A high-density scalp EEG dataset acquired during brief naps after a visual working memory task

Ning Mei\textsuperscript{a}, Michael D. Grossberg\textsuperscript{b}, Kenneth Ng\textsuperscript{a}, Karen T. Navarro\textsuperscript{a}, Timothy M. Ellmore\textsuperscript{a,∗}

\textsuperscript{a} Department of Psychology, The City College of the City University of New York, New York, NY 10031, United States
\textsuperscript{b} Department of Computer Science, The City College of the City University of New York, New York, NY 10031, United States

Article history:
Received 3 January 2018
Received in revised form 9 March 2018
Accepted 18 April 2018
Available online 25 April 2018

Abstract

There is growing interest in understanding how specific neural events that occur during sleep, including characteristic spindle oscillations between 10 and 16 Hz (Hz), are related to learning and memory. Neural events can be recorded during sleep using the well-known method of scalp electroencephalography (EEG). While publicly available sleep EEG datasets exist, most consist of only a few channels collected in specific patient groups being evaluated overnight for sleep disorders in clinical settings. The dataset described in this Data in Brief includes 22 participants who each participated in EEG recordings on two separate days. The dataset includes manual annotation of sleep stages and 2528 manually annotated spindles. Signals from 64-channels were continuously recorded at 1 kHz with a high-density active electrode system while participants napped for 30 or 60 min inside a sound-attenuated testing booth after performing a high- or low-load visual working memory task where load was randomized across recording days. The high-density EEG datasets present several advantages over single- or few-channel datasets including most notably the opportunity to explore spatial differences in the distribution of neural events, including whether spindles occur locally on only a few channels or co-occur globally across many channels.

* Corresponding author.
E-mail address: tellmore@ccny.cuny.edu (T.M. Ellmore).
whether spindle frequency, duration, and amplitude vary as a function of brain hemisphere and anterior-posterior axis, and whether the probability of spindle occurrence varies as a function of the phase of ongoing slow oscillations. The dataset, along with python source code for file input and signal processing, is made freely available at the Open Science Framework through the link https://osf.io/chav7/.

© 2018 The Authors. 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 area                  | Neuroscience                                   |
|------------------------------|-----------------------------------------------|
| More specific subject area   | Cognitive neuroscience of learning and memory |
| Type of data                 | High density scalp electroencephalography     |
| How data was acquired        | actiCHamp active electrode EEG system (Brain Products, GmbH) |
| Data format                  | EEG data are available in the Brainvision raw.eeg format. Python scripts are provided for reading the raw files and optionally performing basic signal processing including filtering and ICA-based artifact correction. |
| Experimental factors         | Twenty-two participants underwent EEG recording on two separate days. On each of the two days, subjects participated between the hours of 10 a.m. and 5 p.m. The night before each recording session, participants were instructed to go to sleep an hour and a half later than usual and wake up at their normal time so that they would be tired during the nap. Each completed a high- or low-load visual working memory task where load was randomized across days. Following the working memory task, participants napped for 30 or 60 min inside a sound-attenuated testing booth. During each nap, signals from 64-channels were continuously recorded at 1 kHz. |
| Experimental features        | The 64 recorded channels included 2 electrodes for electrooculography (EOG). The left EOG electrode was placed under the left eye on the maxilla, while the right EOG electrode was placed above the right eyebrow on the frontal bone. Both EOG electrodes on either side were in line with the middle of the eyes. Before EEG recordings began, the impedance of all electrodes was optimized to be less than 25 kOhms by application of electrolytic gel between the scalp and the electrodes tips. The reference electrode during recording was TP9 (left mastoid). Recorded data was re-referenced to the average of the data during preprocessing. |
| Data source location         | New York, New York, USA 10031                  |
| Data accessibility           | Open Science Framework public repository.      |
| “Nap EEG”                   | https://osf.io/chav7/                           |
**Value of the data**

- High-density EEG acquired during naps is valuable for exploring with high spatial resolution the time-frequency, amplitude, and distribution of spindles as well as their occurrence relative to non-REM sleep slow waves.
- Because we provide manually annotated sleep stage information and 2528 manually annotated spindles, these EEG nap datasets can be used for the development and benchmarking of automated sleep stage classification and automated spindle detection models.
- EEG datasets acquired at multiple times in the same participants can be used for within-subject test-retest reliability as well as exploration of inter-hemispheric and anterior-posterior differences in spindle occurrence.
- Python source code is provided for file input, preprocessing of the EEG data as well as a framework to optimize and validate machine learning models for automatic spindle classification.

