Latin American Data in Science, v.2, n.2
datainscience.com.br
ISSN: 2763-9290

EEG data for motor imagery based brain-computer interface using low-cost equipment.

Recebido: 17/03/2022 | Aceito: 13/03/2023 | Publicado: 15/05/2023
https://doi.org/10.53805/lads.v2i2.49

Gabriel Henrique de Souza[1]*, Gabriel O. M. Faria[1], Luciana P. Motta[1], Heder S. Bernadino[1], Alex B. Vieira[1].

ABSTRACT

EEG-based brain-computer interfaces (BCI) for motor imagery recognition can be used in many applications, including prosthesis control, post-stroke motor rehabilitation, communication, and videogames. Such BCIs usually need to be calibrated with EEG data before being used. The calibration can use data from either a single person, the same person who will use the equipment, or a group of different people. However, although BCIs are increasingly used in research and real-world problems, high equipment costs prevent their popularization in personal use applications. For this reason, there are many ongoing efforts to create more affordable BCI devices. Nevertheless, most public datasets for motor imagery EEG-BCIs still use expensive equipment. Therefore, our work presents a dataset for EEG-based motor imagery BCIs focused on personal use applications. Using a low-cost 16-electrode EEG OpenBCI Cyton+Daisy Biosensing Board, we recorded the brain signals of 6 subjects while they imagined the movements of their hands, resulting in a dataset containing 960 trials of left and right-hand motor imagery. This dataset can be used to calibrate BCIs using similar low-cost equipment as well as study the signals generated by such equipment.

Keywords: Brain-Machine Interface; Electroencephalography; Brain-Computer Interface; Motor Imagery; Low-cost equipment.

[1] Universidade Federal de Juiz de Fora, Juiz de Fora, Brasil. gabriel.souza@engenharia.ufjf.br.

DATA IMPORTANCE

- The data were collected using low-cost EEG equipment. Most datasets in the literature were collected using high-cost equipment. However, low-cost equipment is more suitable for real-world applications. Therefore, this dataset can be used to evaluate the effectiveness of a BCI model in such cases.
- Motor imagery EEG data can be used to evaluate different algorithms for brain-computer interfaces. Thus, these data can be used to develop any BCI that uses the motor imagery paradigm.
- The data can be used for cross-subject BCIs, enabling the creation of subject-independent BCIs.
- The data can be used to analyze the laterality of EEG patterns during motor imagery tasks.

MATERIALS AND METHODS

The experiment aimed to collect EEG signals of left and right-hand motor imagery. The electroencephalogram (EEG) equipment used to record the brain signals was the OpenBCI Cyton+Daisy Biosensing Board (16-Channels) with a signal acquisition frequency of 125 Hz. The EEG equipment is shown in Figure 1.

16 electrodes were placed in the regions Fp1, Fz, C3, C4, P7, P8, Cz, Pz, F7, F8, F3, F4, T7, T8, P3, P4, where the AFz position was used as the bias, and the CPz position as the reference. These positions follow the 10-10 system for EEG, which is illustrated in Figure 2.

Figure 1. OpenBCI Cyton+Daisy Biosensing Boards (16-Channels).

