MetroXRAINE 2022 competition - dataset A (Motor Imagery)

Pasquale Arpaia[1,2], Antonio Esposito[1], and Angela Natalizio[3]

[1] Department of Electrical Engineering and Information Technology (DIETI), University of Naples Federico II, Naples, Italy
[2] Interdepartmental Center for Research on Management and Innovation in Healthcare (CIRMIS), University of Naples Federico II, Naples, Italy
[3] Department of Electronics and Telecommunications (DET), Politecnico di Torino, Turin, Italy

1 Experimental Paradigm

The dataset contains the EEG data of 27 subjects. Five sessions were recorded for each participant. Each session lasted about one hour. A synchronous brain-computer interface (BCI) paradigm was adopted, i.e. the user had to imagine according to an external pace. It consisted of two different motor imagery tasks, namely the imagination of movement of the left hand (class 1), and right hand (class 2). Each session consisted of 6 runs separated by short breaks.

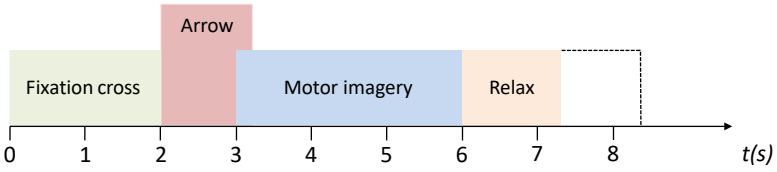

Fig. 1: Timing scheme of the synchronous BCI paradigm for the pure motor imagery phase.

A single run consisted of 30 randomized trials (15 per class), resulting in a total of 180 trials per session. The participants were sitting in a comfortable chair in front of a computer screen.

Participants were divided into two groups: the control group and the neurofeedback group. The control group exclusively engaged in pure motor imagery during each session. In contrast, the neurofeedback group participated in two distinct phases. First, they performed pure motor imagery to train the algorithm, followed by an online phase where EEG signals were used to provide the

feedback as a consequence of their mental task. In both cases, a synchronous paradigm was employed, requiring participants to imagine movements based on external cues.

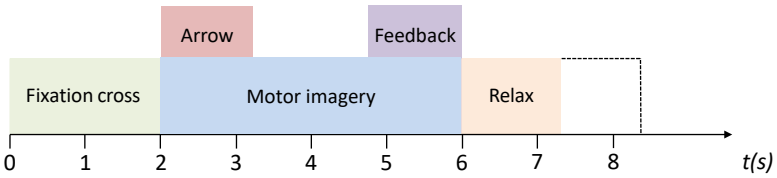

Fig. 2: Timing scheme of the synchronous BCI paradigm for the online phase.

The timing scheme for the pure motor imagery phase is illustrated in Fig. 1.For each trial, a fixation cross from 0.00 s to 2.00 s, a left or right arrow from 2.00 s to 3.25 s, the textual indication "GO!" up to 6.00 s, and the textual indication "RELAX" disappearing randomly between 7.00 s and 8.00 s were shown sequentially. During the the arrow display, participants were asked to prepare for the motor imagery of the right or left hand depending on the orientation of the arrow. Then, they had to actually imagine the movement when the "GO!" was shown. The timing of the relax at the end was randomised to avoid any bias between consecutive trials. Three or six of such runs of pure motor imagery were recorded for the neurofeedback and the control groups, respectively.

Subsequently, the participants of neurofeedback group performed three further runs, during which they received online multimodal feedback. It consisted of a combination of visual and haptic feedback: the visual feedback consisted of a ball rolling left or right on a virtual floor; the haptic feedback was provided by means of a wearable haptic vest and the vibration started from the center of the torso (front side) and it could be moved to the left or to the right in accordance with the movement of the virtual ball. As illustrated in Fig. 2, the timing was slightly adjusted for the online phase. Participants were instructed to start motor imagery upon cue appearance, and feedback was provided from 4.50 s to 6.00 s.

Further details regarding the experimental protocol are available in [1].

2 Data recording

The EEG signals were recorded through the *ab medica Helmate*. The device is class IIA certified according to the EU regulation 2017/745.

