

Data sets 1 *«motor imagery, uncued classifier application»*

Data sets provided by the [Berlin BCI](#) group: Berlin Institute of Technology ([Machine Learning Laboratory](#)) and Fraunhofer FIRST ([Intelligent Data Analysis Group](#)) ([Klaus-Robert Müller](#), [Benjamin Blankertz](#), Carmen Vidaurre, [Guido Nolte](#)), and Campus Benjamin Franklin of the Charité - University Medicine Berlin, Department of Neurology, Neurophysics Group (Gabriel Curio)

Correspondence to Benjamin Blankertz benjamin.blankertz@tu-berlin.de

The Thrill

Most demonstrations of algorithms on BCI data are evaluating classification of EEG trials, i.e., windowed EEG signals for fixed length, where each trial corresponds to a specific mental state. But in BCI applications with asynchronous feedback one is faced with the problem that the classifier has to be applied continuously to the incoming EEG without having cues of when the subject is switching her/his intention. This data set poses the challenge of applying a classifier to continuous EEG for which no cue information is given.

Another issue that is addressed in this data set is that the evaluation data contains periods in which the user has no control intention. During those intervals the classifier is supposed to return 0 (no affiliation to one of the target classes).

Experimental Setup

These data sets were recorded from healthy subjects. In the whole session motor imagery was performed without feedback. For each subject two classes of motor imagery were selected from the three classes *left hand*, *right hand*, and *foot* (side chosen by the subject; optionally also both feet).

Calibration data: In the first two runs, arrows pointing left, right, or down were presented as visual cues on a computer screen. Cues were displayed for a period of 4s during which the subject was instructed to perform the cued motor imagery task. These periods were interleaved with 2s of blank screen and 2s with a fixation cross shown in the center of the screen. The fixation cross was superimposed on the cues, i.e. it was shown for 6s. These data sets are provided with complete marker information.

Evaluation data: Then 4 runs followed which are used for evaluating the submissions to the competitions. Here, the motor imagery tasks were cued by soft acoustic stimuli (words *left*, *right*, and *foot*) for periods of varying length between 1.5 and 8 seconds. The end of the motor imagery period was indicated by the word *stop*. Intermitting periods had also a varying duration of 1.5 to 8s. Note that in the evaluation data, there are not necessarily equally many trials from each condition.

Special Feature: Some of the data sets were *artificially generated*. The idea is to have a means for generating artificial EEG signals with specified properties that is such realistic that it can be used to evaluate and compare analysis techniques. The outcome of the competition will show whether the applied methods perform comparably on artificial and real data. The only information we provide is that there is at least one real and at least one artificial data set, while the true distribution remains undisclosed until the submission deadline. For competition purpose, only results for the real data set(s) are considered. The functions for generating artificial data were provided by [Guido Nolte](#) and Carmen Vidaurre.

Format of the Data

Given are continuous signals of 59 EEG channels and, for the calibration data, markers that indicate the time points of cue presentation and the corresponding target classes.

Data are provided in **Matlab** format (*.mat) containing variables:

- `cnt`: the continuous EEG signals, size [time x channels]. The array is stored in datatype `INT16`. To convert it to uV values, use `cnt= 0.1*double(cnt)`; in Matlab.
- `mrk`: structure of target cue information with fields (the file of evaluation data does *not* contain this variable)
 - `pos`: vector of positions of the cue in the EEG signals given in unit *sample*, length `#cues`
 - `y`: vector of target classes (-1 for class one or 1 for class two), length `#cues`
- `nfo`: structure providing additional information with fields
 - `fs`: sampling rate,
 - `clab`: cell array of channel labels,
 - `classes`: cell array of the names of the motor imagery classes,
 - `xpos`: x-position of electrodes in a 2d-projection,
 - `ypos`: y-position of electrodes in a 2d-projection.

As alternative, data is also provided in zipped **ASC II** format:

- `*_cnt.txt`: the continuous EEG signals, where each row holds the values for all channels at a specific time point
- `*_mrk.txt`: target cue information, each row represents one cue where the first value defines the time point (given in unit *sample*), and the second value the target class (-1 for class one or 1 for class two). For evaluation data *no* `*_mrk.txt` file is provided.
- `*_nfo.txt`: contains other information as described for the matlab format.

Requirements and Evaluation

Please provide an ASC II file (named 'Result_BCIC_IV_ds1.txt') containing classifier outputs (real number between -1 and 1) for each sample point of the evaluation signals, one value per line. The submissions are evaluated in view of a one dimensional cursor control application with range from -1 to 1. The mental state of class one is used to position the cursor at -1, and the mental state of class two is used to position the cursor near 1. In the absense of those mental states (intermitting intervals) the cursor should be at position 0. Note that it is unknown to the competitors at what intervals the subject is in a defined mental state. Competitors submit classifier outputs for all time points. The evaluation function calculates the squared error with respect to the target vector that is -1 for class one, 1 for class two, and 0 otherwise, averaged across time points. In the averaging we will ignore time points during transient periods (1s starting from each cue). For competition purpose, only results for the real data set(s) are considered, but results for artificial data are also reported for comparison.

Optionally, please report which of the data sets you think to be artificially generated.

You also have to provide a description of the used algorithm (ASC II, HTML or PDF format) for publication at the results web page.

Technical Information

The recording was made using BrainAmp MR plus amplifiers and a Ag/AgCl electrode cap. Signals from 59 EEG positions were measured that were most densely distributed over sensorimotor areas. Signals were band-pass filtered between 0.05 and 200 Hz and then digitized at 1000 Hz with 16 bit (0.1 uV) accuracy. We provide also a version of the data that is downsampled at 100 Hz (first low-pass filtering the original data (Chebyshev Type II filter of order 10 with stopband ripple 50dB down and stopband edge frequency 49Hz) and then calculating the mean of blocks of 10 samples).

References

Any publication that analyzes this data set should cite the following paper as a reference of the recording:

- **Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Klaus-Robert Müller, and Gabriel Curio.** [The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects](#). *NeuroImage*, 37(2):539-550, 2007. [[pdf](#)]

