LMFP; (B)resting EEG power; (C) TRSP; (D) ITPC.

Statistical analysis

Paired t tests

Reliability was tested in repetitive tests within same conditions. Differences of means in two visits of the same subject were made.

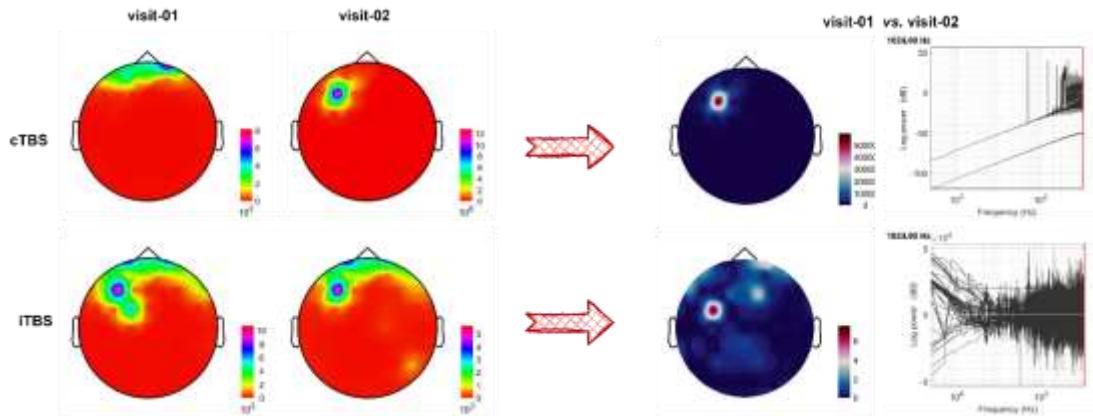


Fig. 34 Comparation between grand average of topological distributions of subjects in visit_01 and visit_02($\text{mean} = -4.66 \times 10^5$, $\text{std} = 3.33 \times 10^7$, $P < 0.05$).

No significant differences were detected between pre- vs. post- TEP data within the same TBS condition($P < 0.05$).

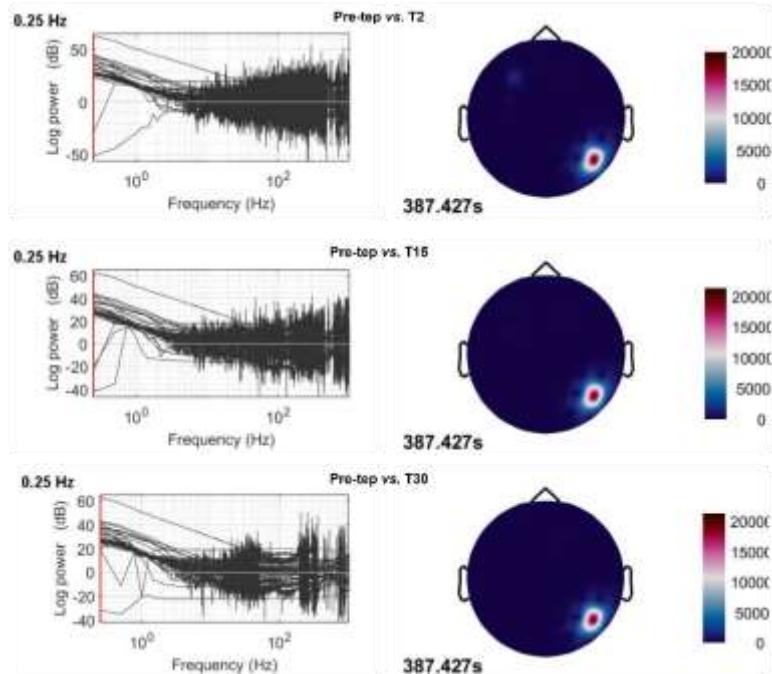


Fig. 35 pre-vs. post- TEP data within the same cTBS condition of sub01. (upper row) pre-TEP Vs. T2-TMS-EEG data. (middle row) pre-TEP Vs. T15-TMS-EEG data. (bottom row) pre-TEP Vs. T30-TMS-EEG data.

Cluster-based permutation tests

Statistics between conditions (cTBS vs. iTBS vs. Sham) for each electrode (FC1, FC3, F1, F3) in a selected a region of interest around the stimulation site and the corresponding contralateral site and each frequency of interests: delta (2~4 Hz), theta (4~7 Hz), alpha (8~12 Hz) beta (13~30 Hz), gamma (30~45 Hz) and gamma2 (46~90 Hz).

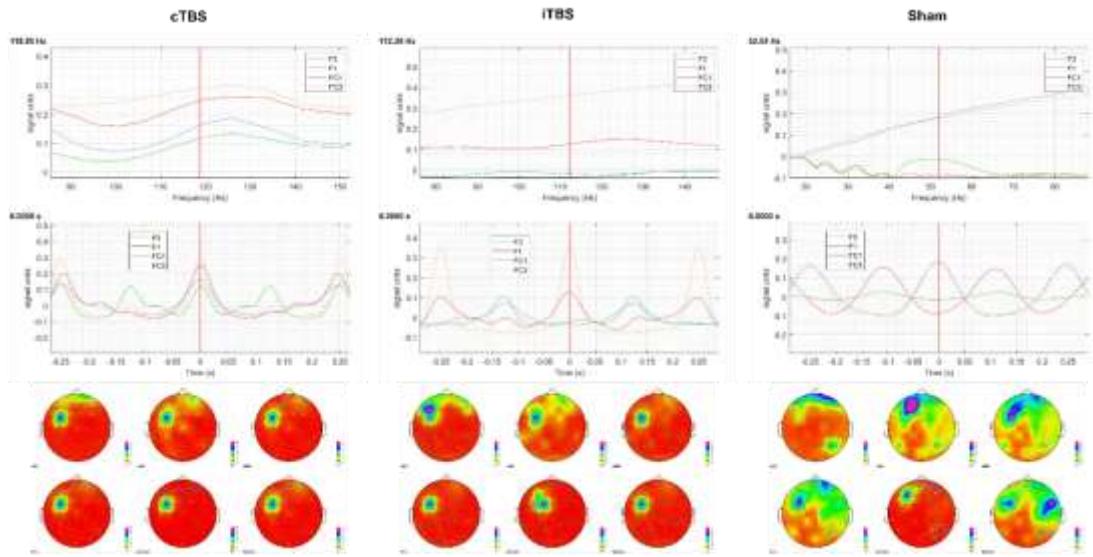


Fig. 36 Time-frequency-channels triplet compared. (upper row) power spectrum density of selected channels; (middle row) time series of selected channels;(bottom row) specific frequency band topological distribution of selected channels.

Conclusion

Assessing and monitoring the influence of pharmacological agents and non-invasive brain stimulation techniques, such as TMS, on brain activity and cortical networks via TEPs data is a widely adopted method to gauge the functional state of both healthy and diseased individuals. In this study, a high-quality, openly accessible TMS-EEG dataset was employed to investigate the neuromodulatory effects of theta burst stimulation on the prefrontal cortex. The Brainstorm and Fieldtrip toolkits integrated into MATLAB R2022b leveraged for data preprocessing, analysis, and visualization. While certain source EEG data were managed to extract and analyze to reveal dynamic changes following diverse events, there is yet exploited potential to extract and reconstruct more data with the obstacles overcome in the advancing of computational resources.

References

- 1 Hallett, M. Transcranial magnetic stimulation and the human brain. *Nature* **406**, 147-150, doi:10.1038/35018000 (2000).
- 2 Hernandez-Pavon, J. C. *et al.* TMS combined with EEG: Recommendations and open issues for data collection and analysis. *Brain Stimulation* **16**, 567-593, doi:<https://doi.org/10.1016/j.brs.2023.02.009> (2023).
- 3 Rogasch, N. C., Daskalakis, Z. J. & Fitzgerald, P. B. Cortical inhibition, excitation, and connectivity in schizophrenia: a review of insights from transcranial magnetic stimulation. *Schizophr Bull* **40**, 685-696, doi:10.1093/schbul/sbt078 (2014).
- 4 Lioumis, P., Kicić, D., Savolainen, P., Mäkelä, J. P. & Kähkönen, S. Reproducibility of TMS-Evoked EEG responses. *Hum Brain Mapp* **30**, 1387-1396, doi:10.1002/hbm.20608 (2009).
- 5 Kallioniemi, E., Saari, J., Ferreri, F. & Määttä, S. TMS-EEG responses across the lifespan: Measurement, methods for characterisation and identified responses. *Journal of Neuroscience Methods* **366**, 109430, doi:<https://doi.org/10.1016/j.jneumeth.2021.109430> (2022).

- 6 Ferreri, F. *et al.* Human brain connectivity during single and paired pulse transcranial magnetic stimulation. *NeuroImage* **54**, 90-102, doi:<https://doi.org/10.1016/j.neuroimage.2010.07.056> (2011).
- 7 Premoli, I. *et al.* TMS-EEG signatures of GABAergic neurotransmission in the human cortex. *J Neurosci* **34**, 5603-5612, doi:10.1523/jneurosci.5089-13.2014 (2014).
- 8 Rogasch, N. C. & Fitzgerald, P. B. Assessing cortical network properties using TMS-EEG. *Hum Brain Mapp* **34**, 1652-1669, doi:10.1002/hbm.22016 (2013).
- 9 Darmani, G. *et al.* Effects of the Selective α5-GABAAR Antagonist S44819 on Excitability in the Human Brain: A TMS-EMG and TMS-EEG Phase I Study. *J Neurosci* **36**, 12312-12320, doi:10.1523/jneurosci.1689-16.2016 (2016).
- 10 Du, X. *et al.* TMS evoked N100 reflects local GABA and glutamate balance. *Brain Stimul* **11**, 1071-1079, doi:10.1016/j.brs.2018.05.002 (2018).
- 11 Belardinelli, P. *et al.* TMS-EEG signatures of glutamatergic neurotransmission in human cortex. *Sci Rep* **11**, 8159, doi:10.1038/s41598-021-87533-z (2021).
- 12 Fecchio, M. *et al.* The spectral features of EEG responses to transcranial magnetic stimulation of the primary motor cortex depend on the amplitude of the motor evoked potentials. *PLoS One* **12**, e0184910 (2017).
- 13 Tremblay, S. *et al.* Clinical utility and prospective of TMS-EEG. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* **130**, 802-844, doi:10.1016/j.clinph.2019.01.001 (2019).
- 14 Bai, Z., Zhang, J. J. & Fong, K. N. K. Intracortical and intercortical networks in patients after stroke: a concurrent TMS-EEG study. *J Neuroeng Rehabil* **20**, 100 (2023).
- 15 Rosanova, M. *et al.* Natural frequencies of human corticothalamic circuits. *J Neurosci* **29**, 7679-7685, doi:10.1523/jneurosci.0445-09.2009 (2009).
- 16 Ferreri, F., Vecchio, F., Ponzo, D., Pasqualetti, P. & Rossini, P. M. Time-varying coupling of EEG oscillations predicts excitability fluctuations in the primary motor cortex as reflected by motor evoked potentials amplitude: an EEG-TMS study. *Hum Brain Mapp* **35**, 1969-1980, doi:10.1002/hbm.22306 (2014).
- 17 Pellicciari, M. C., Veniero, D. & Miniussi, C. Characterizing the Cortical Oscillatory Response to TMS Pulse. *Front Cell Neurosci* **11**, 38, doi:10.3389/fncel.2017.00038 (2017).
- 18 Pellicciari, M. C., Veniero, D. & Miniussi, C. Characterizing the Cortical Oscillatory Response to TMS Pulse. *Frontiers in Cellular Neuroscience* **11**, doi:10.3389/fncel.2017.00038 (2017).
- 19 Dimigen, O. Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage* **207**, 116117, doi:<https://doi.org/10.1016/j.neuroimage.2019.116117> (2020).
- 20 Farzan, F. *et al.* Characterizing and Modulating Brain Circuitry through Transcranial Magnetic Stimulation Combined with Electroencephalography. *Front Neural Circuits* **10**, 73, doi:10.3389/fncir.2016.00073 (2016).
- 21 Farrens, J. L., Simmons, A. M., Luck, S. J. & Kappenman, E. S. *Electroencephalogram (EEG) Recording Protocol for Cognitive and Affective Human Neuroscience Research* (Research Square Platform LLC, 2021).
- 22 Tangwiriyasakul, C. *et al.* Tensor decomposition of TMS-induced EEG oscillations reveals data-driven profiles of antiepileptic drug effects. *Scientific Reports* **9**, 17057, doi:10.1038/s41598-019-53565-9 (2019).
- 23 Olejarczyk, E. *et al.* Statistical Analysis of Graph-Theoretic Indices to Study EEG-TMS Connectivity in Patients With Depression. *Front Neuroinform* **15**, 651082,

- doi:10.3389/fninf.2021.651082 (2021).
- 24 Farzan, F. *et al.* Reliability of Long-Interval Cortical Inhibition in Healthy Human Subjects: A TMS-EEG Study. *Journal of Neurophysiology* **104**, 1339-1346, doi:10.1152/jn.00279.2010 (2010).
 - 25 Paci, M., Di Cosmo, G., Perrucci, M. G., Ferri, F. & Costantini, M. Cortical silent period reflects individual differences in action stopping performance. *Sci Rep* **11**, 15158, doi:10.1038/s41598-021-94494-w (2021).
 - 26 Cash, R. F. H. *et al.* Characterization of Glutamatergic and GABA_A-Mediated Neurotransmission in Motor and Dorsolateral Prefrontal Cortex Using Paired-Pulse TMS-EEG. *Neuropsychopharmacology* **42**, 502-511, doi:10.1038/npp.2016.133 (2017).
 - 27 Noda, Y. *et al.* Characterization of the influence of age on GABA(A) and glutamatergic mediated functions in the dorsolateral prefrontal cortex using paired-pulse TMS-EEG. *Aging (Albany NY)* **9**, 556-572, doi:10.18632/aging.101178 (2017).
 - 28 Noda, Y. *et al.* Evaluation of short interval cortical inhibition and intracortical facilitation from the dorsolateral prefrontal cortex in patients with schizophrenia. *Scientific Reports* **7**, 17106, doi:10.1038/s41598-017-17052-3 (2017).
 - 29 Noda, Y. *et al.* A combined TMS-EEG study of short-latency afferent inhibition in the motor and dorsolateral prefrontal cortex. *Journal of Neurophysiology* **116**, 938-948, doi:10.1152/jn.00260.2016 (2016).
 - 30 Zmeykina, E., Mittner, M., Paulus, W. & Turi, Z. Weak rTMS-induced electric fields produce neural entrainment in humans. *Sci Rep* **10**, 11994, doi:10.1038/s41598-020-68687-8 (2020).
 - 31 Ozdemir, R. A. *et al.* Reproducibility of cortical response modulation induced by intermittent and continuous theta-burst stimulation of the human motor cortex. *Brain Stimul* **14**, 949-964, doi:10.1016/j.brs.2021.05.013 (2021).
 - 32 Veniero, D., Ponzo, V. & Koch, G. Paired associative stimulation enforces the communication between interconnected areas. *J Neurosci* **33**, 13773-13783, doi:10.1523/jneurosci.1777-13.2013 (2013).
 - 33 Moffa, A. H. *et al.* Neuromodulatory effects of theta burst stimulation to the prefrontal cortex. *Scientific Data* **9**, 717, doi:10.1038/s41597-022-01820-6 (2022).
 - 34 Moffa, A. H., Nikolin, S., Martin, D., Loo, C. & Boonstra, T. W. Reliability of transcranial magnetic stimulation evoked potentials to detect the effects of theta-burst stimulation of the prefrontal cortex. *Neuroimage: Reports* **2**, 100115, doi:<https://doi.org/10.1016/j.yrnirp.2022.100115> (2022).
 - 35 Biondi, A. *et al.* Spontaneous and TMS-related EEG changes as new biomarkers to measure anti-epileptic drug effects. *Scientific Reports* **12**, 1919, doi:10.1038/s41598-022-05179-x (2022).
 - 36 Rocchi, L. *et al.* Variability and Predictors of Response to Continuous Theta Burst Stimulation: A TMS-EEG Study. *Frontiers in Neuroscience* **12**, doi:10.3389/fnins.2018.00400 (2018).
 - 37 Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* **134**, 9-21, doi:10.1016/j.jneumeth.2003.10.009 (2004).

EEG-SVRec: An EEG Dataset with User Multidimensional Affective Engagement Labels in Short Video Recommendation

Shaorun Zhang*
zhangshaorun@126.com
DCST, Tsinghua University
Quan Cheng Laboratory
Zhongguancun Laboratory
Beijing, China

Peijie Sun
DCST, Tsinghua University
Quan Cheng Laboratory
Zhongguancun Laboratory
Beijing, China

Zhiyu He*
DCST, Tsinghua University
Quan Cheng Laboratory
Zhongguancun Laboratory
Beijing, China

Qingyao Ai†
DCST, Tsinghua University
Quan Cheng Laboratory
Zhongguancun Laboratory
Beijing, China

Yiqun Liu
DCST, Tsinghua University
Quan Cheng Laboratory
Zhongguancun Laboratory
Beijing, China

Ziyi Ye
DCST, Tsinghua University
Quan Cheng Laboratory
Zhongguancun Laboratory
Beijing, China

Min Zhang†
DCST, Tsinghua University
Quan Cheng Laboratory
Zhongguancun Laboratory
Beijing, China

ABSTRACT

In recent years, short video platforms have gained widespread popularity, making the quality of video recommendations crucial for retaining users. Existing recommendation systems primarily rely on behavioral data, which faces limitations when inferring user preferences due to issues such as data sparsity and noise from accidental interactions or personal habits. To address these challenges and provide a more comprehensive understanding of user affective experience and cognitive activity, we propose EEG-SVRec, the first EEG dataset with User Multidimensional Affective Engagement Labels in Short Video Recommendation.

The study involves 30 participants and collects 3,657 interactions, offering a rich dataset that can be used for a deeper exploration of user preference and cognitive activity. By incorporating self-assessment techniques and real-time, low-cost EEG signals, we offer a more detailed understanding user affective experiences (valence, arousal, immersion, interest, visual and auditory) and the cognitive mechanisms behind their behavior. We establish benchmarks for rating prediction by the recommendation algorithm, showing significant improvement with the inclusion of EEG signals. Furthermore, we demonstrate the potential of this dataset in gaining

insights into the affective experience and cognitive activity behind user behaviors in recommender systems. This work presents a novel perspective for enhancing short video recommendation by leveraging the rich information contained in EEG signals and multidimensional affective engagement scores, paving the way for future research in short video recommendation systems.

CCS CONCEPTS

- Information systems → Users and interactive retrieval.

KEYWORDS

Short video, EEG signal, Recommendation system.

ACM Reference Format:

Shaorun Zhang, Zhiyu He*, Ziyi Ye, Peijie Sun, Qingyao Ai, Min Zhang†, and Yiqun Liu. 2023. EEG-SVRec: An EEG Dataset with User Multidimensional Affective Engagement Labels in Short Video Recommendation. In *Proceedings of The 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*. ACM, New York, NY, USA, 12 pages. <https://doi.org/XXXXXX.XXXXXXX>

1 INTRODUCTION

In recent years, short videos have emerged as a popular medium for entertainment and communication across various social media platforms, attracting millions of users worldwide. These videos typically span from a few seconds to several minutes and encompass a broad spectrum of content. Short video platforms generally gather, process, and analyze user behavior data and video information. To enhance the recommendation quality and retain users, various recommendation strategies are employed, including interest-based recommendations [10, 43], popularity-based recommendations [2, 42],

*These authors contributed equally to this work.

†Corresponding Authors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '24, July 14-18, 2024, Washington D.C., USA

© 2023 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/XXXXXX.XXXXXXX>

and personalized recommendations [47]. The choice of a recommendation strategy can significantly impact users' affective engagement while browsing short videos.

Existing short video recommendation systems mainly focus on behavioral metrics, such as likes, dwell time, view percentage, etc., to improve recommendation performance [16, 30, 33]. These behavior data are usually collected from user logs and applied as implicit feedback signals to infer user preferences. Although these observed data usually contain abundant information, only considering existing information is not enough to gain a comprehensive understanding of users [1, 22]. There still exist challenges in capturing user preference from behavioral data. Firstly, behavioral data, such as likes and comments, is usually sparse. Secondly, the presence of noise, resulting from accidental interactions or personal habits, can affect the reliability of the data.

In order to deeply understand users' cognitive activities, we record EEG (Electroencephalograph) signals during short video browsing. EEG, as a neuroelectrical signal, containing rich spatial, temporal, and frequency band information about human experience, can be used to study the underlying neural mechanisms and can reflect relevant information about user cognition, emotion, and attention [20, 25, 35, 44]. Providing high temporal resolution data, the application of EEG technology in the Information Retrieval (IR) domain has been proven to be useful [5, 45]. At the same time, the latest developments in EEG recording devices are known for their high portability and low operating costs [15, 27], which are necessary for real-world application scenarios. The high temporal resolution of EEG data enables it to effectively address the real-time demands of short video recommendation scenarios.

To further understand the relationship between user behavior and EEG signals, it is essential to incorporate user affective experiences into the annotation of short videos. These affective experiences are from different dimensions. Emotion elicited by short videos plays a significant role in the browsing experience, which is commonly modeled via two dimensions: valence and arousal [36]. Both the degree to which short videos align with user interest and the level of immersion experienced by users while browsing short videos influence user behavior and perception. Besides, short videos serve as a combined visual and auditory medium, so understanding the impact of visual and auditory features on users' perceptions can be helpful. Accordingly, we collect six Multidimensional Affective Engagement Scores (MAES), which are valence, arousal, immersion, interest, visual and auditory, and extract the visual and auditory features of the videos.

By employing self-assessment techniques, we obtain a more detailed and multidimensional perspective of user experience. Furthermore, real-time, low-cost EEG signals can be utilized to gain insights into users' cognitive activity.

Therefore, We proposed to build EEG-VSRec¹, an EEG dataset with user Multidimensional Affective Engagement in Short Video Recommendations. We conducted the user study where participants continuously viewed short videos in several sessions. After each session, the participants rated the MAES for each video. We recruited 30 participants and collected 3,657 interactions, each with temporal EEG signals during viewing as well as user behavior and

multidimensional labels. Finally, we collected three types of data: user behavior log, EEG signals, and self-assessment of six MAES.

Subsequently, we present the statistical information of the dataset and show the rich information contained in the dataset. Besides, we discuss the possible applications for the dataset. We first show its impact on user understanding in the short video recommender system, with some primary discoveries thrown. We also establish benchmarks for rating prediction inferred from EEG signals and prevalent recommendation algorithms. Experiments show significant performance improvement with the inclusion of EEG signals, demonstrating the importance of introducing brain signals to recommender systems.

