Freiburg, Dec. 5, 2016

Information about the Datasets

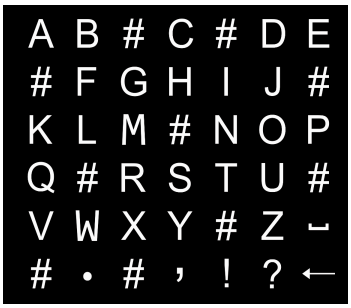
Two studies are available in this data collection. An EEG experiment utilizing visual ERPs with N=13 healthy subjects was conducted in addition to a smaller study with N=5 subjects performing both an auditory and a visual ERP paradigm. They were both recorded at the Brain-State Decoding Lab at University of Freiburg headed by Dr. Michael Tangermann. Details about the study are available in:

> ***Hübner D, Verhoeven T, Schmid K, Müller K-R, Tangermann M and Kindermans P-J. Learning from Label Proportions in Brain-Computer Interfaces: Online Unsupervised Learning with Guarantees. PLOS ONE. 2017. To appear.***

We kindly ask you to cite this paper if you use the data. All subjects gave their written consent and the study was approved by the Ethics Committee of the University Medical Center Freiburg.

Dataset 1: Online Visual ERP Experiment (N=13)

The data of the online study is split into the files '**online_study_1-7.zip**' and '**online_study_8-13.zip**', containing the EEG data of subjects 1-7 and subject 8-13, respectively. The individual datasets are called S1.mat until S13.mat. Data for each subject was bandpassed filtered, sub-sampled to 100 Hz and epochs were extracted. 12852 epochs are available for each subject. The subjects performed a visual spelling task with a spelling matrix containing 42 symbols (see Figure on the right) and where 68 stimuli belong to each character. Further information about the paradigm and pre-processing steps are available in the paper above.



The data can be loaded using Matlab's load() function or Python's scipy.io.loadmat() function. The following fields exist in the data:

Field Name	Dimension	Description
fs	double	**Sampling Frequency** = 100 Hz.
clab	{1x31 cell}	**EEG channels** in the same order as in the .x field
t	[1x90 double]	**Sampling time points** in [ms] in the same order as in .x field
y	[2x12852 logical]	**Label information.** A True / 1 value in the first row indicates a target epoch whereas a True / 1 value in the second row indicates a non-target epoch.
x	[90x31x12852 double]	**Amplitude in µV:** Time Points x Channels x Epochs
stimuli	[42x12852 logical]	**Individual stimuli information.** .stimuli(i,j) == True iff the symbol at position i was highlighted in epoch j. Position count from left to right and then top to bottom, e.g. B is at position 2 and F at position 9.

Dataset 2: Offline Visual and Auditory ERP Experiment (N=5)

The EEG data of the offline study is split into '**offline_auditory.zip**' and '**offline_visual.zip**' containing the auditory and visual epoched data of 5 subjects used for creating aritifical datasets for simulations. Each subject performed both, an auditory as well as a visual task, during one session in an alternating fashion. The auditory paradigm (Figure 1) is a modulation of the AMUSE paradigm by Schreuder et al, 2012 with bisyllabic words used as stimuli. The visual paradigm is a 6 class picture selection task (Figure 2). In both scenarios, the SOA was 250ms and 90 epochs were recorded per selection. Full label information are included.

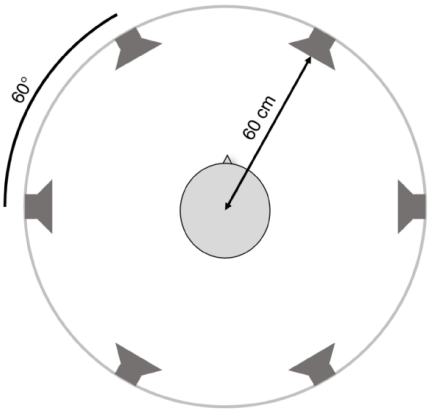


Figure 1: Setup of the auditory paradigm.

Figure 2: Visual stimuli presentation with one highlighted stimulus.

We named the subjects T1 to T5 with suffix _a for an auditory and _v for a visual dataset. The datasets can be loaded with the same functions mentioned above. The same fields are available.

The coding in .stimuli is slightly different, as now numbers from 1-6 directly encode which of the 6 stimuli were played. Here, counting in the auditory case starts from front right and goes clockwise. In the visual case, counting is done as before.

For inquiries, please contact Dr. Michael Tangermann: michael.tangermann@blbt.uni-freiburg.de.