The EEGManyPipelines dataset

Niko A. Busch [1,2]

[1]*Institute of Psychology, University of Münster, Germany*
[2]*Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany*

October 1, 2021

Abstract

This dataset includes data from a study on memory for visual scenes (N=33). EEG data were recorded from 70 channels. The dataset has been only minimally preprocessed. This document provides a description of the paradigm, data recording, and relevant meta-data.

Participants

Thirty-three subjects (22 female; 4 left-handed, mean age 27 years, SD 5.5 years) with normal or corrected-to-normal vision participated in the experiment after providing written informed consent. The study was approved by the ethics committee at the University of Münster.

Apparatus, stimuli, and paradigm

The experiment was written in MATLAB (The Mathworks, Natick, MA, USA) using the Psychophysics Toolbox (Brainard, 1997). Stimuli were presented on a calibrated LCD monitor (VIEWPixx/EEG) with 1920 x 1080 pixels resolution and 100 Hz refresh rate, placed at a distance of 86 cm from the participants' eyes. Head position was stabilized using a chin rest.

The stimuli were 600 images from 4 different scene categories (forests, highways, beaches, buildings) gathered using Google Image Search. All images were 10 x 7.5 $^\circ$ in size and presented in grayscale on black background. A gray fixation cross was displayed at the center of the screen during the inter-stimulus intervals.

Participants performed a continuous recognition task (Shepard and Teghtsoonian, 1961; Friedman, 1990; Isola et al., 2011). In this task, they were presented with a stream of images (Figure 1), some of which were repeated, and they decided for each image whether it was old or new, i.e. whether the presented image had been presented before in the session (old) or whether it was the first time that image was presented (new). Participants were instructed to respond as fast as possible by lifting their
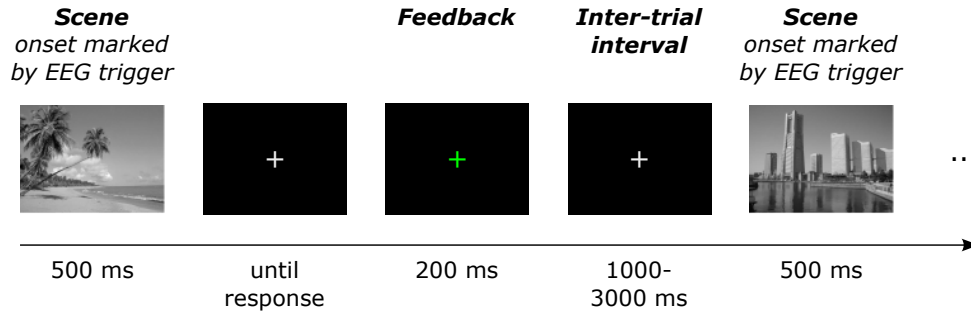
Figure 1: Illustration of the experimental paradigm.

finger from one of two response keys (left/right CTRL key) indicating whether the image was old or new. The assignment of key (left/right) and response category (old/new) was counterbalanced across participants. Images were presented for 500 ms each, followed by an empty screen until the participant had responded. Feedback was provided by turning the fixation cross green or red for correct and incorrect responses, respectively, for a duration of 200 ms. Once both keys were help down again, the next image followed after a variable inter-trial interval, which was drawn from a truncated exponential distribution with a minimum of 1000 ms and a maximum of 2000 ms. The experiment comprised 600 different images; 300 images that were presented only once and another 300 images that were presented three times (first presentation as a new image, second and third presentation as old), resulting in a total of 1200 trials, half of which featured a new image. Image repetitions occurred after a lag of 10 to 60 intervening trials.

EEG Recording and preprocessing

EEG was recorded with a BioSemi Active-Two amplifier system from 64 Ag/AgCl electrodes arranged according to the international 10-10 system and two additional mastoid electrodes (Figure 2). The horizontal and vertical electro-oculograms were recorded from additional electrodes at the lateral canthi of both eyes and below the eyes, respectively. Two additional electrodes located adjacent to electrode POz served as reference and ground. Signals were sampled at 1024 Hz with a 200 Hz low-pass filter.

Data from subjects 1–17 were recorded in an electrically shielded booth, while no shielding was available for subjects 18–33. Thus, stronger power line contamination at 50 Hz is expected for the latter half of subjects.

The dataset we provide here has been only minimally processed, using the following steps:

1. Removal of empty data channels.
2. Re-referencing to channel 30 (POz) – as a result, channel 30 contains only zeros, but is still included in the dataset.
3. Computation of the vertical EOG (channel 71) by subtracting infra-orbital channels IO1 and IO2 (67, 68) from the corresponding super-orbital channels FP1 and FP2 (1, 34) .
4. Computation of the horizontal EOG (channel 72) by subtracting right channel Afp10 (66) from the corresponding left channel Afp9 (65).
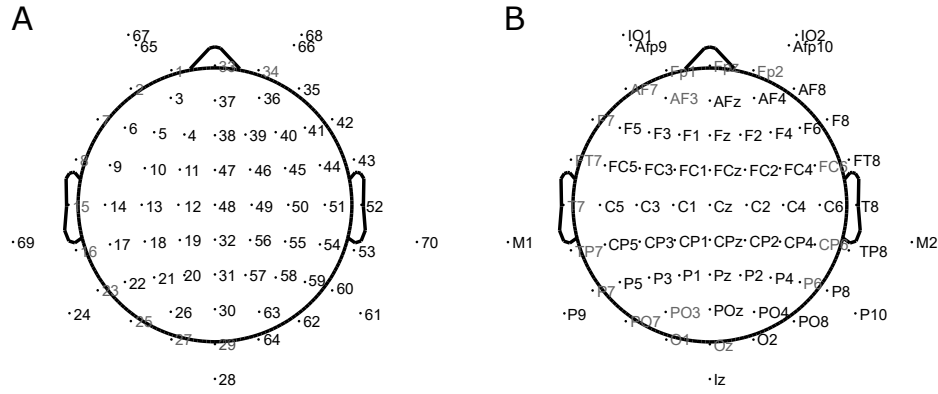
Figure 2: Approximate locations of all 70 EEG channels according to the extended 10-20 system. Note that the vertical EOG (channel 71) and the horizontal EOG (channel 72) do not have associated channel coordinates.