These are our main contributions:

- We proposed the first dataset that contains EEG signals in a real scenario of watching short video streaming. On the basis of user behavior, we provided multidimensional affective engagement scores (MAES), which are valence, arousal, immersion, interest, visual and auditory, as explicit feedback.
- We establish benchmarks for rating prediction by the recommendation algorithm. Comparative experiments show significant performance improvement with the inclusion of EEG signals, demonstrating the importance of introducing brain signals to recommender systems.
- We show the perspective of understanding the affective experience and cognition activity behind user behaviors in the recommender system.

The remainder of this paper is organized as follows: we review related datasets in Section 2. Then we introduce our dataset and its collecting procedure in Section 3. Section 4 presents the combination and the statistical analysis for our dataset. Next, we conducted experiments to show the potential applications in Section 5. Finally, Section 6 and Section 7 discuss and conclude our work.

2 RELATED DATASETS

In this section, we review the work of datasets in the short-video recommendation scenario and EEG datasets in affective computing and compare our dataset with theirs (Table 1).

2.1 Dataset in Short Video Recommendation

Short videos, a new type of online streaming media, have attracted increasing attention and have been one of the most popular internet applications in recent years. Thus, research on short video recommendations has gained traction, and some related datasets have been released.

The datasets in short video recommendations are usually collected from online platforms with user id, item id, and their interaction behavior. An unbiased sequential recommendation dataset KuaiRand [8] contains millions of intervened interactions on randomly exposed videos. Tenrec [46] is a large-scale and multipurpose real-world dataset with the item either a news article or a video. MicroLens [26] consists of one billion user-item interactions with raw modality information about videos. MMSSL [37] is collected from the TikTok platform to log the viewed short videos of users. The multi-modal characteristics are visual, acoustic, and title textual features of videos. Some researchers conduct experiments on the Micro-Video dataset to validate their model [21]. They construct

¹Dataset and codes are available at <https://anonymous.4open.science/r/Z-SV-CFB1>

Table 1: Comparison of the EEG-SVRec with other datasets in the video/music recommendation and the video affective computing domain. U&I represents user and item id. Peri.Bio represents peripheral biosignal (such as, heartbeat, eye tracking, ECG).

Domain	Datasets	Item/Stimulus	U&I	Impression	Ratings	Emotion	Peri. Bio	EEG
Recommendation (open domain)	Movielens	Movie	✓		✓			
	Toffee	Short Video	✓	✓				
	KuaiRand	Short Video	✓	✓				
	MMSSL	Short Video	✓	✓				
	Tenrec	News, Short Video	✓	✓				
	Last.fm	Music	✓	✓				
	MUMR	Music	✓	✓	✓	✓	✓	
Affective computing (closed domain)	DEAP	1min music videos	✓		✓	✓	✓	✓
	SEED	4min movie clips	✓		✓	✓	✓	✓
	AMIGOS	short and long movies	✓		✓	✓	✓	✓
Recommendation (open domain)	EEG-SVRec (ours)	Short Video	✓	✓	✓	✓	✓	✓

a dataset by randomly sampling 100K users and their watched micro-videos over a period of two days. Other researchers crawled micro-videos from Jan 2017 to Jun 2018 from Toffee, a large-scale Chinese micro-video sharing platform [38]. Though the item is not a short video, the dataset Movielens [11] interacting with the movie contains user ratings (ranged 1-5), which have large scale and have had a substantial impact on education, research, and industry.

Different from above, music dataset Last.fm-1k² represents the whole listening habits for nearly 1,000 users. MUMR [18] used a dataset in the music recommendation scenario with the collection of the contexts from low-cost smart bracelets. He et al. [13] consider immersion in online short videos with psychological labels, video features, and EEG signals. In contrast to them, we provide the dataset with various multidimensional affective engagements, giving a deep understanding of users.

Since we collected from user studies, our data contains detailed video and audio features, behavior logs, user multidimensional affection engagement scores, and EEG and ECG signals.

2.2 EEG Dataset in Affective Computing

EEG (Electroencephalogram) has been popular in neuroscience and psychology since it is a non-invasive technique used to measure the electrical activity of the brain. Utilizing physiological signals to help understand people's affection and cognition has become widespread in affection computing for its good balance between mechanistic exploration and real-world practical application. By analyzing EEG signals, researchers can identify patterns that are associated with different emotional states. Researchers collected EEG and peripheral physiological signals when using music, images, and videos as stimulation. Affection is annotated by the participants.

MIIR [32] record the EEG signals from 10 participants when listening to and imagining (by tapping the beat) 12 short music

fragments. Then they rate their taping ability and familiarity. Images can also be the stimulus. A dataset in neuromarketing containing EEG signals of 14 electrodes from 25 participants and their likes/dislikes on e-commerce products over 14 categories with 3 images each [41].

The stimulation of videos includes both visual and auditory aspects, making the information more diverse and rich. DEAP [17] is the dataset of 32 participants whose EEG and peripheral physiological signals were recorded as each watched 40 one-minute excerpts of music videos. The SEED database [49] contains EEG data of 15 subjects, which are collected via 62 EEG electrodes from the participants when they are watching 15 Chinese film clips with three types of emotions, i.e., negative, positive, and neutral. Moreover, AMIGOS [24] collected EEG, ECG, and GSR from 40 participants when watching 16 short videos and 4 long videos. Participants annotate their emotions during watching these videos with self-assessment of valence, arousal, control, familiarity, liking, and basic emotions.

Datasets play a very important role in EEG affective computing. New methods and models have been proposed based on existing datasets to facilitate evaluation. However, the videos given to participants to watch are pre-selected for stimulating different emotions and are consistent between participants (closed domain). Participants were unable to actively influence the video playback, e.g., slide done at any time to switch to the next video. Unlike them, our experiments take place in a real online short video browsing scenario, where videos come from among millions of videos on the platform and are presented to participants through personalized recommendation algorithms or in non-personalized or randomized ways (open domain). During browsing, participants actively engage in behaviors such as swiping and liking videos.

These existing research efforts show the application potential of EEG in various fields. In the context of short video recommendations, there still has no dataset to find the correlation between physiological signals and affective engagement during real short

²<https://www.last.fm/>

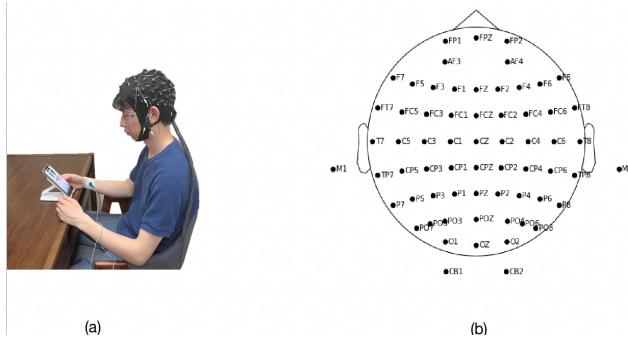


Figure 1: EEG and ECG data acquisition setup: (a) A participant wears an EEG cap while watching short videos in a laboratory setting (Image display has been approved). (b) International 10-20 electrode placement standard for EEG.

video scenarios. What we add on top of these works is that we conduct a user study where participants browse short videos in a real scenario and collect their behavior, multidimensional affective engagement labels, and EEG and ECG signals.

3 DATASET CONSTRUCTION

This section mainly covers ethical and privacy, participants, video stimuli material, apparatus, and experimental procedure (browsing stage and labeling stage).

3.1 Ethical and Privacy

Our user study has undergone review and obtained approval from the institutional ethics committee, xxx University (approve number: xxx³). This study has undergone a rigorous ethical review process to ensure the protection of the participant's rights. In compliance with established ethical guidelines, we have taken multiple measures to protect the participants' privacy, including anonymizing the collected data and obtaining informed consent from all participants before the study. Furthermore, participants were fully informed about the study's objectives, procedures, and potential outcomes. The EEG data collection method employed in this research is non-invasive and poses no harm to the participants. This approach ensures that the study adheres to ethical standards while maintaining the integrity of the research findings. As for the item in the dataset, we only provide anonymized video ids, encoded video tags, and extracted video characteristics (shown in Section 4.3).

3.2 Participants

We recruited 30 college students aged between 18 and 30 ($M=22.17$, $SD=2.20$) for our study. The participant group consisted of 16 males and 14 females, majoring in various fields such as computer science, law, medicine, and sociology. All participants were familiar with at least one short video platform and used it at least once a day. To protect participants' privacy, we provided each participant with a new account on the short video platform. Each participant was required to participate in two experimental settings: a 10-hour

³The protocol ID is hidden for double-blinded review

preference collection phase and a 3-hour lab study phase as Figure 1(a). (including preparation and rest time). Upon completion of the experiments, each participant received approximately 60 dollars in research compensation.

3.3 Video Stimuli Material

Participants browse short videos on a popular video platform, and all items are on the platform. The platform has two settings: personalized and non-personalized. Since they are all affected by the strategy of the platform, we present randomized videos as well. Thus, we categorized the short video stimuli to be presented to the participants into three video pools: personalized, non-personalized, and randomized.

The **personalized video pool** mainly consists of videos selected based on the preference information collected during the 10-hour preference acquisition phase for each participant, obtained through the short video platform's algorithm. The **non-personalized video pool**, with personalized-off, disregards user interaction history and distributes videos may be based on their current popularity ranking. It is worth mentioning that the videos in personalized and randomized pools have a duration of 30-60 seconds, while the non-personalized video pool's time restriction of 30-60 seconds was removed due to the distribution mechanism by platforms. The **randomized video pool** is sampled from the large video platform's video collection, filtered by different popularity levels. We first divided the large video pool into three levels based on view counts, and then randomly selected 100 videos from each level. After that, to ensure the category richness and healthiness of the selected short videos, we filtered 25 videos in each group.

The selection of videos from these three pools results in different session compositions. Four distinct session modes were established: **personalized mode**, **randomized mode**, **mixed mode**, and **non-personalized mode**. It is clear that the personalized and randomized modes consist of 20-30 specific videos from their respective video pools, with a duration of 30-60 seconds each. In the mixed mode, an assortment of 20-30 videos is presented, with an equal proportion of personalized and randomized videos, maintaining a 1:1 ratio. Video sequences are random, ensuring a well-distributed and varied exposure for the study participants. The non-personalized mode involves extracting a certain number of videos from the non-personalized video pool.

3.4 Apparatus

We used a smartphone with a 6.67-inch screen and a 120Hz refresh rate, which connected to a stable local area network (LAN) Wi-Fi to ensure network stability. Participants were allowed to adjust the screen brightness and device volume to a comfortable level before the experiment. They can also adjust the seat position and the angle of the smartphone to a suitable position. During the browsing stage, participants were required to minimize body and head movements to ensure the high quality of the collected physiological signals in Figure 1(a). A Scan NuAmps Express system (Compumedics Ltd., VIC, Australia) along with a 64-channel Quik-Cap (Compumedical NeuroScan) was utilized for recording the participants' EEG data in Figure 1(b) [14]. Some electrode points were also used to eliminate head movement and other artifacts. The impedance of the EEG

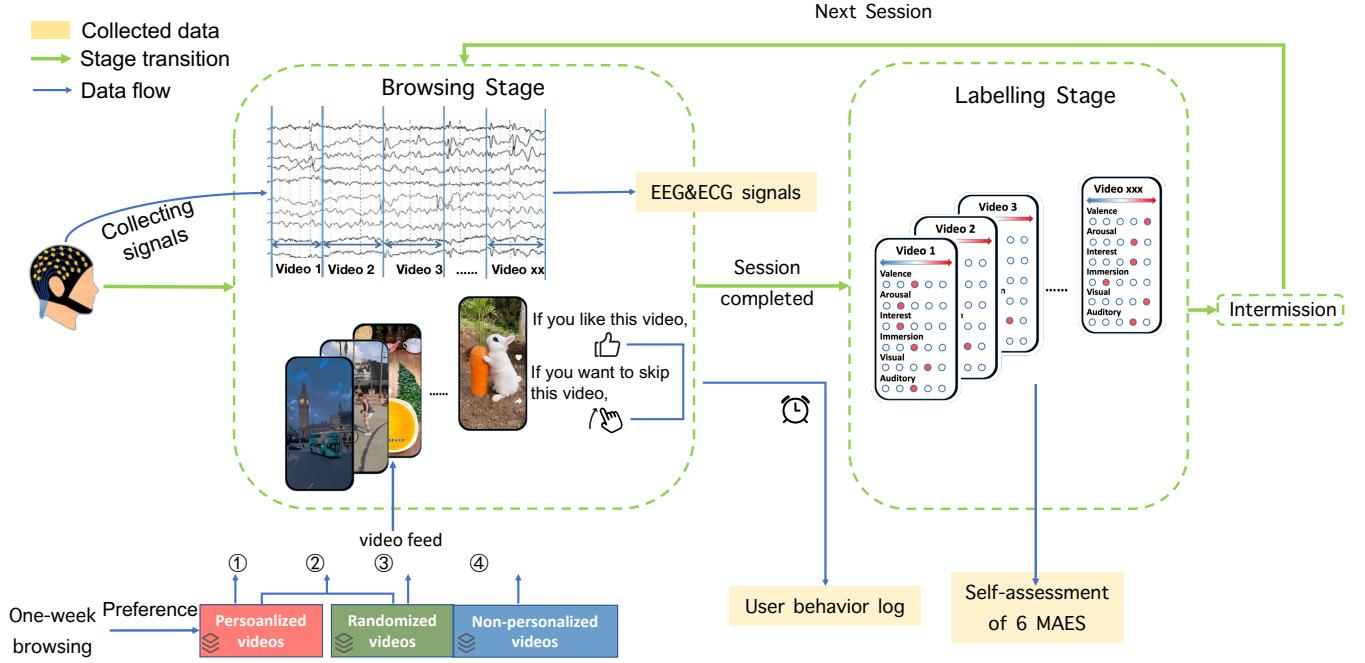


Figure 2: The overall procedure of the lab study for data collection.

channels was calibrated to be under 10 kΩ in the preparation step, and the sampling rate was set at 1,000 Hz.

3.5 Experimental Procedure

Each participant underwent 10-hour preference information collection phase in a week, followed by laboratory experiment phase that included browsing and labelling stages. In the laboratory experiment, participants viewed 4 to 5 sessions of short videos, with each session comprising a 15-minute browsing stage and a roughly 10-minute labelling stage. After completing the video labelling for each session, participants were given a 5-minute rest before proceeding to the next session's browsing stage.

During the **browsing stage**, participants watched sessions of short video sequences distributed from different video pools with each session comprised of 20-30 short videos. Throughout the short video browsing process, participants were allowed to interact with the videos primarily through **liking** and **swiping away** (the video). If participants enjoyed the video they were currently watching, they could click the *like button* at any time during playback. Additionally, if participants did not wish to continue watching the video, they were allowed to *swipe away* anytime. It's noted that the video will be replayed when done without swiping away. Electroencephalogram (EEG) and electrocardiogram (ECG) physiological signals were continuously collected.

After each participant has completed browsing a short video sequence within a specific session, we conducted a video-level multidimensional affective engagement self-assessment **labelling stage**. Participants were given a brief recap of each video chronologically based on their browsing history. Subsequently, they rated each short video on a 5-point Likert scale across six multidimensional affective engagement indicators. the labelling instructions are

given to the participants¹. The six dimensions are valence, arousal, immersion, interest, visual, and auditory. *Valence* represented the positive and negative aspects of emotions, while *Arousal* indicated the intensity of emotions. *Immersion* denoted the degree of the participant's involvement while watching the video, and *Interest* indicated the extent to which the video aligned with the participant's personal interests. *Visual* and *Auditory* scores described the presentation quality of visual elements (e.g., scenery, graphics) and auditory elements, (e.g., voices, music).

Our experiment collected MAES through questionnaires, gathering ratings for videos within each session after its completion. Participants were asked to recall the videos by viewing the first few seconds and to rate them across the six dimensions until they could adequately recall the video. In post-experiment interviews, participants reported that the number of videos per session did not cause memory difficulties, so they could generally recall the browsing history and complete labelling after watching the initial seconds of each video. Having completed the video labelling for each session, participants were given a 5-minute rest before proceeding to the browsing stage in the next session. Thus, the labelling stage generated a corresponding score for each of the six MAES for every short video.

Ultimately, we obtained three types of video-level data: browsing behavior logs, EEG and ECG signals, and multidimensional affective engagement self-assessment labelling.

4 DATASET DESCRIPTION

In this EEG dataset, we ultimately collected 3,657 interactions from 30 users involving 2,636 items (short videos). Due to the different participants watching the same short video in randomized mode, multiple interactions can be associated with the same item. Each

interaction (U-I pair) corresponds to a related EEG and ECG segment. Additionally, each interaction is associated with a behavioral log and a self-assessment of MAES. To further describe the dataset, we introduce it from four aspects: the EEG signals, behavioral, self-assessment data, and characteristics of short videos.

Table 2: The Statistics of Dataset. Each interaction has corresponding MAEs and EEG signals.

	#User	#Item	#Interaction	#EEG datasize
EEG-SVRec	30	2,636	3,657	62GB

4.1 EEG statistics and preprocessing

Here, EEG data are collected through all 3,657 interactions. For each interaction, the size of EEG data is $(Ch, fs \cdot T)$, where fs is the sample rate (1000 Hz), T denotes the recording duration of the interaction, and Ch is the number of electrode channels (62 in total). We preprocess EEG data extract features as follows:

The raw EEG data is subjected to a series of preprocessing steps to eliminate noise and artifacts and enhance the signal quality. The preprocessing pipeline comprises the following stages. First, baseline correction: We first perform baseline correction to remove any constant offsets or drifts in the EEG signals, ensuring that the baseline amplitude is zero. Second, rereferencing: Re-referencing employs the average of M1 and M2 mastoid electrodes as the new reference, minimizing potential bias and improving the signal-to-noise ratio. Third, filtering: Filtering applies a 0.5 Hz to 50 Hz band-pass filter to remove low-frequency drifts (<0.5 Hz) and high-frequency noise (>50 Hz), as well as 50 Hz powerline interference. Last, artifact removal: Artifact removal eliminates abnormal amplitude signals and artifacts induced by eye blinks or head movements.

After the preprocessing steps, we proceed to extract features from the cleaned EEG signals. In this study, we focus on the extraction of differential entropy (DE) as a feature, which has been shown to be useful in characterizing the complexity and information content of EEG signals [6]. Firstly, we estimate power spectral density (noted as $P(f)$) using Welch's method [39] (sampling frequency is 1000) based on sliding window. The window length is two divided by the lower bound of the frequency band. Secondly, we normalized for each band and calculated DE using the following formula:

$$DE = - \int P(f) \log(P(f)) df \quad (1)$$

The frequency bands are delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (25-50 Hz). Finally, for each second of EEG signals, we extract a DE of each electrode and each frequency band.

4.2 User Behavior log and self-assessment of MAES

After integrating the log and label files and corresponding them to the EEG via timestamps, we obtained each subject's interaction

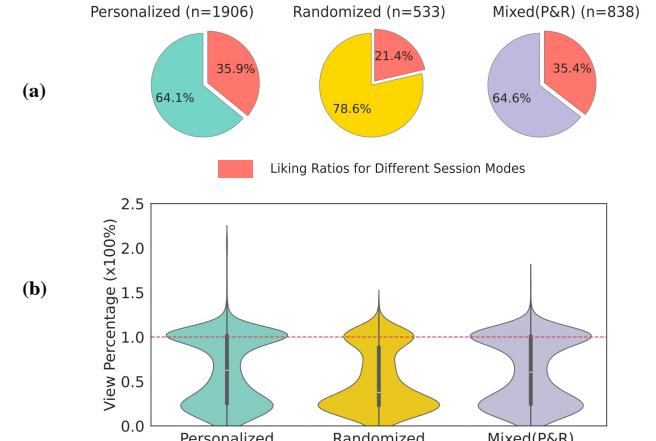


Figure 3: (a) Proportion of likes for short videos: overall and across three session modes (personalized, randomized, and mixed). (b) View percentage distribution across different session modes (View percentage is the viewing duration divided by the video duration. 1.0 represents viewing the video once.)

behavior (liking and viewing duration) and MAES for their video viewing. For each interaction, the UNIX timestamps of browsing are aligned with the start and end time of the corresponding piece of the psychological signals. The video sequence and session mode are also important. Thus, we provide the order of the video in the interaction sequence and session mode (Randomized, Personalized, and Mixed). As for the Mixed mode, we use further distinguish the personalized recommendation video from the random one.

In Figure 3 (a), we present the distribution of the proportion of likes for short videos in both the overall context and across three distinct session modes: personalized, randomized, and mixed (a combination of personalized and randomized). Notably, the like rate in the personalized mode (35.9%) and the mixed mode (35.4%) are relatively similar. In contrast, the like rate in the randomized mode (21.4%) falls below. Same as the like rate, view percentage in personalized mode and mixed mode in Figure 3 (b) is higher than randomized overall. It's surprising that the performance of mixed mode is relatively similar to the personalized mode. Likes and view percentages are presumably influenced by contexts in the session. Focusing on user experience from behavior may shed a little light on recommender systems.

In Figure 4, it can be observed that the distribution of the six MAES exhibits noticeable differences. It can be observed that valence and arousal, commonly utilized as two-dimensional indices in the field of emotion recognition, both exhibit a distinct distribution with 3 being the highest point. Immersion, interest, visual and auditory demonstrate a relatively uniform distribution compared to the former two, indicating a more effective differentiation in representing video content.

4.3 Characteristic of Short Videos

We meticulously extracted comprehensive video features, encompassing both visual and auditory aspects, to further investigate components related to audio and visual ratings.

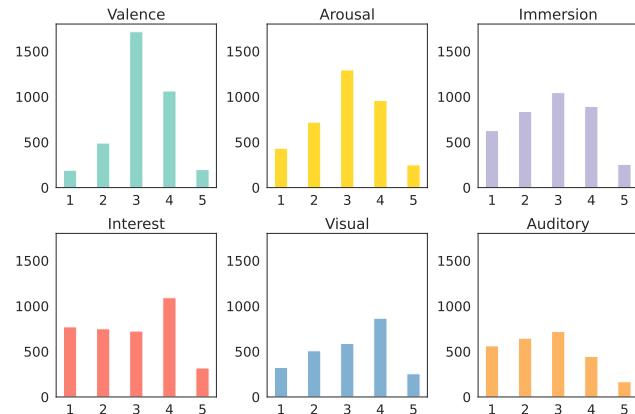


Figure 4: The distribution for six MAES (valence, arousal, immersion, interest, visual and auditory).