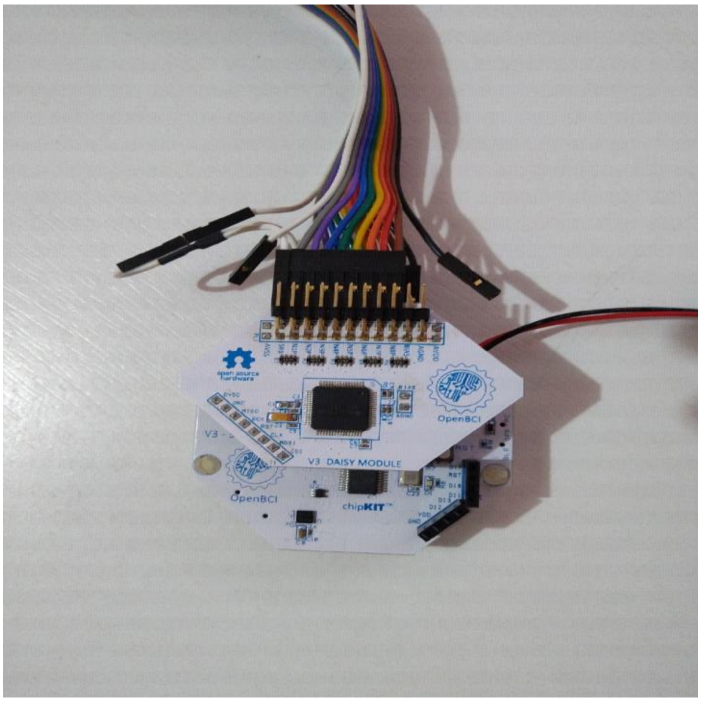
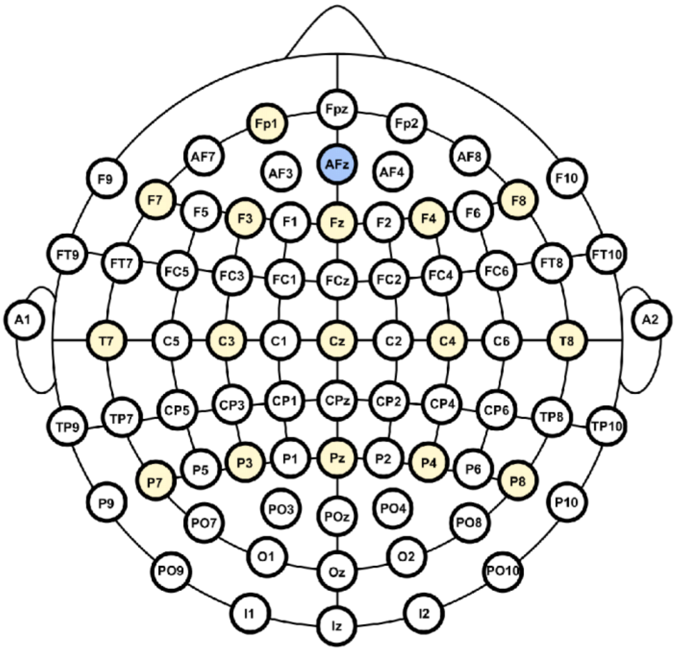


Figure 2. The 10-10 system for EEG. The recorded electrodes are highlighted in yellow, while the reference electrode is highlighted in blue (adapted from SOUZA, 2003).



We collected motor imagery brain signals from 6 subjects. The collection of these data was approved by the ethical committee of Universidade Federal de Juiz de Fora under the number 47866121.5.0000.5147. During the experiments, we instructed each subject to imagine the movement of opening and closing the left or the right hand, while adhering to the following experimental protocol:

1. We hand the informed consent form to the subject, in which the subject agrees with all stages of the experiment.
2. We fill in the subject’s year of birth, sex, and handedness in the recording software.
3. The subject sits down on a chair in front of a laptop, which presents the actions the subject must perform during recording.
4. We place the EEG equipment on the subject's head and apply conductive gel on the equipment.
5. We instruct the subject to move as little as possible and avoid excessive blinking during the experiment, as the equipment is sensitive to noise.
6. The subject executes 4 runs with 40 motor imagery trials each (20 for each hand) while receiving instructions on the screen. Between runs, the subject is given a resting period until he or she feels comfortable enough to resume the experiment. Within a run, each trial follows the protocol presented in Figure 3, which consists of the following steps:
   a. A white cross is displayed at the center of the screen during 3 seconds to maintain the subject’s focus, with a warning beep being played at 2 seconds.
   b. A white arrow pointing either left or right is presented for 2 seconds to indicate the movement to be imagined.
   c. The arrow becomes red and stays on the screen for 4 seconds. The subject is instructed to imagine the indicated movement during this period, as shown in Figure 4.
   d. The screen goes blank and the subject rests for 1 second before the next trial begins.

