Data in Brief

**BIDS-structured resting-state electroencephalography (EEG) data extracted from an experimental paradigm**

Christoffer Hatlestad-Hall\textsuperscript{a,1,∗}, Trine Waage Rygvold\textsuperscript{b,1}, Stein Andersson\textsuperscript{b}

\textsuperscript{a} Department of Neurology, Oslo University Hospital, Oslo, Norway

\textsuperscript{b} Department of Psychology, University of Oslo, Oslo, Norway

**A R T I C L E   I N F O**

Article history:
Received 15 February 2022
Revised 12 September 2022
Accepted 26 September 2022
Available online 1 October 2022

Dataset link: SRM Resting-state EEG (Original data)

Keywords:
Electroencephalography
EEG
Resting-state
BIDS
Preprocessing
Neuropsychological assessment
Neurophysiology

**A B S T R A C T**

Electroencephalography (EEG) offers a unique window into the dynamics of the neuronal symphony that powers our brains. Here, we describe a publicly available dataset of EEG recorded from 111 healthy subjects. The data were recorded with 64 electrodes in a resting-state condition, an approach that offers broad-spectred analysis options, including functional connectivity and graph theory. In a subset of the subjects (n = 42), a second EEG recording was performed, 2-3 months after the initial recording, allowing measurement stability to be assessed. Furthermore, in connection with the EEG acquisition, a range of neuropsychological test scores were obtained for each subject. The dataset is comprehensive and organised according to the Brain Imaging Data Structure (BIDS) specification, providing a valuable starting point for both aspiring and experienced researchers in a range of fields, including cognitive neuroscience, data science, machine learning, and clinical neurophysiology.

© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

\textsuperscript{∗} Corresponding author.

E-mail address: chr.hh@pm.me (C. Hatlestad-Hall).

Social media: @CHatlestadHall (C. Hatlestad-Hall)

\textsuperscript{1} Hatlestad-Hall and Rygvold should be considered joint first authors.

https://doi.org/10.1016/j.dib.2022.108647

2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Specifications Table

| Subject | Neuroscience: Neurophysiology |
|---------|--------------------------------|
| Specific subject area | Non-invasive cognitive neuroscience |
| Type of data | Electroencephalography (EEG): Resting-state |
| How the data were acquired | EEG system: BioSemi ActiveTwo 64-channel system (10-20 electrode positions). Acquisition software: BioSemi ActiView ver. 7.07 Sampling frequency: 1024 Hz Data preprocessing (including BIDS organising): MATLAB ver. 2021a, FieldTrip (repository cloned 30 July 2021). A subset of the subjects (n = 42) underwent a second EEG recording, 2-3 months after the initial. |
| Data format | Raw Preprocessed/cleaned |
| Description of data collection | The EEG data were acquired using a BioSemi ActiveTwo 64-channel system. 111 participants underwent an experimental, non-invasive long-term potentiation-like stimulation paradigm including a segment of four minutes of resting-state (with eyes closed) recording, from which the current data were extracted. During acquisition, the participants were seated in a dimly lit room isolated from environmental noise. |
| Data source location | Institution: Dept. of Psychology, Faculty of Social Sciences, University of Oslo City: Oslo Country: Norway |
| Data accessibility | Repository name: SRM Resting-state EEG Host: OpenNeuro.org OpenNeuro Accession Number: ds003775 Direct URL to the data: [https://openneuro.org/datasets/ds003775](https://openneuro.org/datasets/ds003775) |

Value of the Data

- Well-structured and comprehensive EEG datasets are important for supporting the development of applications that require large amounts of data, such as machine learning approaches. Resting-state data offer a broad-spectred starting point for model development based on features including, but not limited to, functional and effective connectivity, graph theory, and microstate analyses.
- The current data may benefit a range of fields, including cognitive neuroscience, data science, biomedical engineering, and clinical neurophysiology and neurology. In particular, students or other research staff who require data for developing analysis pipelines, prediction models, or other applications, are dependent on publicly available EEG datasets.
- The current dataset can be used in several contexts, including signal processing methods development (source reconstruction, complex feature extraction), brain-behaviour analysis (correlations between electrophysiological features and cognitive test performance), and machine learning (classification of age, sex, or cognitive status).
- Beside the EEG data, the dataset includes scores on a wide range of neuropsychological tests. The tests were carried out by a trained and experienced clinical psychologist.
- The dataset contains high-quality EEG data, acquired in one site with one system.
- In combination with the BIDS-organised data, the preprocessing and cleaning code distributed with the dataset provides a starting point for students and others who are new to and are learning about EEG and signal processing. The code can be reused and modified for other applications.