For video featurization, we sampled each frame per second and computed an array of features. Specifically, we determined the mean (representing brightness) and standard deviation (representing contrast) by converting each frame to grayscale. Additionally, we assessed hue, saturation, value (in terms of HSV), Laplace variation, and color cast for each frame. Regarding audio, we initially extracted audio signals from the short videos using their native sampling rate. We then employed openSMILE to compute features from the ComParE2016 acoustic feature set [29], maintaining the same sampling rate. Subsequently, we utilized the Audio Spectrogram Transformer [9], trained on AudioSet, to classify audio events with a sampling rate of 16,000 to comply with the classifier. If the event was classified as music, we employed Librosa to detect beats and determine the tempo.

5 EXAMPLE APPLICATIONS

5.1 Impact on User Understanding in Recommendation

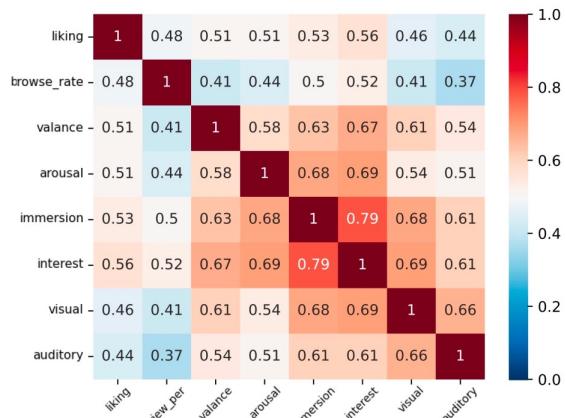


Figure 5: Heatmap presents the correlations of behavior (liking, and view percentage) and MAES (valence, arousal, immersion, interest, visual, and auditory).

5.1.1 Analysis of MAES and Browsing Behavior. Figure 5 presents the correlation between behavioral and MAES attributes. It can be observed that **Liking** has the strongest correlation with Interest (0.56), followed by Immersion (0.53) and Valence (0.51). This suggests that the users' preferences are more closely related to their interest in the content and the degree of immersion they experience while viewing the video, rather than simply the valence or arousal induced by the content. On the other hand, the View Percentage attribute exhibits the highest correlation with Immersion (0.50) and Interest (0.52), indicating that the percentage of browsing is more likely to be influenced by their interest in the content and the level of immersion they experience. This further highlights the importance of considering users' interests and immersion levels when designing recommender systems to improve user engagement and browsing experience. The above findings emphasize the need to consider users' interests and the degree of immersion they experience when designing effective recommendation algorithms. We are expecting more findings to be discovered by the researchers.

5.1.2 The Relation of EEG with MAE and Behaviors. Figure 6 displays the topographical maps illustrating the correlations between EEG signals and the six MAES as well as the two behaviors. These maps reveal distinct correlation patterns for each MAES and behavior. Furthermore, some unique findings emerge, such as the consistent presence of strong correlations between gamma-band electrodes in the frontal lobe area across all six emotion annotations.

Gamma waves are known to play a critical role in numerous brain functions and cognitive processes, including attention, memory, perception, and consciousness [19, 23]. The activation of gamma waves in the frontal lobe suggests the involvement of this region in the associated cognitive processes. As a key area of the brain, the frontal lobe is closely linked to higher cognitive functions, such as decision-making, planning, problem-solving, working memory, and attention control [7]. The observed activation of gamma waves in the frontal lobe may be indicative of the engagement of these higher cognitive functions during the tasks.

5.2 Recommendation in Terms of Various User Feedback Signals

The proposed EEG-SVRec is also feasible for personalized recommendation tasks. Beyond the traditional way of taking *liking* as the user feedback signal, various user feedback signals provided in the dataset can be leveraged as the ground truth. We conduct experiments for item recommendation task while leveraging *liking*, *immersion*, *interest*, *valence*, *arousal*, *visual preference*, and *auditory preference* as user feedback signals, respectively. As an example, we provide the benchmark for item recommendation with and without EEG information. We use the popular recommendation toolkit Recbole [48] for different algorithms, which only support point-wise evaluations for context-aware recommendation models rather than ranking-based evaluation, and report AUC scores but not NDCG performances.

The dataset is split into training set, validation set, and test set by 7:1:2. As for the EEG data, we utilize the 310-dimensional (62 channels * 5 frequency bands) DE (EEG feature described in Section 4.1) corresponding to interactions and project them into an embedding

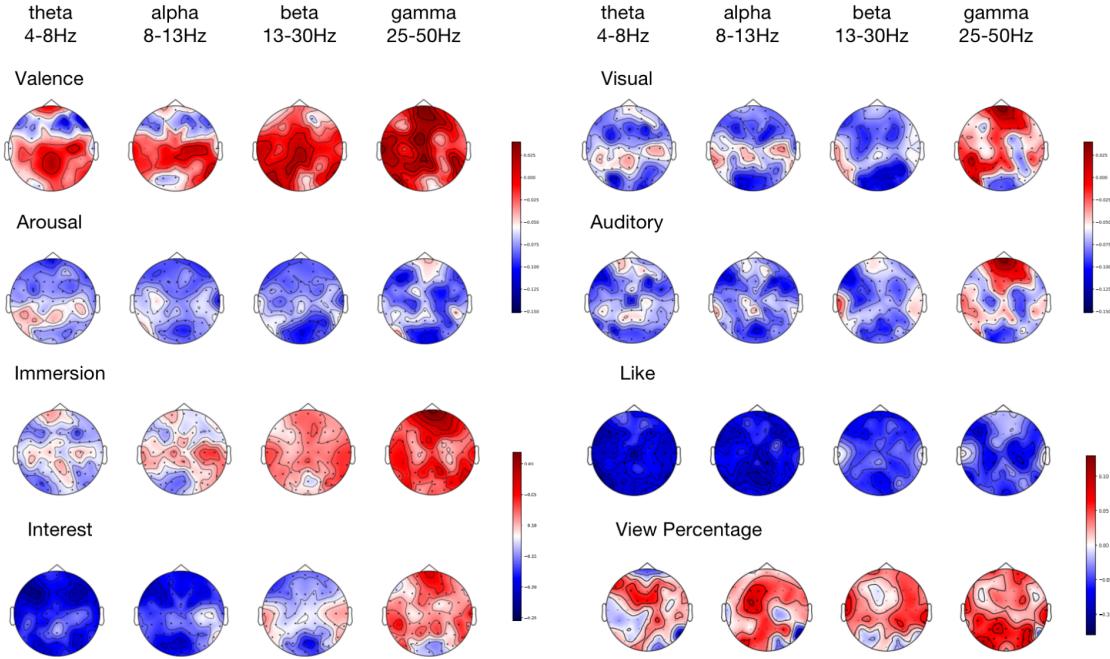


Figure 6: The mean correlations (overall participants) of the MAES (emotion of valence and arousal, immersion, interest, and rating of visual and auditory, ranged 1-5) and behaviors (like and view percentage) with DE in the broad frequency bands of theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-45 Hz). The white circle marks the significant correlation ($p < 0.05$).

Table 3: The recommendation performance (in terms of AUC) that leverages liking, interest, immersion, visual preference, and auditory preference as user feedback respectively. The two-sided t-test is conducted. * indicates p -value < 0.05 . bold shows the higher result of the two settings.

Model	Feature	Like	Immersion	Interest	Valence	Arousal	VisualPref	AudioPref
FM	id	0.7152	0.6776	0.6950	0.6348	0.6917	0.6685	0.6419
	id+EEG	0.7312*	0.6857	0.6933	0.6492	0.6929	0.6690	0.6675*
DeepFM	id	0.7331	0.6869	0.7005	0.6379	0.6930	0.6691	0.6600
	id+EEG	0.7368	0.6927	0.7010	0.6586*	0.7077	0.6711	0.6608
AFM	id	0.7188	0.6774	0.6935	0.6406	0.6962	0.6736	0.6251
	id+EEG	0.7236	0.6955*	0.6910	0.6583*	0.6898	0.6688	0.6578*
WideDeep	id	0.7324	0.7033	0.7027	0.6651	0.7066	0.6718	0.6735
	id+EEG	0.7387	0.7056	0.7121	0.6660	0.7094	0.6978*	0.6767
DCN-V2	id	0.6937	0.6190	0.6698	0.5855	0.6340	0.6443	0.6585
	id+EEG	0.6924	0.6582*	0.6802*	0.6249*	0.6715*	0.6585	0.6440

through a fully connected layer. We tune hyperparameters and choose the best result for each setting (id and id+EEG).

From the Table 3, it is observed that in most instances, models incorporating EEG signals achieve superior results, suggesting the general potential of EEG signals in recommendation tasks. It is worth noting that only simple way of introducing EEG information is implemented in the benchmark experiments, which directly

embeds EEG signals as features, and has already effectively enhanced recommendation performance. This verifies that EEG contains additional valuable information. Thus, leveraging EEG signals presumably assists recommender systems in better understanding user multidimensional affective engagement and behaviors, thereby providing better personalized recommendations.

EEG reflects the cognitive activity of viewing short videos, which can be used as auxiliary information to enhance representations. Thus, a natural idea is to enhance user and item embeddings with their corresponding EEG signals. The idea is widely used in existing recommendation models, such as review-based [3, 34], social-based [40], knowledge graph-based [4], and visual-based [12] models. However, EEG directly reflects the user's brain activities, which can bring more in-depth user understanding beyond the above auxiliary information. This opens a novel avenue to enrich the representation of items and further help the recommender systems understand the users with incognizable, subject, and direct feedback with cognitive information. The comparison of EEG data and other information, as well as more sophisticated recommendation models, are left as future work.

6 DISCUSSIONS AND LIMITATIONS

6.1 Possible Research Directions

In this section, we discuss the potential applications of the EEG-SVRec dataset in various aspects of short video recommendation systems and beyond.

(1) **Human-centric Evaluation Metrics:** The dataset offers a more human-centric perspective on evaluation metrics, going beyond traditional measures such as dwell time and likes. It enables researchers to assess recommender systems based on their ability to enhance users' overall experience, considering multidimensional aspects of user engagement, rather than merely maximizing utility metrics.

(2) **Uncovering the Relationship Between User Behavior and Cognitive Activity:** Utilizing the dataset to study user behavior and cognitive activities during the recommendation process can reveal insights into how brain activity can inform adjustments in recommendations. This knowledge potentially helps reduce information echo chambers and enhance content diversity, leading to a more balanced and varied user experience.

(3) **EEG-guided Recommendation Algorithms:** The EEG-SVRec dataset opens up opportunities to explore the development of EEG-guided recommendation algorithms that incorporate EEG signals for a deeper understanding of user preferences and behavior. By leveraging a smaller labeled EEG dataset alongside a larger unlabeled dataset, algorithms can potentially learn more accurate and personalized recommendations by generalizing the knowledge gained from EEG signals across a broader user base. Furthermore, EEG reflects the cognitive activity of viewing short videos which can be used as auxiliary information to enhance representation.

(4) **Accessibility for Users with Disabilities in Short Video Streaming:** The EEG-SVRec dataset has the potential to facilitate the development of more inclusive recommendation systems tailored for individuals with disabilities. By analyzing the unique cognitive and emotional experiences of these users through EEG data, algorithms can be adapted to better cater to their needs and preferences, ultimately improving their experience with short video recommendations.

In summary, the EEG-SVRec dataset presents an array of potential applications that can contribute to the development of more effective, personalized, and inclusive recommendation algorithms. By focusing on a more human-centric approach and leveraging the

rich information provided by EEG signals, researchers and practitioners can drive innovation in the field of recommender systems and enhance user experiences across various contexts.

6.2 Limitations

In this study, we present the EEG-SVRec dataset. Despite its potential value for the recommender systems community, there are some limitations that should be considered:

(1) **Sample Size:** The dataset was constructed with a scale of 30 participants from the university, which may not fully capture the diversity of users on social media platforms. Although the sample size might seem limited, it is important to note that the high cost associated with EEG data collection can hinder the ability to gather larger sample sizes. Many published EEG datasets are with the same scale of participants universities [17, 28, 49].

(2) **Generalizability:** EEG's applicability in large-scale real-world scenarios could be challenging due to the required equipment and expertise. Meanwhile, personalized and randomized videos are 30-60s, which may differ from general contexts. The reason to choose 30-60s refers to Section 3.3. Despite this, investigating the temporal dynamics of user behavior and emotions in various recommendation settings would be a valuable direction for future research.

(3) **Algorithmic bias:** The EEG-SVRec dataset might contain biases from the underlying recommendation algorithms from the platform, which could impact the generalizability of the findings. However, we provide the interaction with randomized video as unbiased data for this purpose. It is essential for future research to identify and address any potential biases present in the dataset.

Despite these limitations, the EEG-SVRec dataset provides a valuable resource for exploring user behavior and emotions in short video recommendations and can inspire further research in this area.

7 CONCLUSION AND FUTURE WORK

This paper introduces EEG-SVRec, a novel dataset including EEG and ECG signals, multidimensional affective engagement annotations, and user behavior data for short video recommendation. This dataset bridges a critical gap by providing insights into user intrinsic experience and behavior in real-world short video scenarios. Our key contributions include proposing the first EEG dataset in short video streaming scenario, collecting multidimensional affective engagement scores, and providing both implicit and explicit user feedback. We carried out a rigorous experimental process for 30 participants and obtained a dataset, which is highly versatile and applicable to various research problems. We establish benchmarks for rating prediction by including EEG signals and prevalent recommendation algorithms. Experimental results demonstrate the usefulness of EEG signals in recommendation scenarios. It is worth noting that our current application of EEG signals is primary, leaving room for future improvements.

For future work, it is expected that more sophisticated models, such as DGCNN [31], could be employed to utilize electrode position information from the EEG signals and further improve recommendation performance on the EEG-SVRec dataset. By leveraging more advanced techniques, deeper insights into the role that

EEG signals play in short video recommendation systems could be uncovered. Furthermore, the application of EEG and ECG signals could be expanded to a broader range of research areas, such as developing more affective-centric evaluation metrics and applications for individuals with disabilities. Lastly, the dataset holds significant societal value in further exploring the occurrence and changes in user emotions and cognitive behavior within short video recommendation scenarios. We anticipate that our work will inspire further exploration and innovation in the field of recommendation and encourage researchers to delve into these potential applications.

REFERENCES

- [1] Ricardo Baeza-Yates. 2018. Bias on the web. *Commun. ACM* 61, 6 (2018), 54–61.
- [2] Marco Bressan, Stefano Leucci, Alessandro Panconesi, Prabhakar Raghavan, and Eriko Teroli. 2016. The limits of popularity-based recommendations, and the role of social ties. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 745–754.
- [3] Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural attentional rating regression with review-level explanations. In *Proceedings of the 2018 world wide web conference*. 1583–1592.
- [4] Zhongxia Chen, Xiting Wang, Xing Xie, Tong Wu, Guoqing Bu, Yining Wang, and Enhong Chen. 2019. Co-attentive multi-task learning for explainable recommendation. In *IJCAI*. 2137–2143.
- [5] Keith M Davis III, Michiel Spapé, and Tuukka Ruotsalo. 2021. Collaborative filtering with preferences inferred from brain signals. In *Proceedings of the Web Conference 2021*. 602–611.
- [6] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. 2013. Differential entropy feature for EEG-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 81–84.
- [7] Joaquín M Fuster. 2002. Frontal lobe and cognitive development. *Journal of neurocytology* 31, 3–5 (2002), 373–385.
- [8] Chongming Gao, Shijun Li, Yuan Zhang, Jiawei Chen, Biao Li, Wenqiang Lei, Peng Jiang, and Xiangnan He. 2022. KuaRand: An Unbiased Sequential Recommendation Dataset with Randomly Exposed Videos. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 3953–3957.
- [9] Yuan Gong, Yu-An Chung, and James Glass. 2021. Ast: Audio spectrogram transformer. *arXiv preprint arXiv:2104.01778* (2021).
- [10] Liang Gou, Fang You, Jun Guo, Luigi Wu, and Xiaolong Zhang. 2011. Sfviz: interest-based friends exploration and recommendation in social networks. In *Proceedings of the 2011 Visual Information Communication-International Symposium*. 1–10.
- [11] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.
- [12] Ruining He and Julian McAuley. 2016. VBPR: visual bayesian personalized ranking from implicit feedback. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 30.
- [13] Zhiyu He, Shaorun Zhang, Peijie Sun, Jiayu Li, Xiaohui Xie, Min Zhang, and Yiqun Liu. 2023. Understanding User Immersion in Online Short Video Interaction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 731–740.
- [14] Richard W Homan, John Herman, and Phillip Purdy. 1987. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology* 66, 4 (1987), 376–382.
- [15] Xin Hu, Jingjing Chen, Fei Wang, and Dan Zhang. 2019. Ten challenges for EEG-based affective computing. *Brain Science Advances* 5, 1 (2019), 1–20.
- [16] Dietmar Jannach, Lukas Lerche, and Markus Zanker. 2018. Recommending based on implicit feedback. In *Social Information Access: Systems and Technologies*. Springer, 510–569.
- [17] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* 3, 1 (2011), 18–31.
- [18] Jiayu Li, Zhiyu He, Yunmeng Cui, Chenyang Wang, Chong Chen, Chun Yu, Min Zhang, Yiqun Liu, and Shaoping Ma. 2022. Towards Ubiquitous Personalized Music Recommendation with Smart Bracelets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–34.
- [19] Mu Li and Bao-Liang Lu. 2009. Emotion classification based on gamma-band EEG. In *2009 Annual International Conference of the IEEE Engineering in medicine and biology society*. IEEE, 1223–1226.
- [20] Xiang Li, Yazhou Zhang, Prayag Tiwari, Dawei Song, Bin Hu, Meihong Yang, Zhigang Zhao, Neeraj Kumar, and Pekka Marttinen. 2022. EEG based emotion recognition: A tutorial and review. *Comput. Surveys* 55, 4 (2022), 1–57.
- [21] Yiyu Liu, Qian Liu, Yu Tian, Changping Wang, Yanan Niu, Yang Song, and Chenliang Li. 2021. Concept-aware denoising graph neural network for micro-video recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 1099–1108.
- [22] Hongyu Lu, Min Zhang, Weizhi Ma, Yunqiu Shao, Yiqun Liu, and Shaoping Ma. 2019. Quality effects on user preferences and behaviors in mobile news streaming. In *The World Wide Web Conference*. 1187–1197.
- [23] Kimford J Meador, Patty G Ray, Javier R Echauz, David W Loring, and George J Vachtsevanos. 2002. Gamma coherence and conscious perception. *Neurology* 59, 6 (2002), 847–854.
- [24] Juan Abdon Miranda-Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. 2018. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing* 12, 2 (2018), 479–493.
- [25] Yashar Moshfeghi, Peter Triantafillou, and Frank E Pollick. 2016. Understanding information need: An fMRI study. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 335–344.
- [26] Yongxin Ni, Yu Cheng, Xiangyan Liu, Junchen Fu, Youhua Li, Xiangnan He, Yongfeng Zhang, and Fajie Yuan. 2023. A content-driven micro-video recommendation dataset at scale. *arXiv preprint arXiv:2309.15379* (2023).
- [27] Mamunur Rashid, Norizam Sulaiman, Anwar PP Abdul Majeed, Rabiu Muazu Musa, Bifta Sama Bari, Sabira Khatoon, et al. 2020. Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review. *Frontiers in neurorobotics* (2020), 25.
- [28] Arman Savran, Koray Ciftci, Guillaume Chanel, Javier Mota, Luong Hong Viet, Blent Sankur, Lale Akarun, Alice Caplier, and Michele Rombart. 2006. Emotion detection in the loop from brain signals and facial images. In *Proceedings of the INTERFACE 2006 Workshop*. Citeseer.
- [29] Björn Schuller, Stefan Steidl, Anton Batliner, Julia Hirschberg, Judee K Burgoon, Alice Baird, Aaron Elkins, Yue Zhang, Eduardo Coutinho, and Keelan Evanini. 2016. The interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language. In *17TH Annual Conference of the International Speech Communication Association (Interspeech 2016)*, Vols 1–5, Vol. 8. ISCA, 2001–2005.
- [30] Guy Shani and Asela Gunawardana. 2011. Evaluating recommendation systems. *Recommender systems handbook* (2011), 257–297.
- [31] Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. 2018. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing* 11, 3 (2018), 532–541.
- [32] Sebastian Stober, Avital Sternin, Adrian M Owen, and Jessica A Grahn. 2015. Towards Music Imagery Information Retrieval: Introducing the OpenMIIR Dataset of EEG Recordings from Music Perception and Imagination. In *ISMIR*. 763–769.
- [33] Peijie Sun, Le Wu, Kun Zhang, Xiangzhi Chen, and Meng Wang. 2023. Neighborhood-Enhanced Supervised Contrastive Learning for Collaborative Filtering. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [34] Peijie Sun, Le Wu, Kun Zhang, Yanjie Fu, Richang Hong, and Meng Wang. 2020. Dual learning for explainable recommendation: Towards unifying user preference prediction and review generation. In *Proceedings of The Web Conference 2020*. 837–847.
- [35] Michal Teplan et al. 2002. Fundamentals of EEG measurement. *Measurement science review* 2, 2 (2002), 1–11.
- [36] Robert E Thayer. 1990. *The biopsychology of mood and arousal*. Oxford University Press.
- [37] Wei Wei, Chao Huang, Lianghao Xia, and Chuxu Zhang. 2023. Multi-Modal Self-Supervised Learning for Recommendation. *arXiv preprint arXiv:2302.10632* (2023).
- [38] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. 2019. MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In *Proceedings of the 27th ACM international conference on multimedia*. 1437–1445.
- [39] Peter Welch. 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* 15, 2 (1967), 70–73.
- [40] Le Wu, Peijie Sun, Yanjie Fu, Richang Hong, Xiting Wang, and Meng Wang. 2019. A neural influence diffusion model for social recommendation. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*. 235–244.
- [41] Mahendra Yadava, Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, and Debi Prosad Dogra. 2017. Analysis of EEG signals and its application to neuromarketing. *Multimedia Tools and Applications* 76 (2017), 19087–19111.
- [42] JungAe Yang. 2016. Effects of popularity-based news recommendations (“most-viewed”) on users’ exposure to online news. *Media Psychology* 19, 2 (2016), 243–271.
- [43] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 325–334.
- [44] Ziyi Ye, Xiaohui Xie, Yiqun Liu, Zhihong Wang, Xuesong Chen, Min Zhang, and Shaoping Ma. 2022. Towards a Better Understanding of Human Reading Comprehension with Brain Signals. In *Proceedings of the ACM Web Conference 2022*. 380–391.
- [45] Ziyi Ye, Xiaohui Xie, Yiqun Liu, Zhihong Wang, Xuancheng Li, Jiaji Li, Xuesong Chen, Min Zhang, and Shaoping Ma. 2022. Why Don’t You Click: Understanding Non-Click Results in Web Search with Brain Signals. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 633–645.
- [46] Guanghu Yuan, Fajie Yuan, Yudong Li, Beibei Kong, Shujin Li, Lei Chen, Min Yang, Chenyun Yu, Bo Hu, Zang Li, et al. 2022. Tenrec: A Large-scale Multipurpose Benchmark Dataset for Recommender Systems. *arXiv preprint arXiv:2210.10629* (2022).
- [47] Min Zhang and Yiqun Liu. 2021. A commentary of TikTok recommendation algorithms in MIT Technology Review 2021. *Fundamental Research* 1, 6 (2021), 846–847.