5.   Downsampling to 512 Hz.

6.   Simplification of trigger codes (see below).

Otherwise, the data are provided as recorded. Specifically, they have not been pruned for bad channels, epochs, or subjects.

Data structure

EEG data

We provide the dataset in three different formats: as EEGLAB `.set` files (Delorme and Makeig, 2004), as Brainvision `.dat`, `.vhdr` and `.vmrk` files, and as data files formatted according to the BIDS standard (Pernet et al., 2019) with the actual EEG voltage data included in Brainvision format. Each subject's data are stored in a separate file, approximately 700 MB each. We also provide data in ascii format as comma-separated values (`.csv`) files, where each row represents one sampling point, the first column represents the sampling points' time stamps, and subsequent columns represent the 72 data channels. Given that the file size of ascii format data is considerably larger (2–3 GB each), we recommend using one of the other file types if possible.

Channel coordinates

Approximate channel coordinates according to the extended 10-20 system are stored in the folder `channel_locations` in the text file `chanlocs_ced.txt` using polar, cartesian, and spherical coordinates and in the text file `chanlocs_besa.txt` using the notation expected by BESA (and other EEG software). Furthermore, channel coordinates are already included in the EEGLAB-formatted `.set` data files. Note that the bipolar channels 71 (vEOG) and 72 (hEOG) have no associated channel coordinates.

Trigger codes and event information

The data files include information about the timing and type of relevant experimental events in the form of triggers or markers, which mark the time of image onsets and their corresponding experimental conditions. Each trigger is a four-digit number, where each digit codes information about the stimuli and condition. The triggers are coded according to the following scheme (from left-most digit to right):

scene category Did the image on this trial show a scene that can be described as man-made (1; i.e. building or highway) or as natural (2; i.e. beach or forest)?

old Was the image on this trial shown for the first time (0; new) or had it been shown before in this session (1; old)?

behavior If the image on this trial was old, was it correctly recognized as old (1; hit) or was it incorrectly judged as new (2; miss/forgotten)? If the image was new, was it incorrectly judged as old (3; false alarm) or was is correctly judged as new (4; correct rejection)?

subsequent memory If the image on this trial was shown again on a subsequent trial, did the subject correctly recognize it as old (1; subsequently remembered) or did they judge it as new (2; subsequently forgotten) then? The comparison between trials that are associate with future remembering or forgetting are referred to as "subsequent memory" (Paller and Wagner, 2002). Note that subsequent memory is not defined for trials with images that were not shown another time, i.e. for images that had already been repeated or for trials at the very end of the experiment.

Table 1: Trigger codes used to index experimental conditions.

	scene category	old	behavior	subsequent memory
0	—	new	—	subseqnt. forgotten
1	man-made	old	hit	subseqnt. remembered
2	natural	—	miss/forgotten	
3	—	—	false alarm	—
4	—	—	correct rejection	—
5	—	—	—	—
6	—	—	—	—
7	—	—	—	—
8	—	—	—	—
9	—	—	missing/NA	missing/NA

Examples:

2040 A trial with a forest scene, which is a natural category (man-made=2), that was shown for the first time (old=0, meaning it was new) and was correctly rejected, i.e. not incorrectly judged as old (behavior=4). The next time this image was shown, it was not recognized as old, i.e. it was subsequently forgotten (subsequent memory=0).

1129 A trial with a building scene, which is a man-made category (man-made=1), that had been shown before (old=1) and was not recognized as old, i.e. it was forgotten (behavior=2). The image was not shown another time, so subsequent memory is not applicable (subsequent memory=9).

A complete list of all possible trigger values and their associated experimental conditions is provided in the file `TriggerTable.csv` in the `Documentation` folder.

Furthermore, information about each trial is provided for each subject as a text file in the `events` folder with comma-separated values (`.csv` file), where each line represents a single trial. The columns represent the index of the trial, the trigger code, the latency of stimulus presentation in seconds and sampling points relative to the start of the file, and the trial's experiemntal conditions: scene category, old, behavior, subsequent memory. Note that the EEGLAB-formatted and BIDS-formatted data files already include all relevant event information in their respective event structures. BDF formatted data files only include the numeric trigger codes.

Acknowledgments

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10:433–436.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134:9–21.

Friedman, D. (1990). Cognitive event-related potential components during continuous recognition memory for pictures. *Psychophysiology*, 27:136–148.

Isola, P., Xiao, J., Torralba, A., and Oliva, A. (2011). What makes an image memorable? In *CVPR 2011*, pages 145–152.

Paller, K. A. and Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6:93–102.

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., and Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, 6:103.

Shepard, R. N. and Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 62(3):302–309.