1. Data Description

The dataset is structured according to the electroencephalography-extended Brain Imaging Data Structure (BIDS) specification [1]. The reader is referred to [2] for a detailed and updated description of the specification. The general file structure of the dataset is outlined in Fig. 1.
Fig. 1. Directory and file structure. Only two subjects (001 and 002) are shown.
Please note that in the following, ‘###’ (three digits) refers to the sequential subject number (001-111), and ‘#’ (one digit) refers to the session number (1 or 2). Specified paths beginning with ‘/’ originate from the dataset’s root directory.

1.1. Root-level files and directories

The following files and directories are found at the dataset’s root-level:

- **CHANGES**: Text file describing changes made to the dataset from version to version.
- **dataset_description.json**: Metadata file describing the dataset parameters.
- **participants.json**: Metadata file describing the variables contained in participants.tsv.
- **participants.tsv**: Tab-separated text file containing subjects’ values on a range of variables, including sex, age, and neuropsychological test scores.
- **README**: Text file describing the dataset.
- **code**: Directory containing MATLAB code used to process the dataset.
- **derivatives**: Directory containing cleaned data stored as MATLAB/EEGLAB .set files.
- **sub-###**: Subject directories containing raw data. ### refers to the subject number.

1.2. Subject directories and raw data

Raw EEG files (sub-###_ses-t#_task-resteyesc_eeg.edf) are stored in /sub-###/ses-t#/eeg/ in the European Data Format (.edf) [3]. Each raw EEG data file has a sidecar metadata file (sub-###_ses-t#_task-resteyesc_eeg.json) and a tab-separated text file listing all the data channels present in the raw EEG file (sub-###_ses-t#_task-resteyesc_channels.tsv). In this dataset, the two latter files are identical across all subjects, as all raw EEG files are of equal length and sample frequency, contain the same data channels, and are recorded with a single system. In addition, the time of data acquisition (date and time format: yyyy-mm-ddThh:mm:ss) is available from a tab-separated text file (/sub-###/ses-t#/sub-###_ses-t#_scans.tsv).

1.3. Derived data and code

For convenience, a preprocessed and cleaned version of the complete data are found under /derivatives/cleaned_data/. The directory structure is identical to the raw data directory structure. The cleaned EEG data files (sub-###_ses-t#_task-resteyesc_desc-cleaned_eeg.set) are stored in the format of EEGLAB (.set) [4], a MATLAB toolbox. The .set files of EEGLAB are out-of-the-box ready for use with other M/EEG analysis packages in MATLAB such as FieldTrip [5], but also with non-MATLAB alternatives, including MNE-Python [6]. As with the raw data, the cleaned data are accompanied by a metadata sidecar file (.json) and a list of channels (.tsv). Please note that the latter contains a quality evaluation of each channel (‘good’ or ‘bad’).

The code distributed with the dataset (/code/bidsify-srm-restingstate/) includes MATLAB scripts for organising data according to the BIDS specification (s1_extract_rs.m), and for automatic, unsupervised preprocessing and cleaning of the raw data (s2_preprocess.m). In addition, a script for fixing an issue related to the formatting of the /sub-###/ses-t#/sub-###_ses-t#_scans.tsv files, is included (s1b_fix_scans_tsv.m). In the subdirectories /code/bidsify-srm-restingstate/functions and /code/bidsify-srm-restingstate/chanlocs are MATLAB functions required for executing the scripts, and an electrode/channel positions template (X, Y, Z coordinates), respectively.