- [48] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, et al. 2021. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4653–4664.
- [49] Wei-Long Zheng and Bao-Liang Lu. 2015. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development* 7, 3 (2015), 162–175.

EEG-ImageNet: An Electroencephalogram Dataset and Benchmarks with Image Visual Stimuli of Multi-Granularity Labels

Shuqi Zhu

DCST, Tsinghua University
Beijing
zsq19991106@gmail.com

Ziyi Ye

DCST, Tsinghua University
Beijing
yeziyi1998@gmail.com

Qingyao Ai

DCST, Tsinghua University
Beijing
aiqy@tsinghua.edu.cn

Yiqun Liu

DCST, Tsinghua University
Beijing
yiqunliu@tsinghua.edu.cn

Abstract

Identifying and reconstructing what we see from brain activity gives us a special insight into investigating how the biological visual system represents the world. While recent efforts have achieved high-performance image classification and high-quality image reconstruction from brain signals collected by Functional Magnetic Resonance Imaging (fMRI) or magnetoencephalogram (MEG), the expensiveness and bulkiness of these devices make relevant applications difficult to generalize to practical applications. On the other hand, Electroencephalography (EEG), despite its advantages of ease of use, cost-efficiency, high temporal resolution, and non-invasive nature, has not been fully explored in relevant studies due to the lack of comprehensive datasets. To address this gap, we introduce *EEG-ImageNet*, a novel EEG dataset comprising recordings from 16 subjects exposed to 4000 images selected from the ImageNet dataset. EEG-ImageNet consists of 5 times EEG-image pairs larger than existing similar EEG benchmarks. EEG-ImageNet is collected with image stimuli of multi-granularity labels, i.e., 40 images with coarse-grained labels and 40 with fine-grained labels. Based on it, we establish benchmarks for object classification and image reconstruction. Experiments with several commonly used models show that the best models can achieve object classification with accuracy around 60% and image reconstruction with two-way identification around 64%. These results demonstrate the dataset’s potential to advance EEG-based visual brain-computer interfaces, understand the visual perception of biological systems, and provide potential applications in improving machine visual models.

1 Introduction

Recent advancements in reconstructing visual experiences from the human brain have seen significant progress, largely driven by the extensive use of functional magnetic resonance imaging (fMRI) ([8, 22, 23]) and magnetoencephalogram (MEG) [3] datasets. fMRI and MEG are widely used to investigate various cognitive functions, neurological disorders, and brain connectivity patterns ([2, 40, 37, 35]). Driven by the use of deep neural networks, particularly diffusion-based and transformer-based models ([34, 30, 25, 6]), it is even possible to reconstruct human’s visual perceptions from fMRI or MEG recordings. The application of large-scale deep neural networks in neuroscience research has

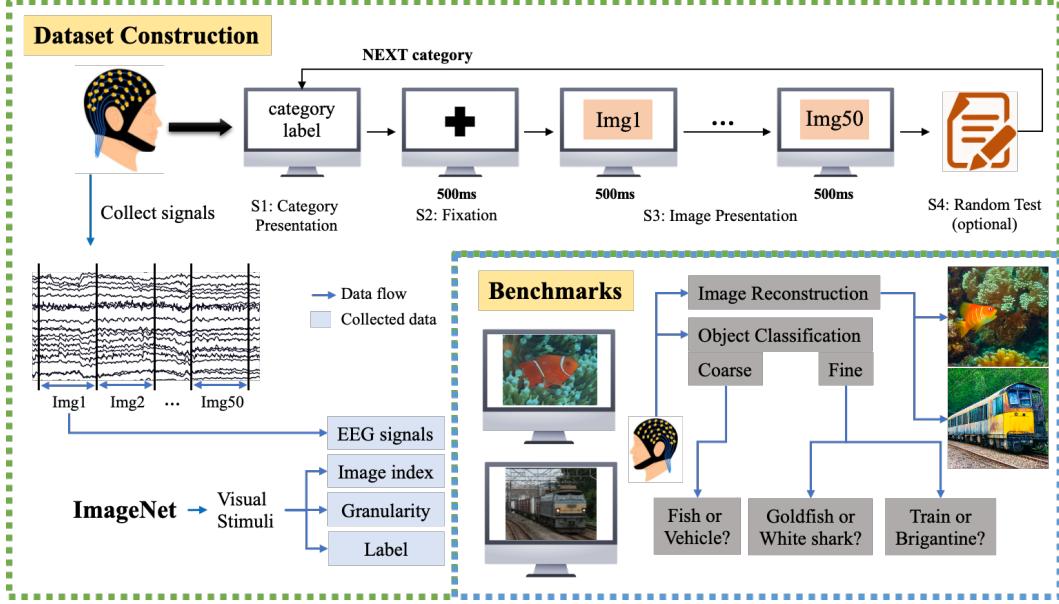


Figure 1: The overall procedure of our dataset construction and benchmark design.

further underscored the importance of large-scale and high-quality datasets [27, 20]. For example, the Natural Scenes Dataset ([1]), contains up to hundreds of thousands of high-quality natural image-fMRI pairs of 8 subjects, providing a solid data foundation for recent work in visual neuroscience. These models and large-scale datasets, in turn, have opened new avenues for understanding the brain’s intricate functions and for developing advanced applications in brain-computer interfaces, neuroimaging, and beyond [33].

On the other hand, electroencephalography (EEG) is another vital tool in neuroscience research. In comparison with fMRI and MEG, EEG is easy to use, cost-efficient, and has superior temporal resolution, making it a valuable tool for capturing rapid and real-time brain dynamics on the order of milliseconds [36]. EEG signals can be obtained non-invasively by placing electrodes on the scalp, making it a less intrusive method for monitoring brain activity. These attributes position EEG as another promising modality for visual neuroscience research.

Although visual reconstruction has been achieved using fMRI and MEG, the high cost and inconvenience of these devices limit their widespread application in practical settings. In contrast, EEG presents advantages over both fMRI and MEG with its cost-efficient and portable features. However, studies on visual perception with EEG signals are limited because of two challenges: (1) the lack of large-scale, high-quality EEG datasets and (2) existing EEG datasets typically featured coarse-grained image categories, lacking fine-grained categories. To the best of our knowledge, the most frequently used dataset is the data set provided by Spampinato et al. [32], which involves 6 participants each watching 2000 image stimuli. However, this dataset’s scale is smaller than existing fMRI datasets, limiting the possibilities for investigating neural aspects related to visual perception and developing deep learning models for relevant visual classification and reconstruction tasks.

On the other hand, the labels in existing EEG datasets are frequently coarse and lack the granularity needed for detailed analysis. Multi-granularity labels are essential because they allow for a more nuanced analysis at different levels of detail. For instance, labels can range from broad categories like “panda” or “golf ball” to more specific attributes like “Rottweiler” or “Samoyed”. These challenges underscore the necessity for new, large-scale EEG datasets with high-quality, multi-granularity labels. Such datasets would enable researchers to explore the intricacies of visual processing with greater accuracy and depth, facilitating advancements in both basic neuroscience and applied fields like brain-computer interfaces and clinical diagnostics.

To address these challenges, we present *EEG-ImageNet*, a novel EEG dataset specifically designed to promote research related to visual neuroscience, biomedical engineering, etc. *EEG-ImageNet* is a comprehensive dataset that includes EEG recordings from 16 subjects, each exposed to 4,000 images

sourced from the ImageNet-21k [28]. These images span 80 different categories, with 50 images per category. The dataset is structured to support multi-granularity analysis, with 40 categories dedicated to coarse-grained tasks and 40 to fine-grained tasks. This dataset aims to bridge the gap in the current research landscape by providing a resource that leverages the strengths of EEG and supports diverse research needs. To demonstrate the utility of this dataset, we establish benchmarks for two primary tasks: object classification and image reconstruction. For the object classification task, we evaluated the dataset using several commonly used models, achieving a best accuracy of 60.88% on the 80-class classification. In the image reconstruction task, our experiments with advanced generative models produced promising results, with the best model achieving a two-way identification of 64.67%. These benchmarks demonstrate the dataset’s potential for advancing EEG-based research. Figure 1 shows the overall procedure of our dataset construction and benchmark design. By addressing the lack of large-scale, high-quality EEG datasets, EEG-ImageNet aims to drive EEG-based visual research forward, improve machine learning models, and provide deeper insights into visual perception and processing.

2 Related Work

In this section, we review some datasets related to visual recognition and neuroscience and compare them with EEG-ImageNet, as shown in Table 1.

2.1 Visual Recognition Dataset

Visual recognition is a cornerstone of computer vision, driven by datasets like ImageNet [28], CIFAR [16], and MS COCO [20]. ImageNet contains over 14 million annotated images, CIFAR-10/100 consist of thousands of 32x32 images in 10 and 100 classes, respectively, and MS COCO is known for its rich annotations supporting tasks such as object detection and segmentation. Later works have matched image datasets with additional modalities.

Efforts to combine visual recognition with neuroscience have led to datasets like the Microsoft COCO Captions [5], which adds textual descriptions to the MS COCO images, enhancing the dataset with a multimodal aspect for evaluating image captioning models. The SALICON dataset [10] extends MS COCO with eye-tracking data, enabling the study of visual attention and saliency through large-scale annotations. Neuroscience datasets utilizing fMRI have further enriched this field. The BOLD5000 dataset [4] includes fMRI data from subjects viewing 5000 images, aiding the exploration of visual perception. The Generic Object Decoding dataset [9] captures brain activity while subjects view and imagine objects, facilitating the decoding of mental images. The VIM-1 dataset [14] from the study “Identifying Natural Images from Human Brain Activity” demonstrates the feasibility of decoding viewed images from brain activity. Additionally, the NSD [1] is a large-scale fMRI dataset in visual neuroscience, recording high-resolution (1.8-mm) whole-brain 7T fMRI data from eight subjects exposed to 9,000–10,000 color natural scenes from the MS COCO dataset over the course of a year. Integrating neuroimaging and eye-tracking datasets with visual recognition tasks has opened new avenues for understanding how the brain interprets visual information, leading to insights and applications in brain-computer interfaces and neural decoding.

2.2 EEG dataset

fMRI is renowned for its high spatial resolution, allowing researchers to obtain detailed images of brain activity by measuring changes in blood flow [23]. In contrast, EEG offers several distinct advantages over fMRI. EEG is relatively easy to use and cost-efficient, with a straightforward setup that involves placing electrodes on the scalp to measure electrical activity. One of the most significant benefits of EEG is its exceptional temporal resolution, which captures neural dynamics on the order of milliseconds [36]. This high temporal resolution makes EEG ideal for studying fast-occurring brain processes and real-time neural responses, providing insights into the timing and sequence of neural events [21]. EEG’s non-invasive nature also makes it suitable for a wider range of participants. However, EEG signals collected using portable devices often have a low signal-to-noise ratio, which can complicate data analysis and reduce the accuracy of the results [13].

Existing EEG datasets span a variety of research areas. The SEED [43] focuses on emotion recognition with detailed EEG recordings from subjects exposed to various emotional stimuli. The BCI

Table 1: Detailed metadata for various neurological datasets based on visual stimuli.

Dataset	#Subjects	Modalities	Visual Stimuli	#Stimuli	#Stimuli per Subject
SALICON [10]	-	Eye-tracking	MS COCO [20]	20,000	-
BOLD5000 [4]	4	fMRI	ImageNet [28], MS COCO, SceneUN [38]	5,254	1,157–1,798
GOD [9]	5	fMRI	ImageNet	1,200	1,200
VIM-1 [14]	2	fMRI	Corel Stock Photo Library [11]	1,870	1,870
NSD [1]	8	fMRI, Eye-tracking	MS COCO	9,000– 10,000	22,000–30,000
Spampinato et al. [32]	6	EEG	ImageNet	2000	2000
EEG-SVRec [42]	30	EEG, ECG	short videos	2636	121.9
EEG-ImageNet	16	EEG	ImageNet	4,000	4,000

Note: The original SALICON dataset lacks information on the number of subjects involved.

Competition IV datasets [41] provide EEG data for motor imagery tasks, while the TUH EEG Corpus [31] is a large clinical EEG collection often used for benchmarking EEG data quality across different conditions. The DEAP [15] collects EEG and peripheral physiological signals from 32 participants as they watch 40 one-minute music videos, providing comprehensive emotional responses. Similarly, the AMIGOS [24] captures EEG and physiological responses from participants watching short video clips designed to evoke specific emotional states. In the realm of visual recognition, datasets like the EEG-Classification dataset [32] involve 6 subjects viewing 2,000 images across 40 object classes from the ImageNet10k. Another significant dataset is the EEG-SVRec [42], which includes EEG recordings from 30 participants interacting with short videos, aiming to capture detailed affective experiences.

In comparison, the EEG-ImageNet dataset offers several advantages. Firstly, its larger scale is conducive to training deep learning models. Secondly, the greater number of subjects facilitates inter-subject experiments and analyses. Additionally, the dataset features multi-granularity image labels, and the images are of high quality, selected from ImageNet21k and filtered to exclude small, blurry, or watermarked pictures.

3 Dataset Construction

During the data collection process of our user study, participants are presented with a visual stimuli dataset containing 4000 natural images from ImageNet. Throughout this process, we continuously record their EEG signals. The whole experimental process is carried out in the laboratory environment. This section describes the entire process of EEG-ImageNet dataset construction. The EEG-ImageNet dataset can be accessed openly through the url <https://github.com/Promise-Z5Q2SQ/EEG-ImageNet-Dataset>.

3.1 Ethical and Privacy

To protect participants' privacy and physical health, our user study adheres to strict ethical guidelines for human research, with approval from the ethics committee of the School of Psychology at Tsinghua University. The study has undergone a comprehensive ethical review to safeguard participants' rights. In accordance with ethical standards, we have taken several steps to protect participants' privacy, including data anonymization and obtaining informed consent from all participants. Additionally, participants are thoroughly informed about the study's objectives, procedures, and potential outcomes. The EEG data collection method employed in this research is non-invasive and poses no risk to participants. This approach ensures compliance with ethical standards while maintaining the integrity of the research findings.

3.2 Participants

We enlist a total of 16 participants via social media, including 10 males and 6 females. These participants are all college students aged between 21 and 27, with an average age of 24.06 and a standard deviation of 1.69. Their majors encompass computer science, mechanical engineering, chemistry, and environmental engineering, and they range from undergraduate to postgraduate levels. All participants are right-handed and assert their proficiency in utilizing image search engines in their daily routines. Each participant dedicates approximately 2 hours to complete the experiment, which includes 30 minutes for equipment setup and task instructions. Before the experiment, participants are informed of a compensation of US\$11.8 per hour upon completion, to ensure the quality of the data collected for the study.

3.3 Stimuli Dataset

The dataset used for visual stimuli was a subset of ImageNet21k, containing 80 categories of objects. Each category comprises 50 manually curated images, ensuring that each image has a width and height greater than 300 pixels and prominently features an object corresponding to its class label in ImageNet. Additionally, every image is free of watermarks. In this manner, we have selected a total of 4000 high-quality natural images as our visual stimulus dataset.

Among all categories, the first half is consistent with the EEG-Classification dataset ([32]), comprising 40 significantly distinct categories from ImageNet1k. We treat these as *coarse-grained* tasks. The latter 40 categories are designed as a *fine-grained* task, divided into 5 groups with 8 categories each. The categories within the same group share the same parent node in WordNet, and each category label is either a leaf node or a sub-leaf node in WordNet. This selection ensures that the chosen categories represent similar granularity while avoiding overly obscure categories, thereby minimizing potential biases in the experimental results. For instance, coarse-grained categories include items such as African elephants, pandas, mobile phones, golf balls, bananas, and pizzas. Under the parent node "musical instruments," the fine-grained categories include accordions, cellos, flutes, oboes, snare drums, and trombones. Detailed information about all the visual stimuli categories and their respective WordNet IDs can be found in our GitHub repository.

3.4 Procedure

Before engaging in the user study, participants are required to fill out an entry questionnaire and sign a consent about the protection of privacy security. They will receive an orientation regarding the primary tasks and operational procedures. Additionally, they will be notified of their right to withdraw from the study at any point. Before the main trials, participants will undergo a series of training trials designed to acquaint participants with the procedures of the formal experiments.

Every participant is required to select a random seed before the experiment to randomize the order of the categories. This randomization guarantees a fair distribution of categories and images among participants. The experimental platform follows a sequential and repetitive process as illustrated in Figure 1. (S1) The experimental platform presents the current category label. Participants can proceed to the next stage by pressing the space key. (S2) A fixation cross is shown at the center of the screen, ensuring attention is drawn when images are displayed. This fixation period lasts for 500 ms. (S3) The 50 images of this category are sequentially presented using the Rapid Serial Visual Presentation (RSVP) paradigm, which is commonly employed in psychological experiments. Each image is presented for a duration of 500 ms [12]. (S4) Random tests are conducted to verify the participant's engagement in the experiment after the presentation. Data from categories for which participants fail the test will not be included in final analyses. The EEG signals of the participant will be captured and recorded continuously during the entire process. The program will cycle back to step S1 and display the next category, repeating this process until all the images have been presented.

3.5 Dataset Description

The EEG-ImageNet dataset contains a total of 63,850 EEG-image pairs from 16 participants. Each EEG data sample has a size of $(n_{\text{channels}}, f_s \cdot T)$, where n_{channels} is the number of EEG electrodes, which is 62 in our dataset; f_s is the sampling frequency of the device, which is 1000 Hz in our dataset; and T is the time window size, which in our dataset is the duration of the image stimulus

presentation, i.e., 0.5 seconds. Due to ImageNet’s copyright restrictions, our dataset only provides the file index of each image in ImageNet and the wnid of its category corresponding to each EEG segment. Additional information about the dataset is shown in Appendix A.1.

4 Benchmarks

In this section, we detail the benchmarks of our study shown in Figure 1 by outlining the preprocessing steps, feature extraction methods, task definitions, and models used.

4.1 Preprocessing

We perform a series of preprocessing steps for the raw EEG data we collect to eliminate noise and artifacts and improve signal quality. The preprocessing pipeline includes the following stages: First, re-referencing: Re-referencing is done using the offline linked mastoids method, which uses the average of the M1 and M2 mastoid electrodes as the new reference point [39]. This minimizes potential biases and improves the signal-to-noise ratio. Then, filtering: Filtering is performed using a 0.5 Hz to 80 Hz band-pass filter to remove low-frequency drifts (<0.5 Hz) and high-frequency noise (>80 Hz). Additionally, 50 Hz environmental noise is eliminated. Finally, artifact removal: Artifact removal eliminates abnormal amplitude signals and artifacts caused by blinks or head movements.

4.2 Feature Extraction

In our benchmarks, for models requiring time-domain signals as direct input, we extract the 40ms-440ms segment of each EEG signal as the feature input. This approach helps to minimize the influence of preceding and subsequent image stimuli on the current stimulus. For models requiring frequency-domain features as input, we extract the differential entropy (DE) of the extracted time-domain signals as features, as this characteristic effectively captures the complexity and variability of brain activity in the frequency domain [7]. According to the general division in neuroscience, the frequency bands are categorized as delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-80 Hz). We use the Welch method with a sliding window to estimate the power spectral density $P(f)$ in each frequency band. Then, we normalize the data and calculate the differential entropy (DE) using the formula, $DE = -\int P(f) \log(P(f))df$. Consequently, for each segment of EEG signals, we obtain the differential entropy (DE) for each electrode and each frequency band.

4.3 Task Definition

In our benchmarks, we test our dataset on two tasks: *object classification* and *image reconstruction*. The object classification task aims to classify the category of the corresponding image stimulus the participant is exposed to with their EEG signals. We evaluate the models using classification accuracy. In the image reconstruction task, given a specific EEG segment, the goal is to reconstruct the image stimulus the participant is exposed to. We evaluate the generated results using two-way identification [30] under different visual neural networks (refer to Appendix A.3 for details.).

On our multi-granularity labeled image dataset, we test various tasks with different levels of granularity. Additionally, due to the inherent significant inter-individual differences in EEG signals, all models in our benchmarks are trained exclusively in an intra-subject experimental setup. Furthermore, to mitigate the significant temporal effects [19] observed in the EEG-Classification dataset, all our experimental setups strictly adhere to a dataset split where the first 30 images of each category are used as the training set, and the last 20 images are used as the test set. We strongly recommend that all researchers using the EEG-ImageNet dataset adopt a similar dataset split methodology. We also explain the impact of this split in Appendix A.5.

4.4 Models

In the object classification task, we employ simple machine-learning classification models such as ridge regression, KNN, random forest, and SVM. Additionally, we implement deep learning models including MLP, EEGNet [17], and RGNN [44]. MLP consists of two hidden layers, while EEGNet and RGNN are implemented using their original architectures. We use cross-entropy as the loss function. We train the models for each participant for 2000-3000 epochs on an NVIDIA 4090 GPU.

Table 2: The average results of all participants in the object classification task. * indicates the use of time-domain features, otherwise it indicates the use of frequency-domain features. † indicates that the difference compared to the best-performing model is significant with p-value < 0.05.

	Model	Acc (all)	Acc (coarse)	Acc (fine)
Classic model	Ridge	0.2859†	0.3944†	0.5833†
	KNN	0.3037†	0.4012†	0.6954†
	RandomForest	0.3489†	0.4535†	0.7288†
	SVM	0.3919	0.5057†	0.7784†
Deep model	MLP	0.4037	0.5339	0.8163
	EEGNet*	0.2604†	0.3030†	0.3645†
	RGNN	0.4050	0.4703†	0.7057†

In the image reconstruction task, we use a frozen Stable Diffusion 1.4 [29] as the backbone and train a two-layer MLP as an encoder to generate the prompt embeddings for input. For our reconstructions, we use 50 denoising timesteps with PNDM noise scheduling to generate 512x512 images. We use MSE as the loss function to align the encoder’s output with the CLIP (ViT-L/14) [26] embeddings of the captions of real images obtained through BLIP [18]. The model for each participant is trained for 2000 epochs with on an NVIDIA 4090 GPU.

For model and training details, please refer to the open-source code and Appendix A.3.