1. Data

The dataset consists of continuous EEG data obtained during naps taken by healthy adult participants after performance of a visual working memory task. Given the restricted age range of the sample, this dataset is not suited to address questions related specifically to childhood brain development or normal aging or specific life styles that may influence napping and memory. Each participant took part in two recording sessions during which each completed a high- or low-load scene working memory task followed by a 30 or 60-min nap on a bed inside a sound-attenuated recording chamber. During each recording session participants underwent continuous recording from 64-channels, including 2 channels for electrooculography (EOG). The dataset includes raw Brainvision format .eeg, .vdhr, and .vmrk files, manually annotated sleep stage information, manually annotated spindles occurrences, and python source code for file input/output, minimal signal pre-processing including filtering and independent component analysis-based artifact correction, automatic identification of spindles using a filter based thresholding approach, and a framework for model validation and optimization using a supervised machine learning approach [1].

2. Experimental design, materials and methods

2.1. Subjects

The dataset consists of 41 recordings from 22 subjects (mean age 25.5 ± 7.03 standard deviation SD, range 18–42 years old) who each provided written informed consent and completed study procedures according to a protocol approved by the Institutional Review Board of the City College of New York. Of these 22 subjects, 16 were male (mean age 26.5 ± 7.80 SD) and 6 were female. Subjects were compensated $15 per hour for participation. Experiments were carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Nine EEG recordings (including both days of subject 7, 26, 27, day 1 of subject 5 and 12, day 2 of subject 20) were acquired but excluded because the subjects failed to reach stage 2 sleep or because their data were too noisy for processing.

2.2. EEG acquisition

A total of 64-channels of data, including 2 EOG electrodes, were continuously recorded at 1 kHz using an antiCHamp active electrode system (Brain Products, GmbH). Before EEG recordings were initiated, the impedance of all electrodes was optimized to be < 25 kOhms by application of electrolytic gel between the scalp and the electrodes tips. The reference electrode during recording was TP9 (left mastoid) and two additional flat electrodes were placed around the eyes for EOG recording.
2.3. Experimental design

Subjects completed memory tasks before and after the nap EEG recordings. Over the course of two days, subjects participated between the hours of 10 a.m. and 5 p.m. Each session lasted two hours. Before each session, subjects were instructed to go to sleep an hour and a half later than their usual sleep time so that they would be tired at the start of the experiment and therefore more likely to fall asleep. During each session, subjects performed a low or high load scene working memory task. The scene working memory task was a variant of the well-known Sternberg paradigm [2], which consists of i) an encoding period during which new stimuli are sequentially presented, ii) a delay period during which subjects must maintain the stimuli presented during the trial’s encoding period in memory for a brief time, and ii) a probe period during which a positive or negative probe stimulus is presented. The subject pressed a button indicating whether the probe was or was not presented during the trial’s encoding period. A positive probe was a stimulus shown during the trial’s encoding period. A negative probe was a new stimulus not shown during the trial’s encoding period. The ratio of positive to negative probes was 50:50 in both the low and high load tasks. The low load task consisted of 100 trials during which subjects encoded two pre-experimentally novel outdoor scenes presented each for 2 s, maintained them in memory for 6 s, and then responded by pressing one of two buttons indicating whether a subsequently presented probe scene displayed for 2 s was previously presented in the encoding set. After the probe, a phase-scrambled scene was presented for 5 s after which the next trial began. The high load task was similar, but the encoding set on each trial included five scenes and the total number of trials presented was 40 so that the number of pre-experimentally novel stimuli participants saw on each day was equal. Immediately before and after napping, subjects completed a yes/no recognition task to assess memory for the scenes presented during the working memory task.

2.4. Manually annotated sleep EEG data

A single rater (KTN) manually annotated sleep stages and spindles in each EEG dataset. The EDF browser (http://www.teuniz.net/edfbrowser/) was used for sleep staging and annotating spindles.