The preprocessing pipeline includes (a) re-referencing of the data with an iterative re-referencing procedure (ch_iterative_reref.m), (b) detection and removal of bad segments and channels (ch_ampstat_badsegments.m), (c) 1 Hz high-pass filtering (default EEGLAB filter design), (d) removal of power line noise with ZapLine [7], (e) independent component estimation
with the SOBI algorithm [8], (f) subtraction components related to ocular or muscular activity with ILabel [9], (g) interpolation of removed channels (marked as 'bad' in the .tsv sidecar file), (h) 45 Hz low-pass filtering (default EEGLAB filter design), and (i) detection of small, yet repetitive signal glitches (ch_find_glitches.m), where the worst channels are interpolated. After the preprocessing and cleaning, the continuous data were segmented into non-overlapping 4-seconds epochs. Please note that most pipeline parameters are adjustable, should the user want to experiment with various configurations. The interested reader is encouraged to inspect the pipeline code and documentation.

2. Experimental Design, Materials and Methods

2.1. Subjects

The dataset comprises resting-state EEG, age, sex, and raw (not standardised) neuropsychological assessment scores from 111 healthy control subjects (mean age = 37.6 years; range = 17-71; SD = 14.0). All subjects provided written informed consent prior to participation. Normal or corrected-to-normal vision and normal hearing were required, and none of the subjects reported severe psychiatric or neurological symptoms. All subjects were screened using a basic audiometry protocol. Subjects were recruited through social media platforms (Facebook and Instagram), and local advertisement.

2.2. Experimental protocol outline

The current data were recorded in the context of an event-related potentials (ERP) experiment investigating the long-term potentiation-like (LTP-like) modulation of sensory evoked potentials following high-frequency repetitive stimulus exposure, termed the sensory response modulation (SRM) phenomenon [10]. The complete EEG data acquisition included three experimental paradigms: a visual evoked potentials (VEP) paradigm, an auditory evoked potentials (AEP) paradigm, and a loudness-dependent AEP paradigm. For a detailed description of the two former paradigms, the reader is referred to [10]. For a review of the latter paradigm, see e.g., [11]. Please note that the data from these paradigms are not included in the current dataset.

The resting-state (eyes closed) EEG was acquired towards the end of the above-described data acquisition, approximately 45 minutes into the session. The segment was introduced with standardised written introductions presented on a 24” LCD screen (BenQ, model ID: XL2420-B). The instructions read (translated from Norwegian): “Please close your eyes, and remain seated with your eyes closed for approximately four minutes. You will not see any patterns, hear any sounds, nor do you have to press any buttons during this segment”.

2.3. EEG acquisition

The EEG data were recorded using a 64-channel (Ag-AgCl electrodes) BioSemi ActiveTwo system (BioSemi B.V., Amsterdam). The electrodes were positioned according to the extended 10–20 system (10-10) [12]. Data acquisition was done at a sampling rate of 1024 Hz without online filters, except for the default hardware anti-aliasing filter. Event markers were sent from the MATLAB platform to the EEG data through a 25-pin serial port.

The EEG recordings were conducted by experienced operators and provide data of high quality. In two previous studies, our group analysed event-related data from the same recordings and reported the results [10,13]. In these investigations, in addition to visual expert assessment, a preprocessing and cleaning pipeline similar to the one included with the dataset was...
utilised. The current dataset’s resting-state data were collected without interruption from the same recording sessions (no removal or modification of the cap or individual electrodes). Quantitatively, the included fully automated cleaning procedure, which should be considered strict in terms of the data rejection, produced a derived dataset in which 64.1% of the epoched data files kept more than 90% of their channels and 23.5% retained between 75% and 90% of their channels.

2.4. Neuropsychological tests

All subjects underwent neuropsychological testing at the first timepoint. The test battery included Rey Auditory Verbal Learning Test [14] (verbal learning and memory), Wechsler Adult Intelligence Scale-IV [15] Digit Span (attention span and working memory), and the Delis-Kaplan Executive Function System (D-KEFS) [16,17] tests Trail Making Test (psychomotor speed and executive functioning), Colour-Word Interference Test (reading speed and executive functioning), and Verbal Fluency (phonemic and semantic processing). Table 1 contains descriptive statistics and variable descriptions for each test.