Figure 3. Experimental protocol

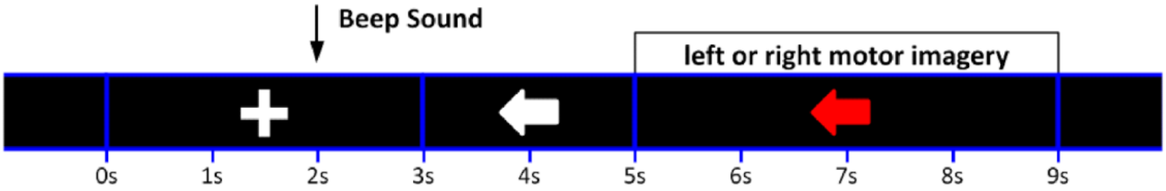


Figure 4. A subject imagining the movement of his right hand.

DATA DESCRIPTION

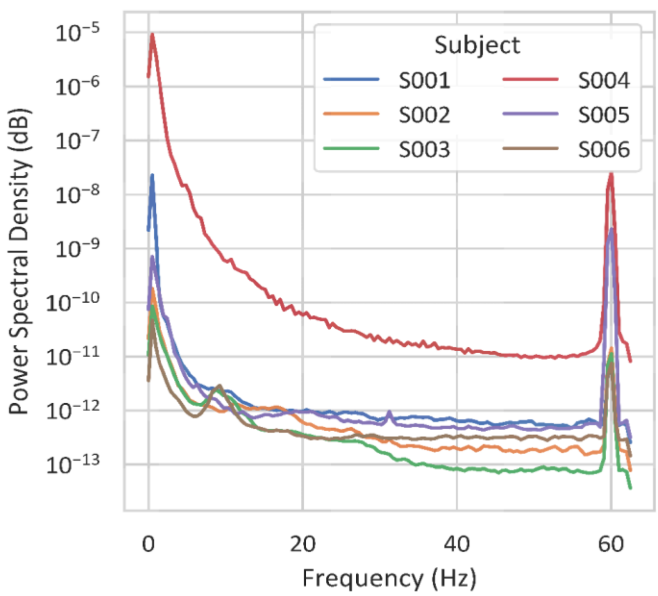
The data comprise EEG recordings from 6 healthy subjects according to the acquisition setup previously described. Each subject's data contain 4 runs with 40 trials each, totaling 160 trials (80 for each motor imagery task). A summary of the subjects who participated in the experiments is shown in Table 1.

To visualize the data, we calculate the average power spectral density (PSD) of the EEG signals over electrodes and trials using Welch's method. Importantly, we disregard signals from resting periods in the calculation. The results are shown in Figure 5. We can see that the PSD of the subjects consistently peaks at 60 Hz. This effect occurs due to powerline interference during recording and can be mitigated using a notch filter. We can also see that the PSD values of subject 004 are much larger than the others, which indicates that this subject's signals contain more noise.

Table 1. Summary of the recorded subjects. The subjects were aged 32.5 ± 9.6, from which 67% were male, 33% were female, 50% were left-handed, and 50% were right-handed.

Subject Code	Age	Sex	Handedness
001	41	Male	Right
002	44	Male	Right
003	24	Male	Left
004	29	Female	Left
005	37	Female	Left
006	20	Male	Right

Figure 5. Mean power spectral density (over electrodes and trials) of each subject in the dataset.



Dataset

The dataset consists of 6 files, each containing a subject's EEG recordings along with additional information. The files are stored in the European Data Format (EDF) (KEMP; OLIVAN, 2003), the de-facto standard for EEG data, which is supported across multiple programming languages and software tools.

Specifically, each EDF file in our dataset can be divided into three sections of data: EEG recordings, event annotations, and metadata, as Figure 6 illustrates. We describe each section below.

- **EEG recordings**: correspond to all EEG signals recorded during the experiment. These signals can be read into a $C \times T$ matrix M, where C is the number of electrode channels and T is the number of time samples recorded. Each element m_{ct} of the matrix is a number representing the electrical potential difference, measured in microvolts, between an electrode c and the reference electrode CPz at a time t.

- **Event annotations**: indicate at which moments of the recording an event occurred, delimiting the EEG signals. Our dataset includes the following events:

NewRun: indicates that a new run is starting.