5 Experiments

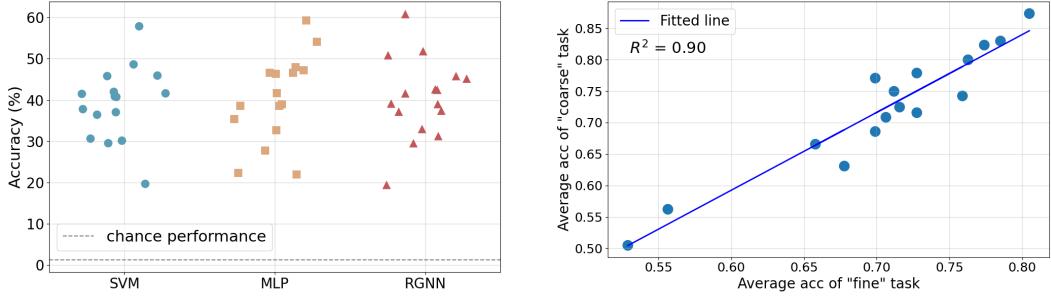
In this chapter, we present the experimental results of our study, focusing on two benchmark tasks: object classification and image reconstruction.

5.1 Object Classification

In the object classification task, we conduct experiments at three levels of granularity: all, coarse, and fine. The “all” category represents the 80-class classification accuracy, the “coarse” category represents the 40-class classification accuracy, and the “fine” category represents the average accuracy of five 8-class classification tasks. The performance of each model is detailed in Table 2, which shows the average results of all participants.

The table shows that RGNN achieves the highest accuracy in the “all” task among all models, with an 80-class classification accuracy reaching 40.50%. MLP slightly lags behind RGNN in the “all” task accuracy, but it significantly outperforms in the “coarse” and “fine” tasks, with a 40-class accuracy of 53.39% and an average 8-class accuracy of 81.63%. Among the classical machine learning models, SVM exhibits the highest classification performance, only slightly trailing RGNN and MLP across the three different granularity scenarios. The lower accuracy of RGNN in the “fine” task might be due to the complex model structure, which may require more fine-tuning to better adapt to easier tasks. Meanwhile, the lower performance of time-domain features compared to frequency-domain features suggests that frequency-domain features might be more informative for the classification tasks in this study. Specific performance details for each participant in each task are provided in Appendix A.4. We find that the ranking of participants’ accuracy is relatively consistent across different models.

We compile the accuracy of each participant for the “all” task under the SVM, MLP, and RGNN models, as shown in Figure 2a. We observe significant differences between participants, indicating high variability in individual responses. The best-performing participant achieves an accuracy of 60.88%. To better compare the differences between “fine” and “coarse” tasks in EEG-ImageNet, we modify the coarse-grained task. We randomly select 8 coarse-grained categories and use the RGNN model for training and testing. This process is repeated 5 times, and the average accuracy for each participant is calculated and plotted alongside their average “fine” task accuracy in Figure 2b. We then perform linear regression on the data points, and the resulting function has a slope greater than 1, indicating that participants generally achieve better results on the “coarse” classification tasks. This finding is also consistent with intuition.



(a) Classification accuracy for each participant in the object classification task (all) across different Models.

(b) Classification accuracy of the “coarse” (vertical axis) and “fine” task (horizontal axis).

Figure 2: Classification performance of the object classification task, each dot represents the classification performance of a single participant.



Figure 3: The image reconstruction results of a single participant (S8).

5.2 Image Reconstruction

Figure 3 shows some of the results from our image reconstruction pipeline of a single participant (S8). From the selected images, it can be seen that the reconstruction pipeline can effectively restore the category information of the image stimuli. However, restoring low-level details such as color, position, and shape is inaccurate. This may be because we align the EEG signals to the image captions obtained from BLIP. Due to the limited descriptive precision of BLIP, we cannot fully leverage the diffusion model’s generative capabilities.

Table 3 shows the average two-way identification (chance=50%) of the images generated by our reconstruction pipeline using different visual neural networks. Two-way identification refers to comparing the generated image with the original and one distractor to evaluate accuracy. The highest accuracy achieved by CLIP suggests that integrating vision transformers with extensive pre-training on diverse image-text data can significantly enhance model performance for complex tasks like image reconstruction from EEG signals. Additional results can be found in Appendix A.4.

6 Discussion and Conclusion

In this section, we outline the limitations of our dataset and explore how it could guide future research efforts to advance machine learning and brain-computer interface design.

Limitation. Firstly, while our dataset is more comprehensive than similar works, each participant’s data is still relatively limited. This necessitates the development of inter-subject models to overcome this limitation and enhance generalizability. Additionally, it is limited in representation, as participants

Table 3: The average results of all participants in the image reconstruction task.

Method	Alex(2)	Alex(5)	Incep	CLIP(ViT-L/14)
Two-way Identification	56.05%	62.99%	56.75%	64.67%

were drawn from a convenience sample at our university. This results in an age distribution skewed towards college-aged individuals and a racial composition predominantly White and Asian, which limits the dataset’s generalizability. Future work should aim to include a more diverse and extensive participant pool. Secondly, while we employed methods such as reducing the segment length of each EEG recording and sequentially splitting the training and test sets to mitigate the temporal effect, we were unable to eliminate it completely. Future work should explore more sophisticated techniques to address this issue. Lastly, our benchmarks did not incorporate many of the latest deep-learning methods. We believe that recent advancements in deep learning could greatly benefit from our comprehensive dataset, potentially leading to significant breakthroughs in visual neuroscience.

Insight for ML. The EEG-ImageNet dataset provides a comprehensive resource for developing models in visual recognition tasks, enabling the development of sophisticated deep-learning models capable of capturing intricate patterns within EEG data. Future research could leverage the dataset to enhance domain adaptation and transfer learning techniques, facilitating effective inter-subject task completion. Researchers might develop state-of-the-art models for the benchmarks we defined, or even create new tasks. By offering a diverse set of visual stimuli and supporting multi-level classification tasks, EEG-ImageNet could foster the creation of hierarchical models that mirror human cognitive processes and improve the generalization capabilities of machine learning algorithms.

Insight for BCI. As hardware technology progresses, portable EEG devices are becoming increasingly feasible, offering new opportunities for real-time BCI applications. Researchers could use the dataset to develop robust BCI systems that accurately interpret user intent from EEG signals. The comprehensive size and diverse visual stimuli in EEG-ImageNet allow for the creation of adaptive BCI systems that learn and respond to individual user patterns. This paves the way for personalized neurotechnology solutions, particularly enhancing human-computer interaction for individuals with disabilities. Furthermore, addressing privacy protection and ethical concerns will be crucial as BCI technology advances, ensuring user data is securely handled and individual rights are respected.

References

- [1] Emily J Allen, Ghislain St-Yves, Yihan Wu, Jesse L Breedlove, Jacob S Prince, Logan T Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, et al. A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, 25(1):116–126, 2022.
- [2] Richard Antonello, Aditya Vaidya, and Alexander Huth. Scaling laws for language encoding models in fmri. *Advances in Neural Information Processing Systems*, 36, 2024.
- [3] Yohann Benchetrit, Hubert Banville, and Jean-Rémi King. Brain decoding: toward real-time reconstruction of visual perception. *arXiv preprint arXiv:2310.19812*, 2023.
- [4] Nadine Chang, John A Pyles, Austin Marcus, Abhinav Gupta, Michael J Tarr, and Elissa M Aminoff. Bold5000, a public fmri dataset while viewing 5000 visual images. *Scientific data*, 6(1):49, 2019.
- [5] Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*, 2015.
- [6] Fan L Cheng, Tomoyasu Horikawa, Kei Majima, Misato Tanaka, Mohamed Abdelhack, Shuntaro C Aoki, Jin Hirano, and Yukiyasu Kamitani. Reconstructing visual illusory experiences from human brain activity. *Science Advances*, 9(46):eadj3906, 2023.
- [7] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th international IEEE/EMBS conference on neural engineering (NER)*, pages 81–84. IEEE, 2013.
- [8] David J Heeger and David Ress. What does fmri tell us about neuronal activity? *Nature reviews neuroscience*, 3(2):142–151, 2002.
- [9] Tomoyasu Horikawa and Yukiyasu Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, 8(1):15037, 2017.

- [10] Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao. Salicon: Saliency in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1072–1080, 2015.
- [11] Ujash Joshi and Michael Guerzhoy. Automatic photo orientation detection with convolutional neural networks. In *2017 14th Conference on Computer and Robot Vision (CRV)*, pages 103–108. IEEE, 2017.
- [12] Blair Kaneshiro, Marcos Perreau Guimaraes, Hyung-Suk Kim, Anthony M Norcia, and Patrick Suppes. A representational similarity analysis of the dynamics of object processing using single-trial eeg classification. *Plos one*, 10(8):e0135697, 2015.
- [13] N Kannathal, U Rajendra Acharya, Choo Min Lim, and PK Sadasivan. Characterization of eeg—a comparative study. *Computer methods and Programs in Biomedicine*, 80(1):17–23, 2005.
- [14] Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, 2008.
- [15] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.
- [16] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.
- [17] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.
- [18] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*, pages 12888–12900. PMLR, 2022.
- [19] Ren Li, Jared S Johansen, Hamad Ahmed, Thomas V Ilyevsky, Ronnie B Wilbur, Hari M Bharadwaj, and Jeffrey Mark Siskind. Training on the test set? an analysis of spampinato et al.[31]. *arXiv preprint arXiv:1812.07697*, 2018.
- [20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.
- [21] Yiqun Liu, Jiaxin Mao, Xiaohui Xie, Min Zhang, and Shaoping Ma. Challenges in designing a brain-machine search interface. In *ACM SIGIR Forum*, volume 54, pages 1–13. ACM New York, NY, USA, 2021.
- [22] Nikos K Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878, 2008.
- [23] Nikos K Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *nature*, 412(6843):150–157, 2001.
- [24] Juan Abdon Miranda-Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE transactions on affective computing*, 12(2):479–493, 2018.
- [25] Furkan Ozcelik and Rufin VanRullen. Natural scene reconstruction from fmri signals using generative latent diffusion. *Scientific Reports*, 13(1):15666, 2023.
- [26] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

- [27] Blake A Richards, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, et al. A deep learning framework for neuroscience. *Nature neuroscience*, 22(11):1761–1770, 2019.
- [28] Tal Ridnik, Emanuel Ben-Baruch, Asaf Noy, and Lihi Zelnik-Manor. Imagenet-21k pretraining for the masses. *arXiv preprint arXiv:2104.10972*, 2021.
- [29] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- [30] Paul Scotti, Atmadeep Banerjee, Jimmie Goode, Stepan Shabalin, Alex Nguyen, Aidan Dempster, Nathalie Verlinde, Elad Yundler, David Weisberg, Kenneth Norman, et al. Reconstructing the mind’s eye: fmri-to-image with contrastive learning and diffusion priors. *Advances in Neural Information Processing Systems*, 36, 2024.
- [31] Vinit Shah, Eva Von Weltin, Silvia Lopez, James Riley McHugh, Lillian Veloso, Meysam Golmohammadi, Iyad Obeid, and Joseph Picone. The temple university hospital seizure detection corpus. *Frontiers in neuroinformatics*, 12:83, 2018.
- [32] Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. Deep learning human mind for automated visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6809–6817, 2017.
- [33] Ghislain St-Yves, Emily J Allen, Yihan Wu, Kendrick Kay, and Thomas Naselaris. Brain-optimized deep neural network models of human visual areas learn non-hierarchical representations. *Nature communications*, 14(1):3329, 2023.
- [34] Yu Takagi and Shinji Nishimoto. High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14453–14463, 2023.
- [35] Jerry Tang, Meng Du, Vy Vo, Vasudev Lal, and Alexander Huth. Brain encoding models based on multimodal transformers can transfer across language and vision. *Advances in Neural Information Processing Systems*, 36, 2024.
- [36] Michal Teplan et al. Fundamentals of eeg measurement. *Measurement science review*, 2(2): 1–11, 2002.
- [37] Mariya Toneva, Jennifer Williams, Anand Bollu, Christoph Dann, and Leila Wehbe. Same cause; different effects in the brain. *arXiv preprint arXiv:2202.10376*, 2022.
- [38] Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 3485–3492. IEEE, 2010.
- [39] Dezhong Yao, Yun Qin, Shiang Hu, Li Dong, Maria L Bringas Vega, and Pedro A Valdés Sosa. Which reference should we use for eeg and erp practice? *Brain topography*, 32:530–549, 2019.
- [40] Ziyi Ye, Xiaohui Xie, Qingyao Ai, Yiqun Liu, Zhihong Wang, Weihang Su, and Min Zhang. Relevance feedback with brain signals. *ACM Transactions on Information Systems*, 42(4):1–37, 2024.
- [41] Haihong Zhang, Cuntai Guan, Kai Keng Ang, Chuanchu Wang, and Zheng Yang Chin. Bci competition iv–data set i: learning discriminative patterns for self-paced eeg-based motor imagery detection. *Frontiers in neuroscience*, 6:7, 2012.
- [42] Shaorun Zhang, Zhiyu He, Ziyi Ye, Peijie Sun, Qingyao Ai, Min Zhang, and Yiqun Liu. Eeg-svrec: An eeg dataset with user multidimensional affective engagement labels in short video recommendation. *arXiv preprint arXiv:2404.01008*, 2024.
- [43] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

[44] Peixiang Zhong, Di Wang, and Chunyan Miao. Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 13(3):1290–1301, 2020.

Checklist

1. For all authors...
 - (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? **[Yes]**
 - (b) Did you describe the limitations of your work? **[Yes]** See Section 6.
 - (c) Did you discuss any potential negative societal impacts of your work? **[Yes]** See Section 3.1 and Section 6.
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? **[Yes]** See Section 3.1.
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? **[TODO]**
 - (b) Did you include complete proofs of all theoretical results? **[TODO]**
3. If you ran experiments (e.g. for benchmarks)...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **[Yes]** See Section 3.
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **[Yes]** See Section 4.
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **[Yes]** The reported experimental results were averaged across multiple participants after being conducted for each participant individually.
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **[Yes]** See Section 4.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? **[Yes]** We use images from ImageNet as visual stimuli and cite relevant works.
 - (b) Did you mention the license of the assets? **[Yes]** In the GitHub repository.
 - (c) Did you include any new assets either in the supplemental material or as a URL? **[Yes]** In the GitHub repository.
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? **[Yes]** We registered to access ImageNet21k, and due to copyright protection, we only annotated the dataset with the image indices and the corresponding categories' wnids.
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **[Yes]** See Section 3.1.
5. If you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? **[Yes]** See Section 3.
 - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? **[Yes]** See Section 3.1.
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **[Yes]** See Section 3.

A Appendix

A.1 Additional Information about Dataset

The specific statistics of the dataset are shown in Table 4.

Table 4: The Statistics of EEG-ImageNet Dataset.

	#Categories	#Images	#Subjects	#EEG-image pairs	Datasize
EEG-ImageNet	80	4000	16	63850	15.88GB

As shown in Listing 1, the EEG-ImageNet dataset storage format is provided. The dataset can be accessed through the cloud storage link available in our GitHub repository <https://github.com/Promise-Z5Q2SQ/EEG-ImageNet-Dataset>. Due to file size limitations on the cloud storage platform, we split the dataset into two parts: “EEG-ImageNet_1.pth” and “EEG-ImageNet_2.pth”, each containing data from 8 participants. Users can choose to use only one of the parts based on their specific needs or device limitations.

```
{  
    "dataset": [  
        {  
            "eeg_data": torch.tensor,  
            "granularity": "coarse"/"fine",  
            "subject": 15,  
            "label": 'ne2106550',  
            "image": 'n02106550_1410.JPEG',  
        }, ...  
    ],  
    "labels": [  
        "n02106662", ...  
    ],  
    "images": [  
        "n02106662_13.JPG", ...  
    ]  
}
```

Listing 1: EEG-ImageNet dataset format.

A.2 Apparatus

All the image stimuli are presented on a desktop computer that has a 27-inch monitor with a resolution of 2,560×1,440 pixels and a refresh rate of 60 Hz. Participants are required to use the keyboard to interact with the platform. EEG signals are captured and amplified using a Scan NuAmps Express system (Compumedics Ltd., VIC, Australia) and a 64-channel Quik-Cap (Compumedical NeuroScan). A laptop computer functions as a server to record EEG signals and triggers using Curry8 software. Throughout the experiment, electrode-scalp impedance is maintained under 50Ω , and the sampling rate is set at 1,000Hz.

A.3 Experimental Setup Details

In the object classification task, we conduct experiments under three different granularity settings: the “all” task includes all 80 categories; the “coarse” task includes 40 coarse-grained categories; and the “fine” task includes 8 fine-grained categories that belong to the same parent node, with the average accuracy calculated across 5 groups. Each classification model maintains parameter consistency across the three tasks. We train one model per participant, ensuring that the parameters are consistent between models for different participants as well.

The model structures and hyperparameters are as follows. For SVM, we try linear, polynomial, and radial basis function (RBF) kernels. The regularization parameter is tested from values $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$. For RandomForest, we try to set the number of trees in the

forest from values $\{20, 50, 100, 200, 500\}$, with all other parameters set to their default values. For KNN, we set the number of neighbors to $\{5, 10, 15, 20\}$. For ridge regression, all parameters are set to their default values. For RGNN, when calculating the edge weights between electrodes, we use the hardware parameters of our data collection device to determine the topological coordinates of each electrode. In addition to the standard implementation, we add two batch normalization layers. The main hyperparameters adjusted are the number of output channels of the graph convolutional network (i.e., the hidden layer dimension) and the number of hops (i.e., the number of layers). These are set to $\{100, 200, 400\}$ and $\{1, 2, 4\}$ respectively. For EEGNet, we use the standard implementation and set the length of the first step convolution kernel to half the number of sampling time points, which is 200. The main hyperparameters adjusted are the number of output channels for the first convolutional layer (F1) and the depth multiplier (D), which are set to $\{8, 16, 32\}$ and $\{2, 4, 8\}$ respectively. For MLP, we set two hidden layers with dimensions of 256 and 128, respectively. Each linear layer is followed by a batch normalization layer and a dropout layer with a probability of 0.5.

For all deep models, we use the cross-entropy loss function. In MLP and EEGNet, we use the SGD optimizer with learning rate 10^{-3} , weight decay 10^{-3} , and momentum 0.9, training for 2000 epochs. After that, we adjust the learning rate to 10^{-4} and weight decay to 10^{-4} and continue training for another 1000 epochs. In RGNN, we use the Adam optimizer with learning rate 10^{-3} and weight decay 10^{-3} , training for 2000 epochs. Subsequently, we adjust the learning rate to 10^{-4} and weight decay to 10^{-4} and train for an additional 1000 epochs. The batch size is uniformly set to 80.

In the image reconstruction task, we first use blip-image-captioning-base to obtain captions for each stimulus image, with parameters set to "max length": 200 and "num beams": 20 to achieve relatively detailed descriptions. Then, we use the CLIP ViT-L/14 version's tokenizer and text encoder to compute the CLIP embedding for each caption, resulting in dimensions of (77, 768). Next, we map the frequency domain features of the EEG signals to (77, 768) using an MLP with two hidden layers, having dimensions of 1024 and 2048, respectively. This output is then fed into a frozen Stable Diffusion 1.4 model built with the diffusers library. After 50 steps of inference, we obtain image outputs with dimensions of 512x512. The Stable Diffusion model we built specifically consists of the UNet2DConditionModel and AutoencoderKL, and we use PNDMScheduler as the noise scheduler.

Two-way identification refers to a method where the generated image is compared with the original image and one distractor image. The task is to correctly identify the original image from the pair. This method evaluates the ability of the reconstruction pipeline to produce images that are distinguishable and recognizable as the original stimuli. In this study, we selected the second and fifth convolutional layers of AlexNet, the output before the linear layer of Inception, and the embeddings from CLIP ViT-L/14 as comparison features to calculate two-way identification.

All the implementations mentioned above are open-sourced and available in the GitHub repository.

A.4 Additional Experimental Results

Table 5 shows the performance of the best-performing participant across all models and tasks.

Table 5: The best results of all participants in the object classification task.

	Model	Acc (all)	Acc (coarse)	Acc (fine)
Classic model	Ridge	0.4550	0.5375	0.7200
	KNN	0.5025	0.6063	0.8013
	RandomForest	0.5006	0.6488	0.8450
	SVM	0.5794	0.7038	0.8588
Deep model	RGNN	0.6088	0.6525	0.8050
	EEGNet*	0.4413	0.5213	0.5988
	MLP	0.5925	0.7413	0.8875

Figure 4 shows the accuracy for each participant in the object classification task across SVM, MLP, and RGNN models. We find that the ranking of participants' accuracy is relatively consistent across different models.

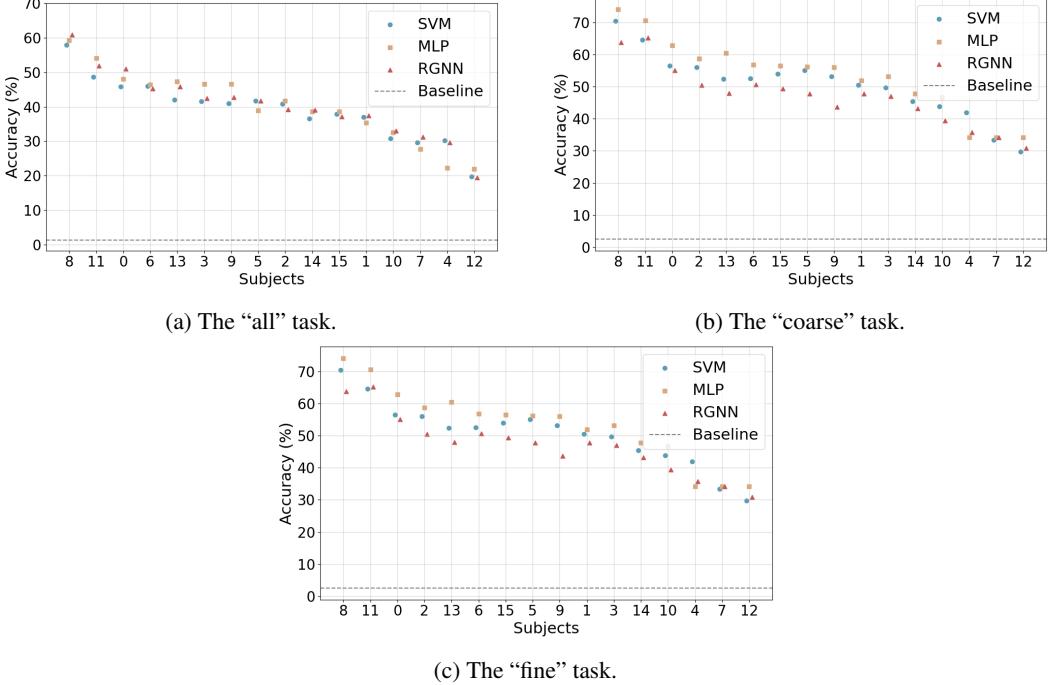


Figure 4: Acc for each participant in the object classification task across SVM, MLP, and RGNN Models.