| Test | Variable | Description | Mean ± SD | Range |
|------|----------|-------------|-----------|-------|
| RAVLT | ravlt_1 | Number of items correctly recalled after first learning trial | 8.4 ± 2.2 | 4 - 14 |
|      | ravlt_5 | Number of items correctly recalled after fifth learning trial | 13.9 ± 1.3 | 10 - 15 |
|      | ravlt_tot | Total number of items correctly recalled across all five learning trials | 60.5 ± 7.6 | 43 - 74 |
|      | ravlt_imm | Number of items correctly recalled immediately after the final learning trial | 12.9 ± 2.3 | 5 - 15 |
|      | ravlt_del | Number of items correctly recalled 30 minutes after the final learning trial | 13.0 ± 2.3 | 5 - 17 |
|      | ravlt_rec | Number of items correctly recognised from a list | 14.4 ± 1.6 | 12 - 15 |
|      | ravlt_fp | Number of false positive responses during the recognition task | 0.3 ± 0.8 | 0 - 7 |
| WAIS-IV DS | ds_forw | Achieved score in the test’s forward condition | 9.6 ± 2.2 | 5 - 15 |
|      | ds_back | Achieved score in the test’s backward condition | 9.1 ± 2.0 | 4 - 16 |
|      | ds_seq | Achieved score in the test’s sequencing condition | 9.5 ± 2.2 | 5 - 15 |
|      | ds_tot | Total score across all conditions | 28.2 ± 5.0 | 17 - 44 |
| TMT | tmt_2 | Number of seconds elapsed before completion of the test’s number condition | 28.3 ± 11.4 | 13 - 75 |
|      | tmt_3 | Number of seconds elapsed before completion of the test’s letter condition | 27.4 ± 12.9 | 11 - 77 |
|      | tmt_4 | Number of seconds elapsed before completion of the test’s switching condition | 70.3 ± 30.4 | 29 - 202 |
| CWIT | cw_1 | Number of seconds elapsed before completion of the test’s colour only condition | 29.1 ± 5.7 | 20 - 53 |
|      | cw_2 | Number of seconds elapsed before completion of the test’s reading only condition | 21.0 ± 3.4 | 14 - 30 |
|      | cw_3 | Number of seconds elapsed before completion of the test’s interference condition | 49.4 ± 12.5 | 31 - 100 |
|      | cw_4 | Number of seconds elapsed before completion of the test’s interference and switching condition | 56.7 ± 14.9 | 37 - 130 |

(continued on next page)
Table 1 (continued)

| Test | Variable | Description | Mean ± SD | Range |
|------|----------|-------------|-----------|-------|
| VF   | vf_1     | Number of words correctly listed in the test’s phonemic condition | 47.8 ± 12.5 | 19 - 84 |
|      | vf_2     | Number of words correctly listed in the test’s semantic condition | 51.2 ± 11.6 | 14 - 88 |
|      | vf_3     | Number of words correctly listed in the test’s switching condition | 16.1 ± 3.3  | 7 - 24  |

Note. RAVLT = Rey Verbal Learning Test; WAIS-IV DS = Wechsler’s Adult Intelligence Scale-IV Digit Span; TMT = Trail Making Test; CWIT = Colour-Word Interference Test; VF = Verbal Fluency.

Ethics Statements

The data collection was conducted in accordance with the Declaration of Helsinki, and informed consent was obtained from all participants. The procedures were approved by the Regional Ethics Committee of South-Eastern Norway (reference number: 2016/2003).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

SRM Resting-state EEG (Original data) (OpenNeuro.org).