NewTrial: indicates that a new trial is starting. This event corresponds to the moment the white cross appears on the screen.

LeftPrepare and *RightPrepare:* indicate that the preparation period for either left or right-hand motor imagery is starting. These events correspond to the moment the white arrow appears on the screen.

LeftExec and *RightExec*: indicate that the motor imagery period for either left or right-hand movement is starting. These events correspond to the moment the arrow turns red.

Resting: indicates that a resting period is starting. This event corresponds to the resting period between either trials or runs.

- **Metadata**: comprise information about the experiment, the subject, and the signals. Our files provide the following metadata:

Patient code: a string value containing the code of the subject (the same shown in Tab. 1).

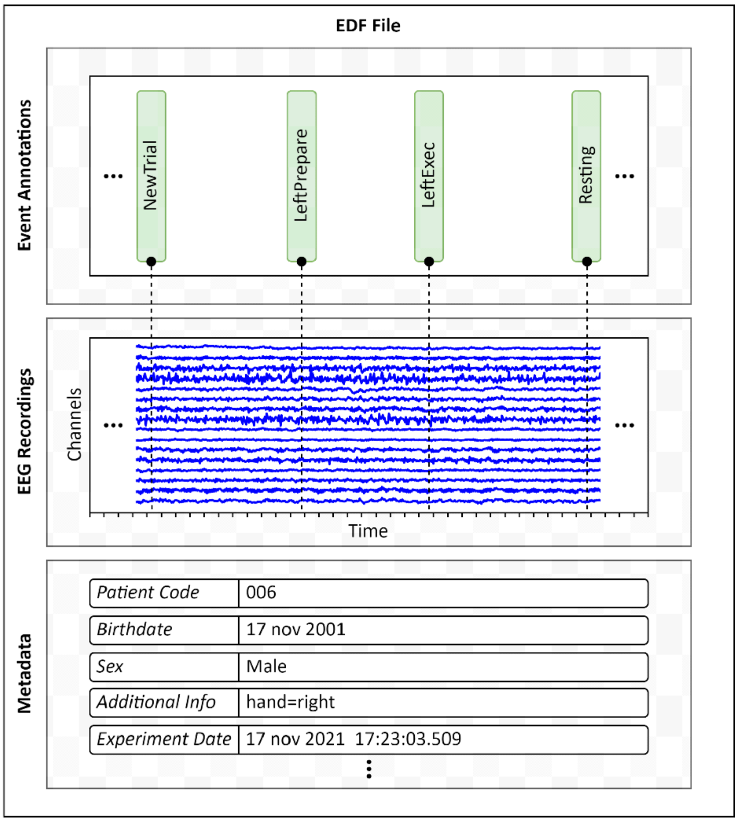
Birthdate: the date of birth of the subject. For the sake of anonymity, we set the subject's birth to the experiment date minus the subject's age.

Sex: the sex of the subject. The possible values of this field are "male" and "female".

Additional info: a string value containing arbitrary information. In particular, we use this field to indicate the subject's handedness. In our dataset, this value is formatted as "hand=·", where · may assume the values "right", "left", or "both".

Experiment date and time: a value that represents the date and time when the recording was started (UTC).

Figure 6. A simplified scheme of an EDF file. The information shown in the image belongs to subject 006.



SUPPLEMENTARY MATERIALS

001.edf
002.edf
003.edf
004.edf
005.edf
006.edf

ACKNOWLEDGEMENTS

The authors thank the financial support provided by CAPES, CNPq, FAPEMIG, FAPESP, OpenBCI, and UFJF.

REFERENCES

KEMP, B.; OLIVAN, J. European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data. Clinical Neurophysiology, [S. l.], v. 114, n. 9, p. 1755–1761, 2003. DOI: 10.1016/S1388-2457(03)00123-8.

SOUZA, G. H.; BERNARDINO, H. S.; VIEIRA, A. B. Single Electrode Energy on Clinical Brain–Computer Interface Challenge. Biomedical Signal Processing and Control, v. 70, n. 9, p. 102993, 2021. DOI: 10.1016/j.bspc.2021.102993