Associated Publication

Crell, M. R. & Müller-Putz, G. R. Handwritten character classification from EEG through continuous kinematic decoding. *Comput. Biol. Med.* **182**, 109132 (2024)
DOI: https://doi.org/10.1016/j.compbiomed.2024.109132

Experimental Paradigm

Twenty participants wrote ten different letters (a,d,e,f,j,n,o,s,t,v) with the index finger of their right hand. Participants were equipped with 64 EEG channels to measure EEG and EOG data. Movements of the right index finger were recorded with a motion capture system which tracked a marker attached to the top of the finger. The paradigm consisted of two rounds comprising the paradigm runs. During each run, the target letter would be faded onto the screen over 2 s, stay on the screen at full opacity for 0.5 s and fade out again over 2 s (see Figure 1). As soon as the letter was faded out completely, the participant would start to draw the letter with their index finger. When they finished writing the letter, they were instructed to stop the movement at the last point and remain there until the next letter faded onto the screen. The next letter would fade onto the screen 4 s after the last letter faded out completely. In every trial except for the first one in every run, participants were instructed to move their hand to a comfortable position in the middle of the monitored area.
In every run, the set of ten letters was written four times by shuffling the set of ten letters randomly and concatenating the shuffled sets. Thus, 40 letters were written per run.

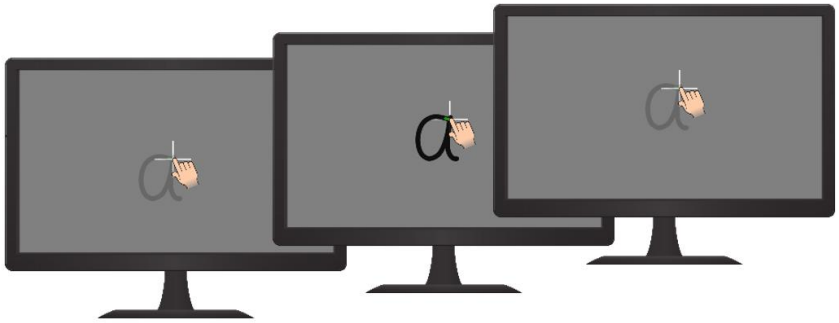

Figure 1 Target letter displayed on the screen during the paradigm. The letter is faded onto the screen during 2 s, stays on the screen at full opacity for 0.5 s and fades out over 2 s. The screen additionally displays a fixation cross at all times.

The complete paradigm consisted of an SGEYESUB run, a training round of the paradigm, a round of seven runs of the paradigm, a longer break, a second SGEYESUB run and a second round of eight paradigm runs.
The SGEYESUB runs were implemented to generate data to train a model for the elimination of eye artifacts from the data. A detailed description of the paradigm can be found in [1].
During the training, participants were instructed to move their hand correctly in four iterations of the ten letters. In the first iteration, a red dot would follow the trajectory of the letter so that participants would know how to write the letters. In the second iteration, participants were instructed to follow the red dot by executing the motion with their index finger. The participant's motion would be displayed on the screen with the image of a hand (see Figure 1) that moved according to their hand. In the third iteration, the red dot would not be visible but feedback about the participant's movement would still be given on the screen and in the last iteration, the paradigm would be executed normally without any visual feedback.
After the training, the paradigm runs would be executed as described above. Between runs, a break of 60 s was established to allow the participant to move and relax. After the seventh run, a

longer break according to the participant's choosing was held, before continuing with another SGEYESUB run and the remaining eight runs of the paradigm.
15 paradigm runs were executed to achieve a total of 60 repetitions per character. The duration of each paradigm run was approximately 340 s with the whole experiment lasting for about 150 min. The complete experiment is also explained in detail in the associated publication.

Dataset

We provide 20 .mat files each containing the recorded data from one participant. Each file comprises a struct for round one paradigm data, round two paradigm data, round one SGEYESUB data and round two SGEYESUB data. The structs contain the following fields:

Round one/two SGEYESUB (*sgeyesub_one*/*sgeyesub_two*)

- *BrainVisionRDA_data*
- *BrainVisionRDA_time*
- *EyeblockStimulus_data*
- *EyeblockStimulus_time*
- *EyeblockMarker_data*
- *EyeblockMarker_time*

Round one/two paradigm (*paradigm_one*/*paradigm_two*)

- *BrainVisionRDA_data*
- *BrainVisionRDA_time*
- *MoCap_data*
- *MoCap_time*
- *ParadigmMarker_data*
- *ParadigmMarker_time*

All arrays are extracted from raw *.xdf* files recorded with LabRecorder from LSL data/marker streams. The *_time* arrays contain $1\ x\ n_{times}$ LSL timestamps associated with the recorded data points in *_data*.

BrainVisionRDA_data contains the recorded EEG data from 64 channels (EOGL1, EOGL2, EOGL3, EOGR1, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PPO1h, PPO2h, PO7, PO3, POz, PO4, PO8, O1, Oz, O2). Data is sampled at $500\ Hz$ and represented in microvolts (µV). For all EEG recordings, the Reference electrode was positioned on the right mastoid of the participant and the Ground electrode at FPz. EOG electrodes were positioned 2cm left (EOGL1) and right (EOGR1) to the left/right eye and above (EOGL2) and below (EOGL3) the left eye.

MoCap_data contains the recorded movement data of the index finger of the participants at a sample rate of 30 Hz. The index finger was equipped with a marker on top of the fingernail of the participant which was tracked by the system. Channel 1 and 2 correspond to x- and y- coordinates of the marker in pixel coordinates, respectively. Channel 3 is a binary flag stating whether the marker was detected (1) or not detected (0), in which case the x/y coordinate of the last detected position was streamed.

ParadigmMarker_data and *ParadigmMarker_time* contain the markers and corresponding timestamps from the paradigm. The markers are encoded numerically according to the following schematic:

- 10: start of the paradigm
- 20: start of the training
- 30: end of the training
- 40: start of the actual paradigm rounds
- 50: start of a single paradigm round
- 60: end of a single paradigm round
- 70: end of the paradigm
- 1: target letter starts to fade in (start of fade-in phase)
- 2: target letter is completely faded in (end of fade-in phase)
- 3: target letter starts to fade out (start of fade-out phase)
- 4: target letter is completely faded out (start of writing phase)
- 5: target letter should be completely written (end of the writing phase)
- 100+x: target letters are numerically encoded as 100 + index of the letter in lowercase ascii (see python-function string.ascii_lowercase.index). For the utilized letters a,d,e,f,j,n,o,s,t,v the numerical values were 100,103,104,105,109,113,114,118,119,121.

EyeblockStimulus_data and *EyeblockMarker_data* are recorded from the streams of the SGEYESUB algorithm paradigm. A detailed description can be found in [1] and the source code is available in [2].

[1] Kobler, R. J., Sburlea, A. I., Lopes-Dias, C., Schwarz, A., Hirata, M. and Müller-Putz, G. R. "Corneo-retinal-dipole and eyelid-related eye artifacts can be corrected offline and online in electroencephalographic and magnetoencephalographic signals.", 218 (2020). https://doi.org/10.1016/j.neuroimage.2020.117000

[2] https://github.com/rkobler/eyeartifactcorrection