CRediT Author Statement

Christoffer Hatlestad-Hall: Data curation, Methodology, Software, Writing – original draft; Trine Waage Rygvold: Investigation, Project administration, Writing – review & editing; Stein Andersson: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] C.R. Pernet, S. Appelhoff, K.J. Gorgolewski, G. Flandin, C. Phillips, A. Delorme, R. Oostenveld, EEG-BIDS, an extension to the brain imaging data structure for electroencephalography, Sci Data 6 (2019) 103, doi:10.1038/s41597-019-0104-8.
[2] Brain Imaging Data Structure, Brain Imaging Data Struct. (n.d.). https://bids.neuroimaging.io. Accessed February 9, 2022.
[3] B. Kemp, A. Värrri, A.C. Rosa, K.D. Nielsen, J. Gade, A simple format for exchange of digitized polygraphic recordings, Electroencephalogr. Clin. Neurophysiol. 82 (1992) 391–393, doi:10.1016/013–4694(92)90009-7.
[4] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods. 134 (2004) 9–21, doi:10.1016/j.jneumeth.2003.10.009.
[5] R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data, Comput. Intell. Neurosci. 2011 (2011) 156869, doi:10.1155/2011/156869.
[6] A. Gramfort, M. Luessi, E. Larson, D.A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M. Hamalainen, MEG and EEG data analysis with MNE-Python, Front. Neurosci. 7 (2013) 267, doi:10.3389/fnins.2013.00267.

[7] A. de Cheveigné, ZapLine: A simple and effective method to remove power line artifacts, Neuroimage 207 (2020) 116356, doi:10.1016/j.neuroimage.2019.116356.

[8] A. Belouchrani, K. Abed-Merai, J.F. Cardoso, E. Moulines, Second-order blind separation of temporally correlated sources, in: Proc. Int. Conf. Digital Signal Processing, Citeeseer, 1993, pp. 346–351. https://www.researchgate.net/profile/Adel_Belouchrani/publication/2699542_Second_Order_Blind_Separation_of_Temporally_Correlated_Sources/links/00463517ab3e0aed06/Second-Order-Blind-Separation-of-Temporally-Correlated-Sources.pdf.

[9] L. Pion-Tonachini, K. Kreutz-Delgado, S. Makeig, iLabel: An automated electroencephalographic independent component classifier, dataset, and website, Neuroimage 198 (2019) 181–197, doi:10.1016/j.neuroimage.2019.05.026.

[10] T.W. Rygvol, C. Hatlestad-Hall, T. Elvsåshagen, T. Moberget, S. Andersson, Do visual and auditory stimulus-specific response modulation reflect different mechanisms of neocortical plasticity? Eur. J. Neurosci. 53 (2021) 1072–1085, doi:10.1111/ejn.14964.

[11] B.V. O'Neill, R.J. Croft, P.J. Nathan, The loudness dependence of the auditory evoked potential (LDAEP) as an in vivo biomarker of central serotonergic function in humans: rationale, evaluation and review of findings, Hum. Psychopharmacol. 23 (2008) 355–370, doi:10.1002/hup.940.

[12] R. Oostenveld, P. Praamstra, The five percent electrode system for high-resolution EEG and ERP measurements, Clin. Neurophysiol. 112 (2001) 713–719, doi:10.1016/s1388-2457(00)00527-7.

[13] T.W. Rygvol, C. Hatlestad-Hall, T. Elvsåshagen, T. Moberget, S. Andersson, Long-Term Potentiation-Like Visual Synaptic Plasticity Is Negatively Associated With Self-Reported Symptoms of Depression and Stress in Healthy Adults, Front. Hum. Neurosci. 16 (2022) 867675, doi:10.3389/fnhum.2022.867675.

[14] J. Espenes, I.V. Eliassen, F. Ohman, E. Hessen, K. Waterlo, M. Eckerstrøm, I.M. Lorentzen, C. Bergland, M. Halvari Niska, S. Timón-Reina, A. Wallin, T. Fladby, B.-E. Kirsebom, Regression-based normative data for the Rey Auditory Verbal Learning Test in Norwegian and Swedish adults aged 49-79 and comparison with published norms, Clin. Neuropsychol. (2022) 1–25, doi:10.1080/13854046.2022.2106890.

[15] D. Wechsler, D.L. Coalson, S.E. Raiford, WAIS-IV technical and interpretive manual, Pearson, San Antonio, Texas, USA, 2008.

[16] D.C. Delis, E. Kaplan, J.H. Kramer, Delis-Kaplan Executive Function System, The Psychological Corporation, San Antonio, Texas, USA, 2001, doi:10.1037/15082-000.

[17] S. Homack, D. Lee, C.A. Riccio, Test review: Delis-Kaplan executive function system, J. Clin. Exp. Neuropsychol. 27 (2005) 599–609, doi:10.1080/13803390490918444.