Figure 5 presents more image generation results selected from other participants, with Figure 5a showing good cases and Figure 5b showing bad cases. We identified **three** main types of bad cases. Similar to the first two images, the reconstructed images lack or misrepresent low-level information such as color and shape. These errors are relatively common and are due to the limitations of our feature mapper and the simple structure of the reconstruction pipeline, resulting in insufficient information restoration. Similar to the latter two images, the reconstructed images lack detail. This limitation is due to the number of denoising steps in the diffusion model and the inherently low signal-to-noise ratio of EEG signals.

We also observed that for certain categories, especially fine-grained ones, all test data points resulted in near-noise outputs, which drew our attention. When we directly input category labels as text prompts into Stable Diffusion 1.4, we found that the generated images had poor realism and three-dimensional structure. Figure 5c compares these images with those generated by our reconstruction pipeline from the training set. This improvement suggests that we can use EEG, which can be quickly and extensively obtained as human feedback signals, to enhance the performance of text-image pre-trained models or generative models. This will be the direction of future research.

A.5 Temporal Effect

In our experimental paradigm, to reduce the cognitive load on participants, we group images of the same category together and use the RSVP (Rapid Serial Visual Presentation) paradigm for continuous rapid stimulation. This approach may cause the model to learn temporal continuity features rather than the intrinsic characteristics of the category stimuli. In this study, we use narrower cropped time segments as input features and divide the training and test sets based on temporal order to reduce the impact of temporal effects. In Figure 6, we plotted the average classification accuracy for images at different index positions in the test set under various training and test set splits.

We observed that the first few images in the test set have significantly higher accuracy, indicating a strong temporal effect, while the accuracy tends to stabilize for the subsequent images. Next, we also used several statistical methods for analysis. For the 30-20 split, we first calculated the sliding window standard deviation to measure local volatility. With a sliding window size of 5, the standard deviation at the 14th data point was less than 0.01, indicating a convergence trend at this point. We

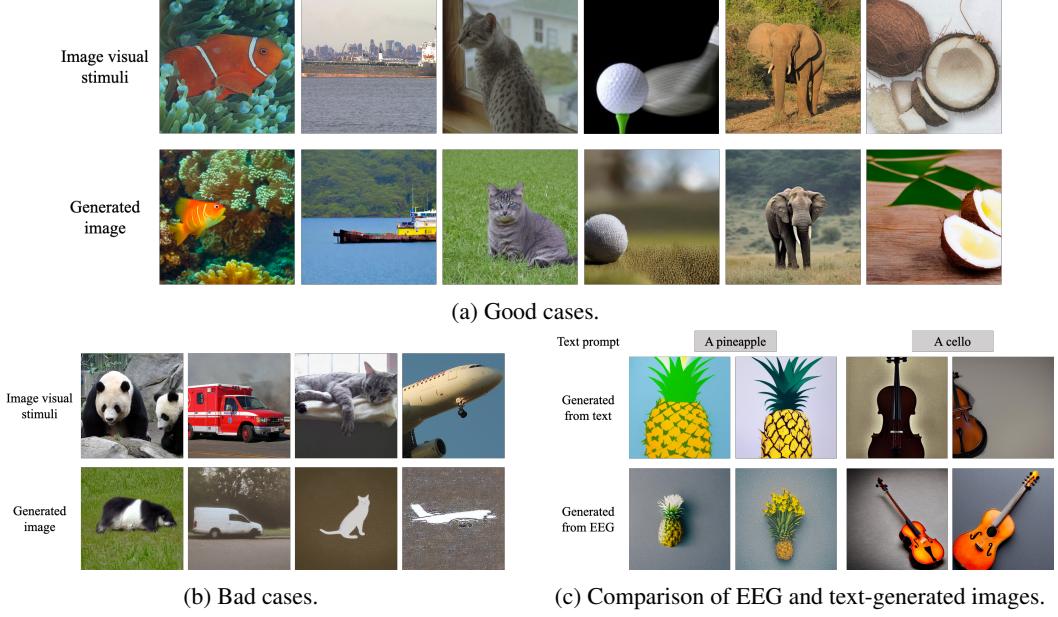


Figure 5: More results in the image generation task.

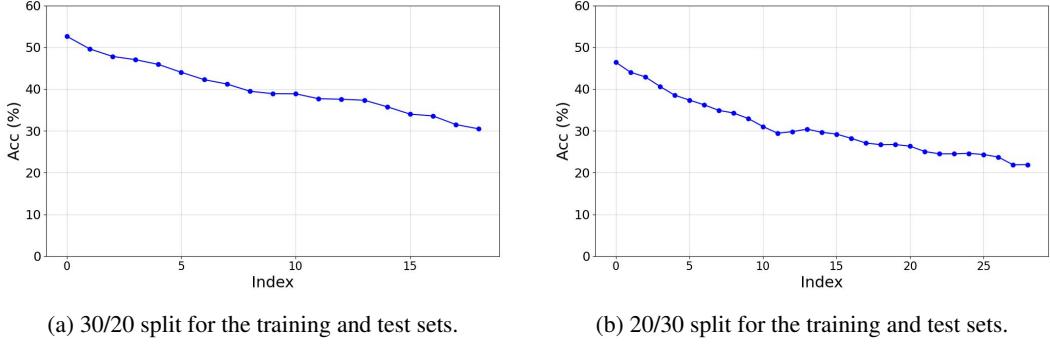


Figure 6: Average classification accuracy under different training and test set splits, with accuracy plotted against the indices of image stimuli in the test set.

then checked the stationarity of the data using the ADF test, obtaining an ADF statistic of -1.1874 and a p-value of 0.6790, suggesting that the data might not have fully converged. For the 20-30 split, the standard deviation at the 16th data point was less than 0.01 with a sliding window size of 5. The ADF test yielded a statistic of -3.8505 and a p-value of 0.0024, indicating that the data is stationary and convergent.

In summary, EEG-ImageNet is subject to certain temporal effects. The methods we employed reduced this impact but did not completely eliminate it. Therefore, we recommend that all researchers using this dataset adopt preprocessing techniques similar to ours.

Mind’s Eye: Image Recognition by EEG via Multimodal Similarity-Keeping Contrastive Learning

Chi-Sheng Chen

Department of Computer Science

National Yang Ming Chiao Tung University

Hsinchu, Taiwan

m50816m50816@gmail.com

Chun-Shu Wei

Department of Computer Science

National Yang Ming Chiao Tung University

Hsinchu, Taiwan

wei@nycu.edu.tw

Abstract

Decoding images from non-invasive electroencephalographic (EEG) signals has been a grand challenge in understanding how the human brain process visual information in real-world scenarios. To cope with the issues of signal-to-noise ratio and nonstationarity, this paper introduces a MUltimodal Similarity-keeping contrastivE learning (MUSE) framework for zero-shot EEG-based image classification. We develop a series of multivariate time-series encoders tailored for EEG signals and assess the efficacy of regularized contrastive EEG-Image pre-training using an extensive visual EEG dataset. Our method achieves state-of-the-art performance, with a top-1 accuracy of 19.3% and a top-5 accuracy of 48.8% in 200-way zero-shot image classification. Furthermore, we visualize neural patterns via model interpretation, shedding light on the visual processing dynamics in the human brain. The code repository for this work is available at: https://github.com/ChiShengChen/MUSE_EEG.

1 Introduction

Understanding visual processing in the human brain remains a profound challenge at the intersection of neuroscience and artificial intelligence. Visual processing involves a complex sequence of neural mechanisms across various brain regions, enabling the intricate processing of visual stimuli [1, 2, 3, 4, 5]. The development of deep learning techniques, such as convolutional neural networks (CNNs), has been significantly inspired by our understanding of these neural mechanisms [6, 7, 8]. Unveiling the brain dynamics of visual processing in real-world contexts holds the potential to inspire future advancements in artificial intelligence (AI), continuing the cycle of innovation driven by biological insights [9, 10]. Recent studies have advanced our understanding of visual processing in the human brain through the observation of brain activity using various neuromonitoring modalities [11]. Electroencephalography (EEG), as a non-invasive, portable modality with high-temporal resolution, offers a unique window into visual processing by revealing the instantaneous neural dynamics of visual perception and recognition in real-world contexts [12, 13, 14].

Decoding images from EEG signals represents a promising approach to study the mechanisms of visual processing. By leveraging EEG, researchers can gain insight into the temporal evolution of neural responses to visual stimuli [15]. However, this endeavor faces significant obstacles, primarily due to the low signal-to-noise ratio and nonstationarity of EEG signals [16, 17]. Addressing these challenges is crucial for advancing our understanding of visual cognition and for developing robust EEG-based image decoding or brain-computer interfacing (BCI) systems. Early studies in EEG-based image decoding have been constrained by the use of small datasets, limiting their ability to develop generalizable models [18, 19]. More recent work has utilized larger datasets collected through the rapid serial visual presentation (RSVP) paradigm, where images are presented in quick succession to

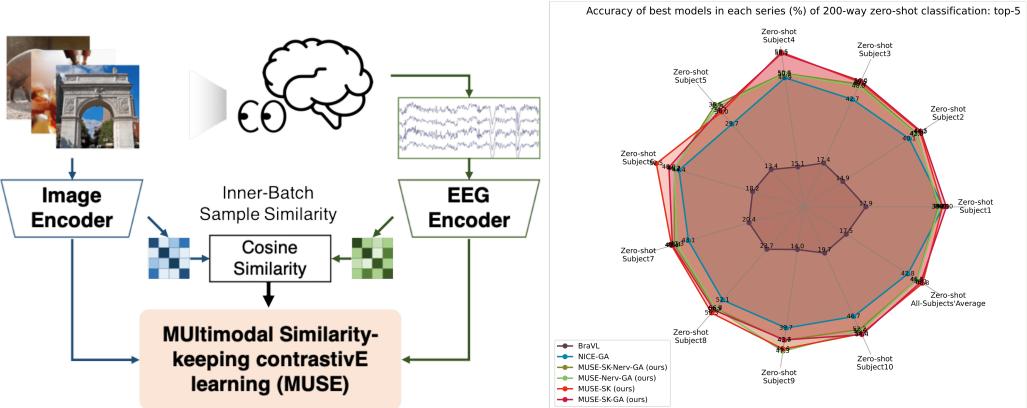


Figure 1: Schematic illustration of the proposed MUltimodal Similarity-keeping contrastive learning (MUSE) framework. During the training phase, EEG-image pairs are independently processed by an EEG encoder and an image encoder. The objectives of the MUSE framework are twofold: 1) maximize the separation between matched and unmatched pairs, and 2) maintain the inner-batch sample similarity within each EEG-image pair (see Algorithm 1 for details). In the test phase, an unseen EEG sample is passed through the EEG encoder, which identifies the most similar image from a set of unseen images based on cross-modality embedding similarity.

elicit brain responses [5, 20]. Despite these advances, the performance of existing methods remains suboptimal, underscoring the need for dedicated design of EEG encoding network architectures that consider the brain’s mechanisms and EEG characteristics.

To address the challenges in EEG-based image decoding, we present a novel self-supervised framework, coined as multimodal similarity-keeping contrastive learning (MUSE), dedicated to cross-modality contrastive learning between EEG and image data. We develop a series of multivariate time-series encoder network architectures tailored for EEG processing that facilitate the cross-modality contrastive learning with an advanced off-the-shelf image encoder (CLIP-ViT [21]). These encoders feature an upstream spatial convolution of EEG data for the sake of feature extraction and noise suppression [22, 23]. Additionally, we propose an innovative similarity-keeping contrastive learning mechanism, inspired by the cortical mapping organization of visual object representation in the inferotemporal (IT) cortex [24], to regularize the contrastive learning process using the information of inter-object relationships within both EEG and image samples.

Furthermore, we employ model interpretation techniques to visualize the neural patterns of image processing, offering a deeper understanding of the underlying dynamics of visual cognition in the human brain. The contributions of this work are threefold:

- We introduce a novel self-supervised multimodal similarity-keeping contrastive learning (MUSE) framework that achieves state-of-the-art performance in zero-shot EEG-based image recognition.
- We propose EEG encoders with upstream spatial convolution and similarity-keeping regularization to enhance EEG-image cross-modality contrastive learning.
- We visualize neural patterns through model interpretation to provide neuroscientific insights into the spatial and temporal brain dynamics of visual processing.

2 Related Works

2.1 Decoding Visual Information from Brain Signals

Interpreting visual data from the human brain has been a longstanding challenge at the intersection of neuroscience and computer science [1, 2, 4, 5]. Despite significant advancements in understanding static visual inputs, rapidly and accurately extracting meaningful information from natural imagery remains difficult [25, 26]. Previous efforts have primarily utilized functional magnetic resonance

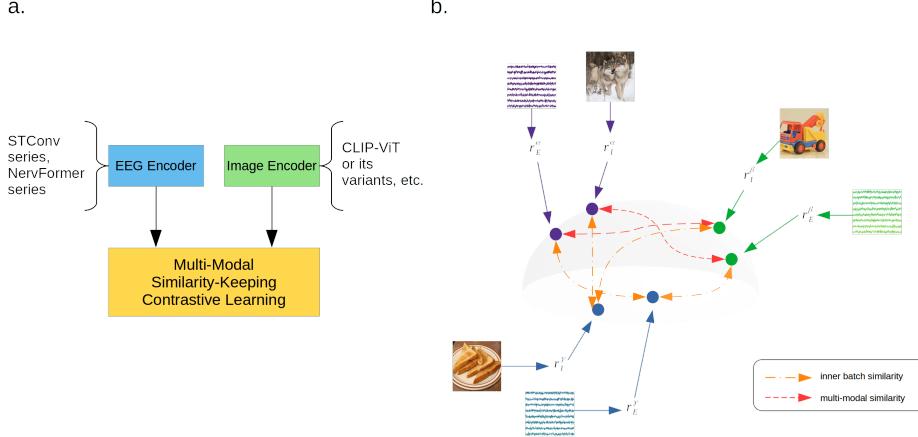


Figure 2: (a.) The whole view of this work. (b.) Illustration on feature space of multimodal similarity-keeping contrastive learning framework (MUSE), different from traditional contrastive learning only focus on multimodal similarity, MUSE both consider the multimodal similarity and inner batch similarity in the loss function. r denotes representation. I and E denotes image and EEG signal, respectively.

imaging (fMRI) [27, 28, 29], which has demonstrated the ability to capture meaningful content and structural details from visual processing in the brain. However, fMRI relies on detecting changes in blood oxygenation, resulting in a temporal lag of several seconds per stimulus, thereby limiting its utility for real-time applications. Additionally, fMRI is expensive and requires large, stationary equipment.

In contrast, electroencephalography (EEG) offers superior temporal resolution, immediate data feedback, and portable, cost-effective hardware. These attributes position EEG as a promising candidate for personal brain-computer interface technology. Nevertheless, current methods for using EEG to extract semantic information for image classification have not achieved satisfactory results [30, 31, 20], highlighting the need for improved approaches. Previous methodologies have often relied on supervised learning techniques with a limited set of image categories, ignoring the intrinsic correlations between visual stimuli and neural responses [31, 18, 32]. These limitations impair their effectiveness in real-world scenarios that require the generalization to recognize novel, unfamiliar object categories. To address these issues, [33] first attempted zero-shot classification using the largest available EEG-image database [5] with a multilayer MLP and joint EEG-image-text representation, while [20] employed a contrastive learning method. However, [20] utilized a basic contrastive learning framework based on CLIP [21]. Our work improves upon this framework and the EEG encoder, introducing a self-supervised learning approach for EEG-based image decoding. This framework allows the model to generalize to object recognition tasks without specific prior training, demonstrating its effectiveness.

2.2 Multimodal Contrastive Learning

In recent years, after the success of the traditional contrastive learning models on the same modal data like text and image [34, 35, 36, 37], the development of multimodal contrastive learning has reached significant advancements in the field of self-supervised learning, particularly in tasks that contain the integration of multiple types of data. This method leverages the strengths of various modalities (e.g., text, images, video) to boost model generalization across diverse datasets. Multimodal contrastive learning aligns representations from different modalities within a shared embedding space, facilitating robust, modality-invariant feature learning. This enhances capabilities in cross-modal retrieval and zero-shot learning. Typically, a two-tower network architecture processes each modality independently, with outputs converging in the embedding space where contrastive loss minimizes distances between similar pairs and maximizes distances between dissimilar ones. One of the most popular and successful multimodal contrastive learning framework is CLIP [21], which project both the image and text to the same feature space. Nevertheless, because datasets containing both time-series signals like EEG and image data are quite rare, there has been little research applying

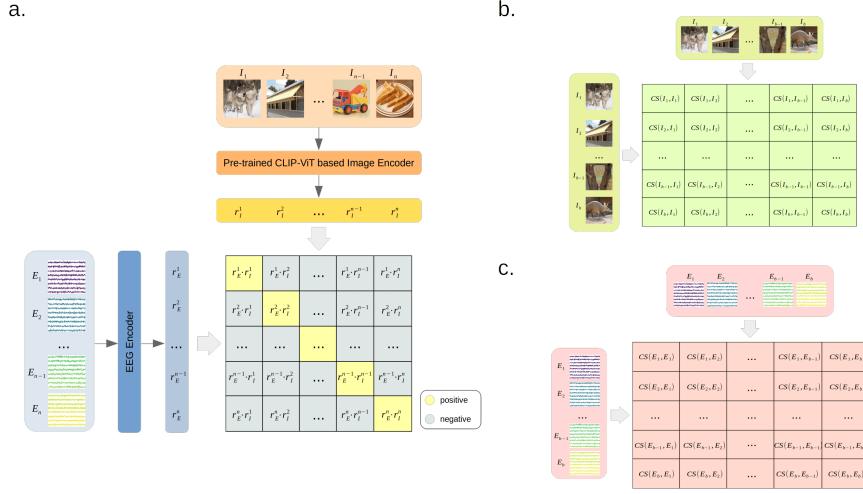


Figure 3: The details of the MUSE. (a.) The contrastive learning loss is calculated from EEG encoding and image encoding. (b.) (c.) The similarity-keeping loss comes from the final similarity of self-batch similarity of the input modal data.

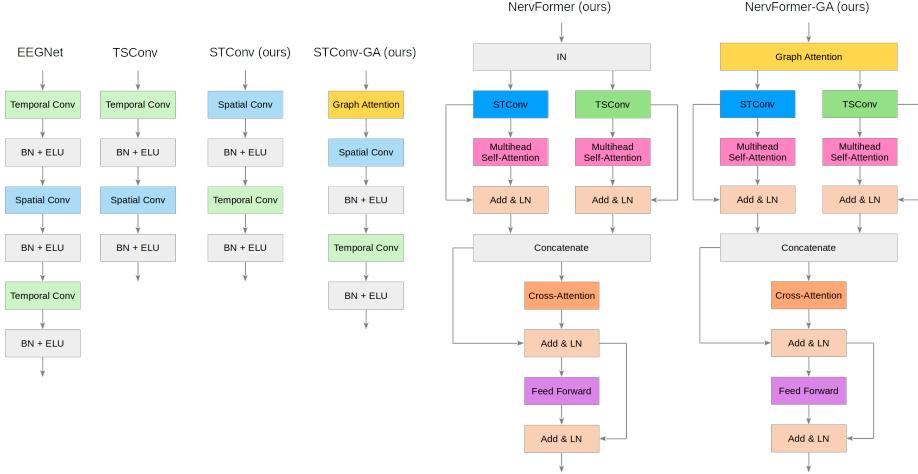


Figure 4: The model structure comparison. Where BN denotes batch normalization, IN denotes instance normalization, LN denotes layer normalization, respectively.

contrastive learning methods to this combination of temporal and visual information. To our best knowledge, [38] is maybe the first work introduced the EEG-image contrastive learning on obtaining the EEG-image representation for image reconstruction downstream task but do not do the zero-shot classification. [32] introduced the EEGClip network for joint representation learning between EEG signal and image but it just do supervised learning. [20] first try to design the EEG encoder on EEG-image contrastive learning , but the work only modified the encoders. This area remains largely uncharted and calls for new, specialized contrastive learning techniques to handle these joint time-series and image modalities effectively.

3 Methodology

3.1 Overview

This section introduces the Multimodal Similarity-Keeping Contrastive Learning (MUSE) framework, comprising the EEG encoder, image encoder, and the contrastive learning method. Our contribution

Algorithm 1 Multimodal Similarity-Keeping Contrastive Learning framework (MUSE)

```
1: Input: (Image, EEG)                                     ▷ stimulus & response
2: Model:  $Enc_{img}$ : CLIP-ViT or its variance,  $Enc_{eeg}$ : STConv or NervFormer

3: # E : (batch, channel, electrode, data sample)           ▷ batch of input EEGs
4: # I : (batch, channel, height, width)                      ▷ batch of input images

5: #  $\tau$  : learned temperature parameter
6: #  $\beta$  : learned inner similarity parameter
7: # CS : Cosine Similarity
8: # SK : Similarity-Keeping

9: # extract normalized representations from the raw image and EEG
10:  $E_f = \text{Norm}(\text{Linear}(Enc_{eeg}(E)))$ 
11:  $I_f = \text{Norm}(Enc_{img}(I))$                                 ▷ can be obtained before training

12: # calculate cosine similarity from the inner batch image and EEG
13:  $E_{CS} = \text{CS}(E_f, E_f)$ 
14:  $I_{CS} = \text{CS}(I_f, I_f)$ 
15:  $loss_{SK} = 1 - \mathbb{E}(\text{CS}(E_{CS}, I_{CS}))$ 

16: # scaled pairwise cosine similarity
17: logits =  $\text{dot}(E_f, I_f.t) \times e^{\tau}$ 

18: # symmetric loss function
19: labels = arange(batch)                                    ▷ self-supervised learning label
20:  $loss_e = \text{CrossEntropyLoss}(\text{logits}, \text{labels}, \text{axis}=0)$ 
21:  $loss_i = \text{CrossEntropyLoss}(\text{logits}, \text{labels}, \text{axis}=1)$ 
22:  $total\_loss = (loss_e + loss_i) / 2 + \beta \times loss_{SK}$ 
```

encompasses cutting-edge EEG encoders tailored for zero-shot classification tasks: the Spatial-Temporal convolution (STConv) and NervFormer architectures, along with a pioneering regularized contrastive learning approach featuring a novel similarity-keeping loss.