![Fig. 1. Example epoch of scalp EEG traces with a highlighted spindle. The 61 scalp channels for one subject are shown for a 30-s epoch. Each channel is shown after re-referencing to the average and filtering between 11 and 16 Hz and a 60 Hz notch filter. Yellow indicates approximate time of a spindle, which is most noticeable in posterior right hemisphere (red traces). Blue and black traces represent left and midline (e.g., Cz) traces respectively. Amplitude scale is set to 60 uV and the time interval separating vertical dotted lines is 1000 ms.](image-url)
Spindles that had dominant power between 11 and 16 Hz with durations between 0.5 and 2 s were annotated. Only those spindles that occurred across several channels (n > 3) with minimal temporal variation (< 1 s) were marked. Spindles were manually annotated by visual inspection of six channels F3, F4, C3, C4, O1, and O2, which were named and placed according to the internationally recognized 10–20 EEG system. The six channels were chosen because they span anterior (i.e., frontal) to posterior

Fig. 2. Example meridian top-down plot illustrating scalp EEG voltage changes around a spindle. The 0 ms contour occurs at the center of the yellow highlighted window in Fig. 1. Time steps of 50 ms before and after 0 ms illustrate voltage changes during the spindle. A single contour step is 1.5 μV with red representing positive change and black negative change. Circles represent individual electrode locations.
EEG signals were filtered using a second order bandpass Butterworth filter. Sleep stages and spindles were marked in each dataset at 30 s intervals. Decisions about spindle occurrences were made based on visible spindle features detected in more than three channels. No information about individual spindle duration was manually annotated. Instead, all spindles were set to 2 s by default, with spindle onset occurring 0.5 s prior to where the spindle was marked. The dataset includes a total of 2528 manually annotated spindles.

**Fig. 3.** Block diagram showing the algorithm used for automated spindle detection. Python source code distributed with the dataset allows for file input/output of the native BrainVision *.eeg files, filtering and artifact rejection, and optionally automated spindle detection, comparison with manually labelled spindles occurring in different non-REM sleep stages, and model validation. Root-mean-square (RMS) is computed and used to represent the feature maps for the automated spindle detection.
2.5. Python code for EEG processing

EEG datasets are provided in their native raw format, which have the Brainvision.eeg extension. The accompanying .vhdr header files contain information about data acquisition rate, etc. The python source code that implements the preprocessing methods is available along with the data at https://osf.io/zxr8m/. The code implements minimal signal preprocessing including down-sampling, low-pass filtering, and artifact correction and depends on the freely available Minimum Norm Estimation (MNE)-python library [4]. During the preprocessing, signals were re-referenced to a common reference, which is the average of activity at all electrodes. To prepare for artifact correction using ICA, the signals were filtered using a bandpass filter between 0.1 and 200 Hz with a notch filter to remove line noise and harmonics at 60, 120, and 180 Hz. The filter is a zero-phase finite impulse response (FIR) filter with transition bandwidth of approximately 5 Hz. We used the MNE-python implemented Independent Component Analysis (ICA) procedures for artifact correction with all default hyper-parameters for the processing except the peak-to-peak amplitude change, which was used for excluding bad segments of data from the ICA computations. Eye blinks and muscle artifacts were identified automatically by correlating each signal to the EOG channels and variations of the signals. These components were removed in the ICA space and projected back to the signal space after removal. A bandpass filter between 0.1 and 50 Hz was applied to the corrected signal. The bandpass filters were zero-phase and the method was overlap-add FIR filtering. The filter length was chosen automatically based on the size of the transition region, which was 6.6 times the reciprocal of the shortest transition band for the Hamming window. In the python source code, the processed data can optionally be saved for later use with no further human input involved in the preprocessing. Additional distributed source code implements automated spindle detection using a filter based and thresholding pipeline with machine learning optimization [1]. An example 30 s epoch highlighting a posterior spindle in one subject is shown in Fig. 1. An example top down meridian view in Fig. 2 shows positive and negative voltage changes at different time steps relative to the center of the time window. A diagram of the processing performed by the different code components is shown in Fig. 3.

Acknowledgements

This work was supported by PSC-CUNY award TRADB #67546-00 45. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number SC2GM109346. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank Forrest Armstrong, Miriam San Lucas, Ben Fernandez, Chelsea Reichert, and Erin Hunt for help with data collection.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.04.073.

References

[1] N. Mei, M.D. Grossberg, K. Ng, K.T. Navarro, T.M. Ellmore, Identifying sleep spindles with multichannel EEG and classification optimization, Comput. Biol. Med. 89 (2017) 441–453.
[2] S. Sternberg, High-speed scanning in human memory, Science 153 (3736) (1966) 652–654.
[3] M. Schabus, G. Gruber, S. Parapatics, C. Sauter, G. Klosch, P. Anderer, et al., Sleep spindles and their significance for declarative memory consolidation, Sleep 27 (8) (2004) 1479–1485.
[4] A. Gramfort, M. Luessi, E. Larson, D.A. Engemann, D. Strohmeier, C. Brodiebeck, et al., MNE software for processing MEG and EEG data, Neuroimage 86 (2014) 446–460.