Self-paced ERP-based BCI speller conducted over multiple sessions spaced weeks apart (BCI-Self-paced Dataset - A unified framework) – Documentation

− *Gabriel Pires, *Institute of Systems and Robotics of the University of Coimbra* and *Polytechnic Institute of Tomar*, Portugal
− Aniana Cruz, *Institute of Systems and Robotics of the University of Coimbra, Portugal*
− Joana Figueiredo, *Institute of Systems and Robotics of the University of Coimbra, Portugal*
− Urbano J. Nunes, *Institute of Systems and Robotics of the University of Coimbra, Portugal*

Corresponding author: gpires@isr.uc.pt

This document provides a detailed description of a BCI-Self-paced Dataset used to test the combination of self-paced control with "One-Time Calibration" (user specific, but session-independent models)

EXPERIMENT DETAILS

This EEG dataset was recorded while participants performed a P300-based BCI speller (Fig. 1) in an asynchronous self-paced mode. The participant is free to initiate an intentional selection (i.e., enter the control (C) state) at any moment during the continuous sequence of trials. When a participant focuses on a desired character during a trial, the system may detect a P300 response and select the corresponding symbol. Conversely, if the participant remained in the non-control (NC) state for the entire trial, the classifier is expected to return a non-control decision. This dataset supports research on P300 variability across multiple sessions and days, the development of adaptive strategies to mitigate performance degradation over time, and the evaluation of self-paced control paradigms.

Figure 1: Lateral single character (LSC) speller paradigm.

EEG RECORDING AND PARTICIPANTS

Eight able-bodied subjects (S1-S8) (mean age = 23.5 ± 0.8 years) participated in the study. The experimental protocol was approved by the Ethics Committee for Health of the University Hospital Center of S. João, Porto (CHUSJ), and all participants provided written informed consent. Half of the participants (n = 4) had no prior experience using a BCI system.

EEG signals were recorded from 12 electrodes placed according to the extended 10–20 international system at the following sites: Fz, Cz, C3, C4, CPz, Pz, P3, P4, PO7, PO8, POz, and Oz. The signals were acquired using a g.USBamp amplifier (g.tec GmbH, Austria), with the reference electrode placed at the right earlobe and ground at AFz. The EEG signals were sampled at 256 Hz and pre-processed using a notch filter at 50 Hz and a band-pass filter [0.1 - 30] Hz.

SESSIONS

The experiment comprised three sessions conducted on different days:

Session 1 - Calibration and Online session:

Calibration (Trainset1): Participants copied the sentence "THE-SLOW-RED-CAMEL" (18 characters, with '-' representing spaces). Each symbol flashed four times, resulting in 72 target and 1944 non-target samples.

Online Test: The classifier was trained with Trainset1, and then participants performed the online session to write the sentence "THENC-NCREDNC-NCCAMELNC," where "NC" indicated non-control periods (participants fixated on the screen center without attending to any symbol). The sentence was repeated twice, separated by a 2-minute break. In total, 26 control and 10 non-control trials were recorded. Participants were instructed not to correct errors.

Session 2 - Calibration and Online session:

Conducted 1–3 weeks after Session 1. Calibration (Trainset2) was repeated under the same conditions as Trainset1. Online spelling followed the same protocol, but the classifier was trained using data from Session 1 (Trainset1).

Session 3 - Online session:

Conducted 1–3 weeks after Session 2. No calibration was performed; the online session used the classifier trained using data from Session 1 (Trainset1).

INSTRUCTIONS TO USE DATASETS

There are the main MATLAB script and two folders.

The main script is Epoch_extraction.m, a MATLAB script used for extracting EEG epochs and visualizing target, non-target, and non-control events.

Folders
- **Data**: Contains EEG recordings from Sessions 1, 2, and 3.
- **DTA_DBST_toolbox:** Contains helper functions for epoch extraction, signal preprocessing, classification, and performance metric computation.

The EEG data are stored separately for each spelled sentence in MATLAB files. For example, S2_wo_Sess1_sentence3.mat contains training data for Subject 2, Session 1, Sentence 3, while S2_w_Sess1_sentence3.mat contains the corresponding testing data.

Each file (e.g., S*_w_Sess*_ sentence*.mat) contains a big matrix signal with the following fields:

Line 1: Timestamp of each sample.

Line 2: EEG samples recorded from Fz.

Line 3: EEG samples recorded from Cz.

Line 4: EEG samples recorded from C3.

Line 5: EEG samples recorded from C4.

Line 6: EEG samples recorded from CPz.

Line 7: EEG samples recorded from Pz.

Line 8: EEG samples recorded from P3.

Line 9: EEG samples recorded from P4.

Line 10: EEG samples recorded from PO7.

Line 11: EEG samples recorded from PO8.

Line 12: EEG samples recorded from POz.

Line 13: EEG samples recorded from Oz.

Line 14: Contains event labels for stimuli:

- Target stimuli: IDs **60** and **70;**
- Non-target stimuli: IDs **1–28;**
- Non-control events: IDs > **77**.

Acknowledgment