3.2 Network Architecture

3.2.1 EEG Encoder

In this study, we introduce a series of multivariate time-series encoding architectures tailored to capture essential features in EEG data. Recent works suggest that upstream spatial convolution serves as an effective spatial filtering method for enhancing feature extraction and noise suppression [22, 23]. Herein, we present the Spatial-Temporal Convolution (STConv) module, which employs spatial convolution to denoise data by referencing between brain electrodes, followed by temporal convolution. Additionally, we extend the capabilities of the STConv and Temporal-Spatial Convolution (TSConv) modules by integrating an attention mechanism, leading to the development of a novel transformer-like EEG encoder, which we refer to as NervFormer. In line with Graph Attention Networks (GATs) principles, we employ the Graph Attention (GA) module (see Appendix) to iteratively refine the state of each node, conceptualized as electrodes, by leveraging the states of all other nodes [39, 40]. The architectures of the baseline and proposed EEG encoders are illustrated and compared in Figure 4.

3.2.2 Image Encoder

For our implementation, we integrate the off-the-shelf CLIP-ViT model [21], which has demonstrated exceptional performance in aligning image and text representations. This model, pre-trained on extensive datasets, captures intricate details and high-level semantic information from images, making it an ideal candidate for our contrastive learning framework.

3.2.3 Similarity-Keeping Contrastive Learning

Inspired by recent neuroscience findings of the cortical network of visual object representation [24, 41], we take the interplay between object categories into account and propose a novel regularized contrastive learning framework. The procedure is outlined in Algorithm 1.

The ordinary contrastive learning uses InfoNCE loss given by [42, 35, 21]:

$$\mathcal{L}_{InfoNCE} = -\mathbb{E} \left[\log \frac{\exp(S_{E,I}/\tau)}{\sum_{k=1}^N \exp(S_{E,I_k}/\tau)} \right] \quad (1)$$

where the $S_{E,I}$ denotes the similarity score between EEG signal E and image I pairing data, the τ is learned temperature parameter, the training process shown in Figure 2.

We introduce regularization to the ordinary contrastive learning by incorporating similarity preservation into the contrastive loss to capture both inter-sample and multimodal similarities. Drawing inspiration from the similarity-keeping (SK) concept used in knowledge distillation between EEG models [43], we propose a novel SK loss to regularize the InfoNCE loss. This involves estimating the inner-batch inter-sample relationship. The SK loss is defined as:

$$\mathcal{L}_{SK} = 1 - \mathbb{E}[S(S_{E,E}, S_{I,I})] \quad (2)$$

We introduce a trainable parameter β to enhance training flexibility. When the $\beta = 0$, the similarity-keeping InfoNCE loss reduces to the standard InfoNCE loss. The combined loss function, which integrates similarity-keeping, is illustrated in Figure 3 and defined as:

$$\mathcal{L}_{SK-InfoNCE} = \mathcal{L}_{InfoNCE} + \beta \times \mathcal{L}_{SK} \quad (3)$$

This integration of similarity-keeping into the contrastive loss framework ensures that the model not only aligns paired EEG and image embeddings effectively but also maintains the intrinsic relationships within the batch.

4 Experiments

4.1 Datasets and Preprocessing

The ThingsEEG dataset [5] comprises extensive EEG recordings gathered through a rapid serial visual presentation (RSVP) paradigm, featuring responses from 10 individuals to 16,740 natural images from the THINGS database [44]. The dataset includes 1654 training classes, each with 10 images, and 200 test classes, each with 1 image. EEG recordings were conducted using 64-channel EASYCAP equipment, and the data were preprocessed by segmenting into trials from 0 to 1000 ms post-stimulus onset, with baseline correction using the pre-stimulus mean. EEG responses for each image were averaged across repetitions, and the images were resized to 224×224 and normalized prior to processing.

4.2 Experiment Details

Experiments were conducted on a GeForce RTX 3090 24G GPU with Pytorch. Training using the MUSE series required approximately 2 to 3 hours per subject, with a batch size of 1000, while NervFormer series models took 40 minutes to 1 hour per subject. Models were saved at 200 epochs when the validation loss reached its lowest point. We use the weighted Adam optimizer with a learning rate of 0.0002 and parameters $\beta_1=0.5$ and $\beta_2=0.999$. The τ in contrastive learning initialized with $\log(1/0.07)$ and $\beta=1$. The NervFormer model achieves the best results with a multiheads number of 5. Results were averaged over five random seeds, and statistical significance was determined using the Wilcoxon Signed-Rank Test.

Table 1: Overall accuracy (%) of 200-way zero-shot classification using CLIP-ViT as image encoder: top-1 and top-5. The parts in bold represent the best results, while the underlined parts are the second best.

Method	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Ave											
	top-1	top-5	Ave																			
Subject dependent - train and test on one subject																						
BraVL	6.1	17.9	4.9	14.9	5.6	17.4	5.0	15.1	4.0	13.4	6.0	18.2	6.5	20.4	8.8	23.7	4.3	14.0	7.0	19.7	5.8	17.5
NICE	12.3	36.6	10.4	33.9	13.1	39.0	16.4	47.0	8.0	26.9	14.1	40.6	15.2	42.1	20.0	49.9	13.3	37.1	14.9	41.9	13.8	39.5
NICE-SA	13.3	40.2	12.1	36.1	15.3	39.6	15.9	49.0	9.8	34.4	14.2	42.4	17.9	43.6	18.2	50.2	14.4	38.7	16.0	42.8	14.7	41.7
NICE-GA	<u>15.2</u>	40.1	13.9	40.1	14.7	42.7	17.6	48.9	9.0	29.7	16.4	44.4	14.9	43.1	20.3	52.1	14.1	39.7	19.6	46.7	15.6	42.8
MUSE-Nerv (ours)	11.0	33.9	12.3	37.4	13.6	39.4	19.1	48.0	10.7	31.9	14.0	41.2	13.0	41.3	21.0	54.6	15.4	38.6	17.1	43.9	14.7	41.0
MUSE-SK-Nerv (ours)	11.6	34.7	14.3	40.4	13.6	38.2	20.8	48.6	12.0	32.2	16.1	41.5	15.7	43.7	24.1	54.4	17.2	41.7	17.1	44.7	16.3	42.0
MUSE-SK-Nerv-GA (ours)	12.1	38.7	15.2	43.0	18.5	48.8	24.4	50.6	14.0	36.0	18.0	46.1	19.7	48.4	24.3	56.9	17.8	43.7	21.9	52.2	18.6	46.5
MUSE-Nerv-GA (ours)	13.4	39.0	<u>17.6</u>	42.8	17.3	48.0	22.6	50.3	14.4	35.9	18.7	46.2	19.2	47.3	26.8	56.7	19.0	47.3	20.6	52.9	19.0	46.6
MUSE (ours)	14.7	39.2	15.2	45.3	19.3	48.7	25.9	61.0	12.6	36.0	18.5	50.6	20.2	50.1	26.3	58.6	19.0	45.7	20.4	54.0	<u>19.2</u>	48.9
MUSE-GA (ours)	14.7	38.3	17.5	47.4	17.1	48.0	24.8	58.2	11.5	34.9	18.5	50.5	19.3	49.1	24.3	55.1	16.9	40.3	24.0	55.8	18.8	47.8
MUSE-SK (ours)	14.4	39.9	16.5	44.2	19.7	49.5	26.4	58.6	13.2	34.0	19.1	52.5	19.5	49.4	26.8	59.3	17.6	46.6	20.1	54.3	19.3	48.8
MUSE-SK-GA (ours)	<u>15.3</u>	41.0	18.1	44.5	20.0	50.0	25.3	58.1	11.2	34.7	17.9	48.0	<u>20.1</u>	49.1	25.4	57.7	17.0	43.6	22.7	54.4	19.3	48.1

Table 2: Ablation Study of MUSE series models, accuracy (%) of 200-way zero-shot classification: top-1 and top-5. The parts in bold represent the best results, while the underlined parts are the second best.

Method	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Ave	Win										
	top-1	top-5	subject score #																			
Subject dependent - train and test on one subject																						
<i>Original MUSE (STConv as EEG encoder & CLIP-ViT as image encoder with InfoNCE loss)</i>																						
MUSE	<u>14.7</u>	39.2	15.2	45.3	19.3	48.7	25.9	61.0	12.6	36.0	18.5	50.6	20.2	50.1	26.3	58.6	19.0	45.7	20.4	54.0	<u>19.2</u>	48.9
<i>Change InfoNCE loss to SK-InfoNCE loss</i>																						
MUSE-SK	14.4	39.9	16.5	44.2	19.7	49.5	26.4	58.6	13.2	34.0	19.1	52.5	19.5	49.4	26.8	59.3	17.6	46.6	20.1	54.3	19.3	48.8
<i>Change STConv to STConv-GA</i>																						
MUSE-SK-GA	<u>15.3</u>	41.0	18.1	44.5	20.0	50.0	25.3	58.1	11.2	34.7	17.9	48.0	<u>20.1</u>	49.1	25.4	57.7	17.0	43.6	22.7	54.4	19.3	48.1
6/20																						
7/20																						
7/20																						

4.3 Performance Comparison

The comparison results presented in Table 1 highlight the performance of various methods, with detailed model abbreviations provided in the appendix. Overall, MUSE-SK achieves the highest average top-1 accuracy at 19.3%, while MUSE attains the highest average top-5 accuracy at 48.9%. Furthermore, MUSE-SK-Nerv-GA, MUSE-Nerv-GA, MUSE, MUSE-SK, MUSE-SK-GA, MUSE-GA, and MUSE-SK-Nerv-GA significantly outperform the NICE-GA model in both top-1 ($p < 0.01$) and top-5 ($p < 0.01$) accuracy. Although individual performance can differ, MUSE-based methods usually do better than others. The GA and SK variants are particularly strong in this evaluation.

4.4 Ablation Study

We conduct ablation studies on both MUSE and MUSE-Nerv series models, with the results of MUSE-Nerv illustrated in Table 3. While the NervFormer EEG encoder does not demonstrate the best average zero-shot performance across all datasets, the MUSE-SK-Nerv-GA model achieves higher individual accuracy for subjects 5 and 10 compared to both MUSE and MUSE-SK. Moreover, beyond the MUSE series models, which solely employ the STConv as the EEG encoder, the MUSE-Nerv series models, incorporating the NervFormer as the EEG encoder, independently validate the efficacy of the similarity-keeping loss architecture and the graph attention module in EEG-image multimodal contrastive learning.

Upon examining the performance metrics of MUSE as depicted in Table 2, it becomes apparent that MUSE, MUSE-SK, and MUSE-SK-GA exhibit similar average performance levels. However, each method demonstrates distinct advantages across the ten subjects studied. For example, MUSE-SK-GA demonstrates superior overall performance in subjects 1, 3, and 10, while MUSE-SK achieves state-of-the-art results in subject 8. Additionally, each method excels uniquely in either top-1 or top-5 rankings in various subjects. This underscores the effectiveness of the SK and GA techniques as enhancements. However, in the context of STConv, these techniques do not demonstrate as clear an advantage as NervFormer does. We also observe that while SK may impact GA performance on NervFormer, both SK and GA enhance performance on STConv, with further details discussed in the model interpretation section.

Table 3: Ablation Study of MUSE-Nerv series models, accuracy (%) of 200-way zero-shot classification: top-1 and top-5. The parts in bold represent the best results, while the underlined parts are the second best.

Method	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Ave	Win											
	top-1	top-5																					
Subject dependent - train and test on one subject																							
<i>Original MUSE-Nerv (NervFormer as EEG encoder & CLIP-ViT as image encoder with InfoNCE loss)</i>																							
MUSE-Nerv	11.0	33.9	12.3	37.4	13.6	39.4	19.1	48.0	10.7	31.9	14.0	41.2	13.0	41.3	21.0	54.6	15.4	38.6	17.1	43.9	14.7	41.0	0
<i>Change InfoNCE loss to SK-InfoNCE loss</i>																							
MUSE-SK-Nerv	11.6	34.7	14.3	40.4	13.6	38.2	20.8	48.6	12.0	32.2	16.1	41.5	15.7	43.7	24.1	54.4	17.2	41.7	17.1	44.7	16.3	42.0	0
<i>Change NervFormer to NervFormer-GA</i>																							
MUSE-SK-Nerv-GA	12.1	38.7	15.2	43.0	18.5	48.8	24.4	50.6	14.0	36.6	18.0	46.1	19.7	48.4	24.3	56.9	17.8	43.7	21.9	52.2	18.6	46.5	20/20

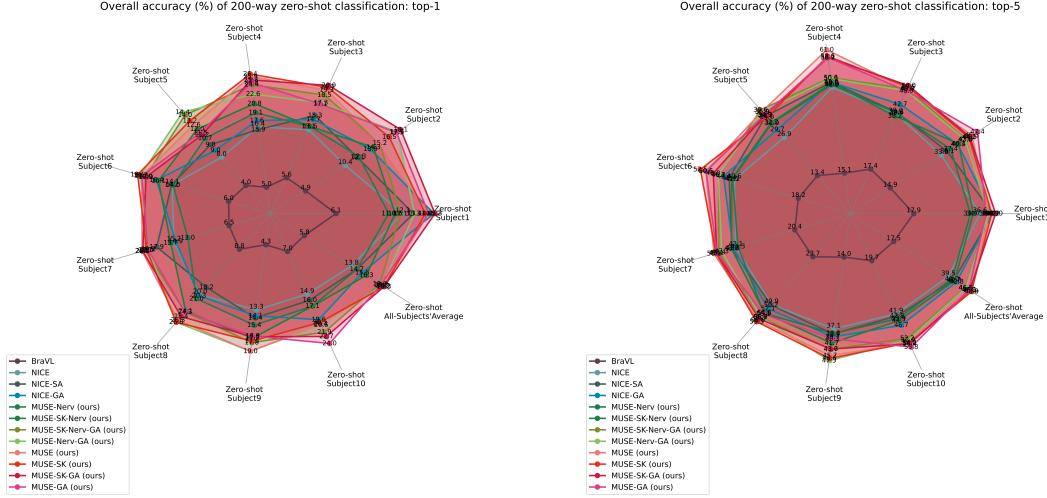


Figure 5: Overall Top-1 zero-shot accuracy comparison of all models.

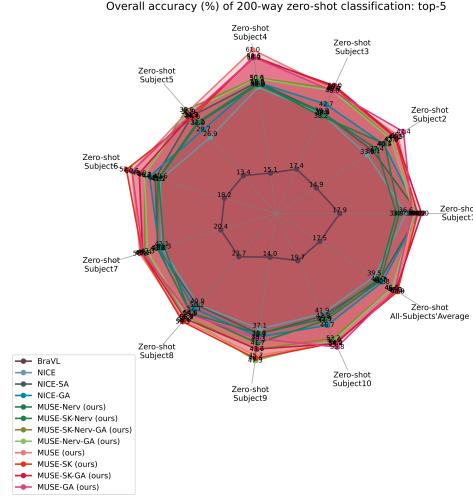


Figure 6: Overall Top-5 zero-shot accuracy comparison of all models.

4.5 Model Interpretation

We conducted model interpretation to uncover the internal mechanisms of our models across three distinct domains: spatial-temporal, brain region topography-temporal, and temporal-frequency. We employed the Grad-CAM analysis method [45] to scrutinize our proposed best MUSE series models.

4.5.1 Spatial-Temporal Dynamics Analysis

To ensure that meaningful signals are preserved during Grad-CAM calculations, we take the absolute value of all Grad-CAM and EEG signal intensities of each trial for further analysis. The spatial-temporal comparison on both training and testing trials is depicted in Figure 7. We note that the higher-performing models, such as MUSE-SK and MUSE-SK-GA, concentrate on the EEG information between the 25th and 125th data points, corresponding to the 100 ms to 500 ms time period. Figure 8 illustrates a distinct response observed in the occipital cortex between 100 and 600 ms after the onset in MUSE-SK. However, the 200 ms stimulus onset asynchrony (SOA) continues to elicit periodic responses in the occipital cortex. Furthermore, a response in the parietal cortex is evident after 100 ms. This observation aligns with the bottom-up hierarchy of the visual system [46], wherein visual stimuli are sequentially processed by V1, V2, and V4 in the occipital cortex, and subsequently by the inferotemporal region in the temporal cortex along the ventral stream for object recognition [24].

5 Conclusion

In summary, this paper introduces the MULTimodal Similarity-keeping contrastivE learning (MUSE) framework, a novel approach tailored specifically for zero-shot EEG-based image classification, thereby addressing the intricate challenge of deciphering visual information from non-invasive EEG signals. Our method, drawing inspiration from established neuroscience findings, achieves state-of-the-art performance in zero-shot classification tasks.

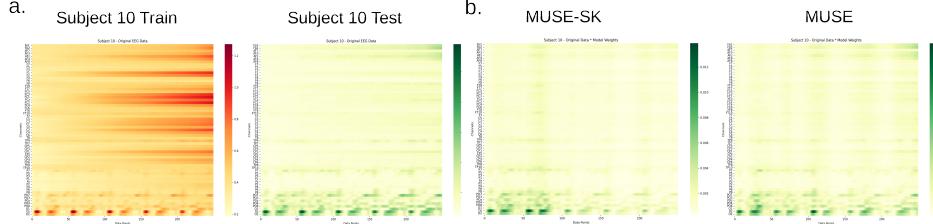


Figure 7: (a) Grad-CAM visualization of the MUSE series model averaged across all trials and repetitions for subject 10. (b) Comparative analysis reveals that MUSE-SK exhibits a heightened focus on the occipital lobes during the 100-500 ms time window compared to MUSE-SK and other models.

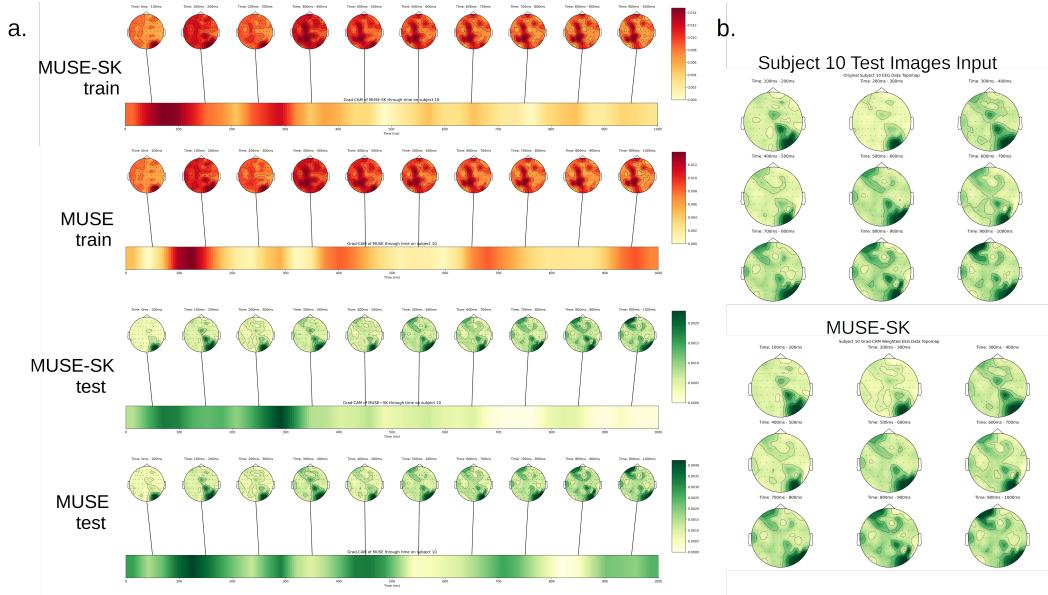


Figure 8: Topomaps depicting the average response over each 100 ms interval across all trials, aggregated over all repetitions for subject 10. (a) Grad-CAM visualization for both MUSE-SK and MUSE models is presented, with the color bar at the bottom indicating the intensity of Grad-CAM for each model over time. Both models predominantly focus on the 100-500 ms time window. (b) A zoomed-in comparison between the input EEG data and the MUSE-SK model highlights the model’s enhanced focus on temporal and occipital areas.

the-art decoding accuracy, as substantiated by rigorous experimental evaluations. We further interpret our models and uncover insights into the spatial-temporal dynamics of EEG responses, shedding light on the neural processes underlying visual perception. We foresee that our work will catalyze further exploration in bridging the gap between EEG decoding and image recognition, advancing our understanding of visual cognition in the human brain.

References

- [1] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025, 1999.
- [2] Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki C Tanabe, Norihiro Sadato, and Yukiyasu Kamitani. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–929, 2008.
- [3] Hesheng Liu, Yigal Agam, Joseph R Madsen, and Gabriel Kreiman. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, 62(2):281–290, 2009.
- [4] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.
- [5] Alessandro T Gifford, Kshitij Dwivedi, Gemma Roig, and Radoslaw M Cichy. A large and rich eeg dataset for modeling human visual object recognition. *NeuroImage*, 264:119754, 2022.
- [6] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.
- [7] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [9] Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258, 2017.
- [10] Shimon Ullman. Using neuroscience to develop artificial intelligence. *Science*, 363(6428):692–693, 2019.
- [11] Bin He, Lin Yang, Christopher Wilke, and Han Yuan. Electrophysiological imaging of brain activity and connectivity—challenges and opportunities. *IEEE transactions on biomedical engineering*, 58(7):1918–1931, 2011.
- [12] Guillaume A Rousselet, Jesse S Husk, Patrick J Bennett, and Allison B Sekuler. Single-trial eeg dynamics of object and face visual processing. *Neuroimage*, 36(3):843–862, 2007.
- [13] Jason Samaha and Bradley R Postle. The speed of alpha-band oscillations predicts the temporal resolution of visual perception. *Current Biology*, 25(22):2985–2990, 2015.
- [14] Chun-Shu Wei and Tzyy-Ping Jung. Towards real-world neuromonitoring and applications in cognitive engineering. *Handbook of Neuroengineering*, pages 3387–3404, 2023.
- [15] Amanda K Robinson, Praveen Venkatesh, Matthew J Boring, Michael J Tarr, Pulkit Grover, and Marlene Behrmann. Very high density eeg elucidates spatiotemporal aspects of early visual processing. *Scientific reports*, 7(1):16248, 2017.
- [16] Alexander Ya Kaplan, Andrew A Fingelkurts, Alexander A Fingelkurts, Sergei V Borisov, and Boris S Darkhovsky. Nonstationary nature of the brain activity as revealed by eeg/meg: methodological, practical and conceptual challenges. *Signal processing*, 85(11):2190–2212, 2005.
- [17] Jose Antonio Urigüen and Begoña Garcia-Zapirain. Eeg artifact removal—state-of-the-art and guidelines. *Journal of neural engineering*, 12(3):031001, 2015.
- [18] Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. Deep learning human mind for automated visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6809–6817, 2017.
- [19] Praveen Tirupattur, Yogesh Singh Rawat, Concetto Spampinato, and Mubarak Shah. Thoughtviz: Visualizing human thoughts using generative adversarial network. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 950–958, 2018.
- [20] Yonghao Song, Bingchuan Liu, Xiang Li, Nanlin Shi, Yijun Wang, and Xiaorong Gao. Decoding Natural Images from EEG for Object Recognition. In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024.