The headset is equipped with 10 dry electrodes made of conductive-rubber and coated with Ag/AgCl. Channel placement follows the 10/20 International

(a) EEG cap

(b) Dry electrodes

Fig. 3: EEG acquisition system.

Positioning System and the provided positions are: Fp1, Fp2, Fz, Cz, C3, C4, O1, and O2. AFz is the reference electrode while Fpz is the ground. The electrode placement is shown in Fig. 4.

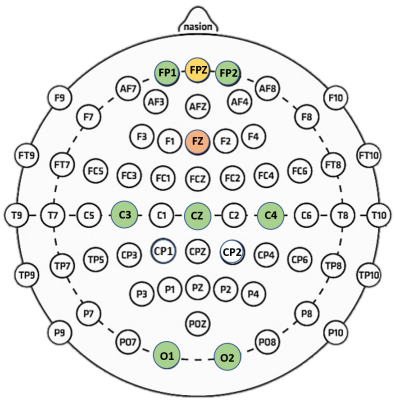

Fig. 4: Electrodes placement

Electrodes have different shapes to reach the skin passing through the hair. The EEG acquisition system and the dry electrodes are shown in Fig.3a and Fig.3b, respectively.

The Helmate integrates the ADS1298 analog front-end from Texas Instruments, with a multi-channel simultaneous sampling, and a 24-bit $\Delta\Sigma$ analog-to-digital converter (ADC). Analog signals are analog filtered and amplified with a nominal pass-band from 0.2 Hz to 70 Hz with the 50 Hz notch filter enabled. The EEG signals are acquired at a sampling frequency of 512 Sa/s. The device

has a rechargeable battery and a Bluetooth connection for data transmission. The Helm8 software manager allows to check the contact impedance between the electrodes and the scalp and the real-time visualization of the EEG data. It also allows a simple pre-processing of the EEG signals.

3 Data file description

Data are stored with the *.mat* extension. For each subject, a folder contains six *.mat* files related with the sessions 1 to 5.

As previously specified, the participants were divided into two groups: the control group and the neurofeedback group. Assignment to these groups can be determined on the basis of the naming of the data files for each session. In particular, subjects with files containing both "B0" and "B1" belong to the neurofeedback group. Those with files containing only "B0" belong to the control group. The grouping of subjects can also be determined using the following ordered vector for subjects 1 to 27:

$$grouping = [0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1]; \quad (1)$$

where subjects belonging to the control group are labeled with "0", while subjects in the neurofeedback group are designated as "1".

Once the *.mat* file is loaded, the workspace will contain a cell "data" and a structure "notes" with meta-information.

The data cell contains six structures associated with the six runs acquired for each session. Each structure contains:

- X: the EEG data stream given as a $N \times M$ matrix, where N is the number of channels and M is the number of samples;
- *trial*: a vector of length 30 containing the starting sample per each trial;
- y: a vector of length 30 containing the labels per each trial;
- *fs*: the sampling frequency;
- *classes*: a 1×2 cell containing the motor imagery task associated to the two possible labels (1 and 2);
- *artifacts*: a logical vector of length 30. It flags artifact marked by visual inspection. Specifically, 0 corresponds to a clean trial and 1 corresponds to a trial containing an artifact.

Meanwhile, the "notes" structure contains general information like the subject ID, the gender, the age of the subject, the session number, a 1×8 cell containing the name of the channels (sorted according to the rows of the matrix X), reference and ground electrode positions, and the date on which the signals were acquired.

Finally, it is worth noting that due to technical issues, the following data is not available:

- the second session of subject 2;

– the third run of fourth session of subject 2;
– the fifth session of subjects 6, 17, and 27;
– the sixth run of fifth session of subject 18.

Cite

If you find useful the EEG data provided by this dataset, please add [1] to your publications.

References

1. P. Arpaia, D. Coyle, A. Esposito, A. Natalizio, M. Parvis, M. Pesola, and E. Vallefuoco, "Paving the way for motor imagery-based tele-rehabilitation through a fully wearable bci system," *Sensors*, vol. 23, no. 13, p. 5836, 2023.