- [21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.
- [22] Chun-Shu Wei, Toshiaki Koike-Akino, and Ye Wang. Spatial component-wise convolutional network (scenet) for motor-imagery eeg classification. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 328–331. IEEE, 2019.
- [23] Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. Matt: A manifold attention network for eeg decoding. *Advances in Neural Information Processing Systems*, 35:31116–31129, 2022.
- [24] Pinglei Bao, Liang She, Mason McGill, and Doris Y Tsao. A map of object space in primate inferotemporal cortex. *Nature*, 583(7814):103–108, 2020.
- [25] Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, 2008.
- [26] Sung-Yu Chen, Chi-Min Chang, Kuan-Jung Chiang, and Chun-Shu Wei. Ssvep-dan: A data alignment network for ssvep-based brain computer interfaces. *arXiv preprint arXiv:2311.12666*, 2023.
- [27] Weijian Mai, Jian Zhang, Pengfei Fang, and Zhijun Zhang. Brain-conditional multimodal synthesis: A survey and taxonomy. *arXiv preprint arXiv:2401.00430*, 2023.
- [28] Yu Takagi and Shinji Nishimoto. High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14453–14463, 2023.
- [29] Paul S Scotti, Mihir Tripathy, Cesar Kadir Torrico Villanueva, Reese Kneeland, Tong Chen, Ashutosh Narang, Charan Santhirasegaran, Jonathan Xu, Thomas Naselaris, Kenneth A Norman, et al. Mindeye2: Shared-subject models enable fmri-to-image with 1 hour of data. *arXiv preprint arXiv:2403.11207*, 2024.
- [30] Hamad Ahmed, Ronnie B Wilbur, Hari M Bharadwaj, and Jeffrey Mark Siskind. Object classification from randomized eeg trials. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3845–3854, 2021.
- [31] Dongjun Liu, Weichen Dai, Hangkui Zhang, Xuanyu Jin, Jianting Cao, and Wanzeng Kong. Brain-machine coupled learning method for facial emotion recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10703–10717, 2023.
- [32] Prajwal Singh, Dwip Dalal, Gautam Vashishtha, Krishna Miyapuram, and Shanmuganathan Raman. Learning robust deep visual representations from eeg brain recordings. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7553–7562, 2024.
- [33] Changde Du, Kaicheng Fu, Jinpeng Li, and Huiguang He. Decoding visual neural representations by multimodal learning of brain-visual-linguistic features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [34] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 776–794. Springer, 2020.
- [35] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.
- [36] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.
- [37] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- [38] Zesheng Ye, Lina Yao, Yu Zhang, and Sylvia Gustin. See what you see: Self-supervised cross-modal retrieval of visual stimuli from brain activity. *arXiv preprint arXiv:2208.03666*, 2022.

- [39] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *The Sixth International Conference on Learning Representations (ICLR)*, 2018.
- [40] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *The Tenth International Conference on Learning Representations (ICLR)*, 2022.
- [41] Liang She, Marcus K Benna, Yuelin Shi, Stefano Fusi, and Doris Y Tsao. Temporal multiplexing of perception and memory codes in it cortex. *Nature*, pages 1–8, 2024.
- [42] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.
- [43] Xin-Yao Huang, Sung-Yu Chen, and Chun-Shu Wei. Enhancing low-density eeg-based brain-computer interfacing with similarity-keeping knowledge distillation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023.
- [44] Martin N Hebart, Adam H Dickter, Alexis Kidder, Wan Y Kwok, Anna Corriveau, Caitlin Van Wicklin, and Chris I Baker. Things: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PloS one*, 14(10):e0223792, 2019.
- [45] RR Selvaraju, M Cogswell, A Das, R Vedantam, D Parikh, and D Batra. Grad-cam: visual explanations from deep networks via gradient-based localization. *arxiv [cs. cv]*, 2016.
- [46] James J DiCarlo and David D Cox. Untangling invariant object recognition. *Trends in cognitive sciences*, 11(8):333–341, 2007.
- [47] Wolfgang Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2-3):169–195, 1999.
- [48] Kyle E Mathewson, Alejandro Lleras, Diane M Beck, Monica Fabiani, Tony Ro, and Gabriele Gratton. Pulsed out of awareness: Eeg alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. *Frontiers in psychology*, 2:99, 2011.
- [49] Pascal Fries, John H Reynolds, Alan E Rorie, and Robert Desimone. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291(5508):1560–1563, 2001.

A Appendix

A.1 The model abbreviations details

The abbreviations detail is shown as Table 4.

Table 4: The detail of all the model

Method	EEG Encoder	Image Encoder	Loss Function
BraVL [33]	MLP	MLP	ELBO
NICE [20]	TSCConv	CLIP-ViT	InfoNCE
NICE-SA [20]	TSCConv-SA	CLIP-ViT	InfoNCE
NICE-GA [20]	TSCConv-GA	CLIP-ViT	InfoNCE
MUSE (ours)	STConv	CLIP-ViT	InfoNCE
MUSE-GA (ours)	STConv-GA	CLIP-ViT	InfoNCE
MUSE-Nerv (ours)	NervFormer	CLIP-ViT	InfoNCE
MUSE-Nerv-GA (ours)	NervFormer-GA	CLIP-ViT	InfoNCE
MUSE-SK (ours)	STConv	CLIP-ViT	SK-InfoNCE
MUSE-SK-GA (ours)	STConv-GA	CLIP-ViT	SK-InfoNCE
MUSE-SK-Nerv (ours)	NervFormer	CLIP-ViT	SK-InfoNCE
MUSE-SK-Nerv-GA (ours)	NervFormer-GA	CLIP-ViT	SK-InfoNCE

A.2 Graph Attention

In line with Graph Attention Networks (GATs) principles, we employ the Graph Attention (GA) module to iteratively refine the state of each node, conceptualized as electrodes, by leveraging the states of all other nodes [39, 40]. Through these mechanisms, the GA module dynamically adjusts the importance of each node based on the contextual information proffered by its neighbors, ensuring an attention-weighted update that underscores the interconnectivity of node features within the graph’s architecture. Each node’s representation is denoted by $n_i \in \mathbb{R}^{1 \times T}$, indexed by i for $i = 1, \dots, ch$, signifying an electrode that establishes connections with a defined set \mathcal{N}_i of adjacent nodes, thus forming a fully connected graph. The update mechanism for an individual node n_i is formalized as:

$$n'_i = \alpha_{i,i} W n_i + \sum_{j \in \mathcal{N}_i} \alpha_{i,j} W n_j \quad (4)$$

where n'_i designates the updated node, $\alpha_{i,j}$ encapsulates the attention coefficients indicative of the feature significance from node j to node i , and W is the weight matrix of the linear transformation. The attention coefficients $\alpha_{i,j}$ are computed via the equation:

$$\alpha_{i,j} = \frac{\exp(a^T \cdot \text{LeakyReLU}(W[n_i \| n_j]))}{\sum_{k \in \mathcal{N}_i \cup \{i\}} \exp(a^T \cdot \text{LeakyReLU}(W[n_i \| n_k]))} \quad (5)$$

In this expression, $a \in \mathbb{R}^{2T}$ represents the weight vector of a feedforward attention mechanism, $(\cdot)^T$ indicates the transpose operation, and $\|$ signifies concatenation. LeakyReLU is introduced as the non-linear function with a negative slope coefficient of 0.2, facilitating computational stability and non-linearity.

A.3 Time-Frequency Dynamics Analysis

We took the best SK model, MUSE-SK, to perform time-frequency analysis and found that the alpha wave, gamma wave, and theta wave signals were concentrated on the occipital and parietal lobes in both the training and testing topomaps. This finding aligns with medical literature, where the alpha wave is associated with visual attention [47, 48], and the gamma wave is related to higher cognitive functions, attention, and visual processing [49]. This also indicates that our designed model has indeed learned some neural behaviors related to the human brain.

A.4 Limitation

In our framework, we have not changed the image encoder to the more powerful CLIP, but we focus on comparing different EEG encoders under the same image encoder and the reliability of our proposed brain-



Figure 9: Time-Frequency map of MUSE-SK on averaging all of subject 10’s training trials. We can see that the MUSE-SK can focus on alpha band and gamma band, where is related to vision attention and high-level visual recognition in neural science.

inspired similarity-keeping framework. After demonstrating that this work can indeed improve the performance of contrastive learning, replacing the image encoder with a more powerful one would be a better direction.

A.5 Table of Testing Object Categories

We also try to use Grad-CAM method doing model interpretation on testing sets with our-selected category.

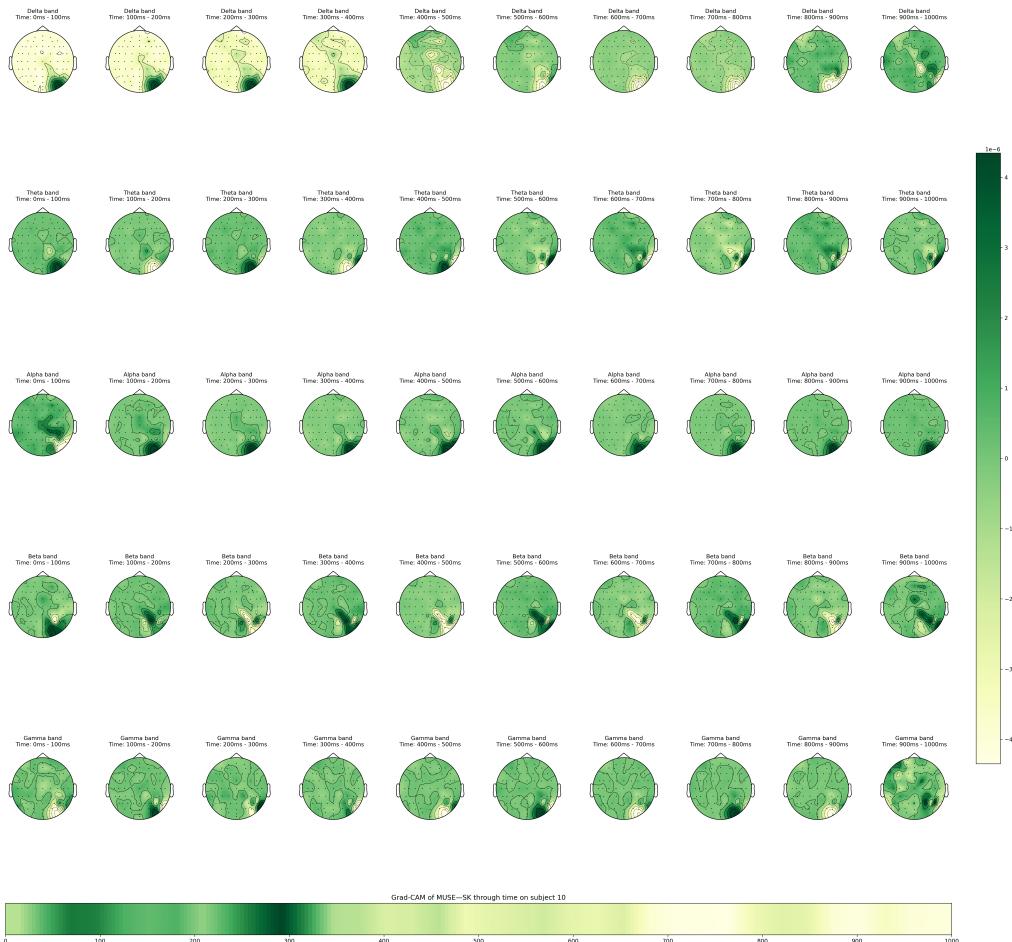


Figure 10: Time-Frequency map of MUSE-SK on averaging all trials in the testing set of subject 10. It is evident that MUSE-SK focuses on the alpha and gamma bands, which are associated with visual attention and high-level visual recognition in neuroscience.

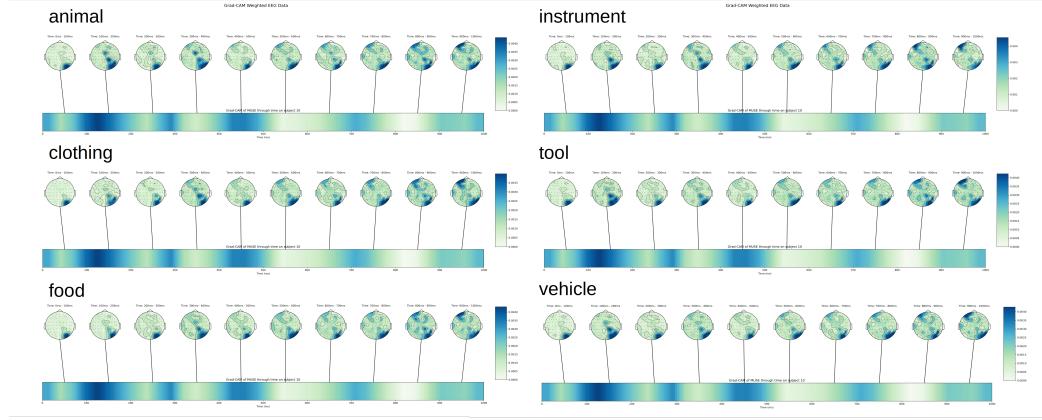


Figure 11: MUSE model interpretation on our-selected category.

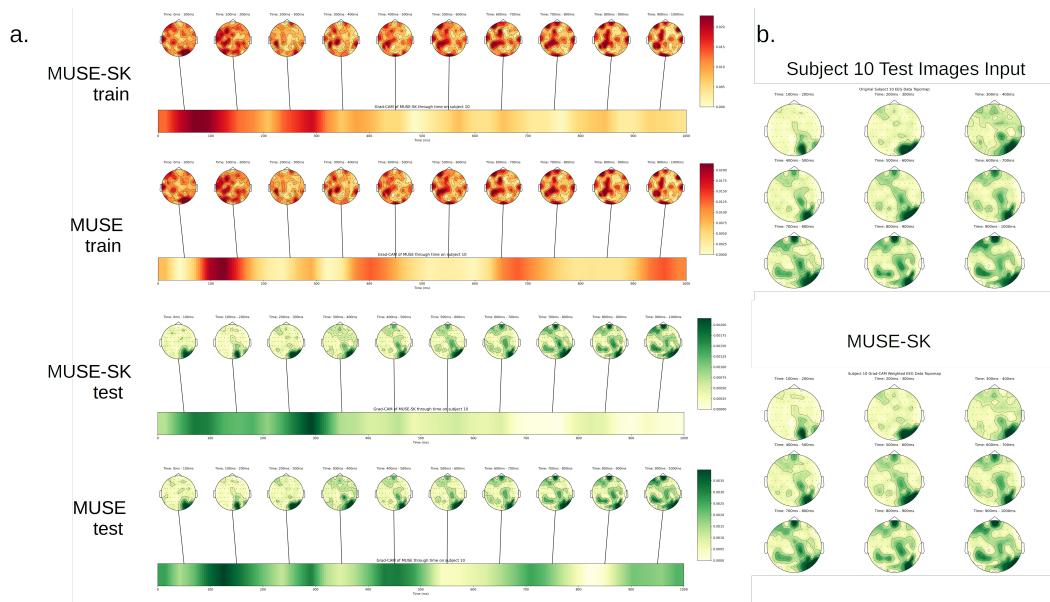


Figure 12: Topomap of each 100 ms by one trial averaging through all the repetition on subject 10. (a.) On MUSE-SK and MUSE models, the color bar on the bottom is the Grad-CAM of each model through time. Most of the model focus on the 100-500ms. The (b.) Zoom-in and compare the input EEG data and the MUSE-SK, can see that the model can more focus on temporal and occipital areas.

Table 5: Test images on THINGSEEG dataset categories

Category	Items
animal	00002_antelope, 00012_beaver, 00024_bug, 00033_cat, 00034_caterpillar, 00039_cheetah, 00046_cobra, 00053_crab, 00058_crow, 00063_dalmatian, 00065.dragonfly, 00069_eagle, 00070_eel, 00072_elephant, 00076_flamingo, 00086_goose, 00087_gopher, 00088_gorilla, 00089_grasshopper, 00097_hummingbird, 00106_lamb, 00110_lightning_bug, 00111_manatee, 00117_mosquito, 00127_ostrich, 00129_panther, 00133_pheasant, 00136_pigeon, 00137_piglet, 00142_possum, 00144_pug, 00150_rhinoceros, 00152_rooster, 00161_seagull, 00183_tick, 00190_turkey
clothing	00019_bonnet, 00037_chaps, 00043_cleat, 00045_coat, 00052_coverall, 00074_face_mask, 00083_glove, 00094_headscarf, 00096_hoodie, 00104_kneepad, 00107_lampshade, 00128_pajamas, 00138_pocket, 00155_sandal, 00169_snowshoe, 00176_suit, 00177_t-shirt, 00182_tiara, 00187_top_hat, 00189_tube_top
instruments	00009_bassoon, 00041_chime, 00067_drum, 00080_french_horn, 00119_music_box, 00149_recorder
food	00005_banana, 00007_basil, 00011_batter, 00015_birthday_cake, 00018_bok_choy, 00022_bread, 00027_bun, 00029_calamari, 00032_cashew, 00038_cheese, 00047_coconut, 00048_coffee.Bean, 00050_cookie, 00051_cordon_bleu, 00054_creme_brulee, 00055_crepe, 00057_croissant, 00060_crumb, 00061_cupcake, 00064_dessert, 00071_egg, 00073_espresso, 00081_fruit, 00082_garlic, 00091_hamburger, 00098_ice_cube, 00101_jelly.Bean, 00109_lettuce, 00112_marijuana, 00113_meatloaf, 00120_mussel, 00122_okra, 00123_omelet, 00124_onion, 00125_orange, 00126_orchid, 00131_pear, 00132_pepper1, 00135_pie, 00140_popcorn, 00141_popsicle, 00143_pretzel, 00147_radish, 00148_raspberry, 00157_sausage, 00158_scallion, 00159_scallop, 00162_seaweed, 00163_seed, 00174_strawberry, 00184_tomato_sauce, 00195_walnut, 00196_wheat, 00199_wine
tool	00003_backscratcher, 00006_baseball_bat, 00016_blowtorch, 00020_bottle_opener, 00021_brace, 00023_breadbox, 00026_bullet, 00030_candlestick, 00035_cd_player, 00042_chopsticks, 00044_cleaver, 00049_coffeemaker, 00062_dagger, 00078_fork, 00079_freezer, 00090_grenade, 00092_hammer, 00093_handbrake, 00103_kettle, 00105_ladle, 00114_metal_detector, 00118_muff, 00130_paperweight, 00134_pickax, 00139_pocketknife, 00145_punch2, 00168_slingshot, 00170_spatula, 00171_spoon, 00173_stethoscope, 00185_tongs, 00186_tool, 00192_vise, 00197_wheelchair, 00200_wok
vehicle	00001_aircraft_carrier, 00014_bike, 00017_boat, 00025_buggy, 00031_cart, 00059_cruise_ship, 00075_ferry, 00084_golf_cart, 00085_gondola, 00100_jeep, 00115_minivan, 00154_sailboat, 00160_scooter, 00164_skateboard, 00165_sled, 00172_station_wagon, 00175_submarine, 00191_unicycle
other	Other categories in test images.

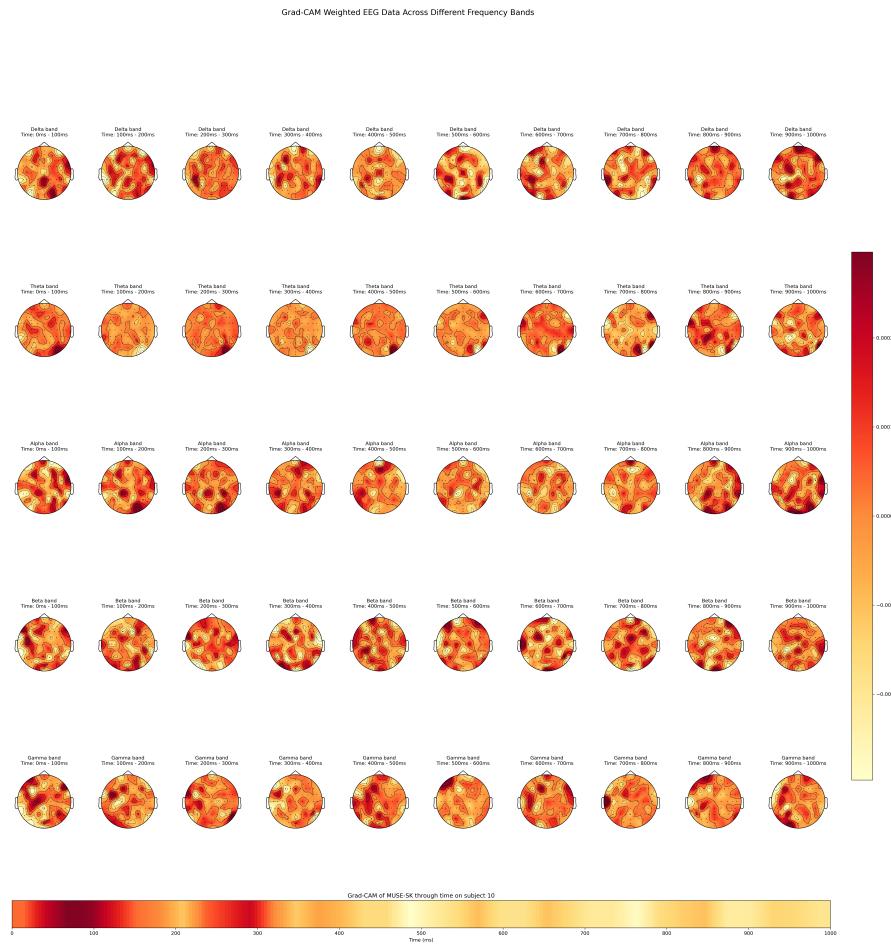


Figure 13: Time-Frequency map of MUSE-SK on one of subject 10's training trial. We can see that the MUSE-SK can focus on alpha band and gamma band, where is related to vision attention and high-level visual recognition in neural science.



Figure 14: Time-Frequency map of MUSE-SK on one trial in the testing set of subject 10. It is evident that MUSE-SK focuses on the alpha and gamma bands, which are associated with visual attention and high-level visual recognition in neuroscience.