CRCNS.org fcx-1 (frontal cortex number 1) data description
Version 0.7 (July 26, 2016)

Summary

Data was recorded using silicon probe electrodes in the frontal cortices of male Long Evans rats between 4-7 months of age. The design was to have no specific behavior, task or stimulus, rather the animal was left alone in it's home cage (which it lives in at all times). Data includes both local field potentials (LFP) and spikes. 11 total animals, 27 recording sessions, 1360 total units recorded, 1121 units considered stable, 995 putative excitatory units and 126 putative inhibitory units. Only recordings including a "WAKE-SLEEP" episode wherein at least 7 minutes of wake are followed by 20 minutes of sleep. On average 2 such WAKE-SLEEP episodes per recording session.

Dataset is that used in the following publication:

Network Homeostasis and State Dynamics of Neocortical Sleep
Watson BO, Levenstein D, Greene JP, Gelinas JN, Buzsáki G.
Neuron. 2016 Apr 27. pii: S0896-6273(16)30056-3.
doi: 10.1016/j.neuron.2016.03.036

Of note, Table 1 and Supplementary Figure 1 contain histological and overview data as well as definitions of brain states used.

Conditions for using the data

If you publish any work using the data, please cite the publication above (Watson et al, 2016) and also cite the data set in the following recommended format:

Watson BO, Levenstein D, Greene JP, Gelinas JN, Buzsáki G. (2016); Multi-unit spiking activity recorded from rat frontal cortex (brain regions mPFC, OFC, ACC, and M2) during wake-sleep episode wherein at least 7 minutes of wake are followed by 20 minutes of sleep. CRCNS.org
http://dx.doi.org/10.6080/K02N506Q

Methods

Methods for data acquisition and processing are fully described in Watson et al, 2016 (above).

Data files organization

Data are organized into a series of folders. Each folder contains the data for one recording session (lasting between about 1 and 8 hours duration). Each folder contains a series of files with suffixes including name-specified aspects of the data. See below for details.

While the number of file subtypes is large, that is due to the fact that we provide a range of data subtypes from most processed to least and also multiple aspects of our data. One could ignore the least processed versions and jump to things like fully identified spike trains. See "Getting Started" below.

Included also are
WatsonSleepHomestasis2016Table.xlsx – a summary of all recordings
LoadBinary_bw.m – matlab file for reading .eeg binary files
TStoolbox.zip – zip of matlab object-oriented programming package that may be used for processing tsdArray, intervalSet and tsd format versions of data.

Data format

Data are mostly in .mat matlab format, there are also .csv files and lfp data in binary files called either .eeg or .lfp. All folders have a name signifying the name of the recording session, ie BWRat17_032413, which serves as a basename for all files. Data files inside each folder have names in the format basename_XXXXX.yyy where XXXXX will signify various data types and .yyy is a suffix. Details below.

Also of note: we actually used tstoolbox format for data processing, but I have included both tsobjects format and more typical matlab formats of the same data for those who do not use this data toolbox. TStoolbox.zip is included for those who want to use those methods for analysis.

1. Metadata
basename_BasicMetaData.mat: Matlab file with the following variables:

- basename - recording name
- basepath - could/should be replaced by user with path to data folder for this session folder
- goodeegchannel - eeg channel selected for default lfp analyses. Numbering based on the first channel being 1 (not 0)
- goodshanks - shanks with good units. Shanks are the same as electrode groups in .xml/Par (below)
- Par - basic recording parameters including which channels are grouped together, sampling rates. Par is a duplicate of the .xml file
- RecordingFileIntervals - [start stop] times for each recording file in seconds (original recordings were concatenations of sub-recordings)
- Thetachannel - LFP channel used for thetaband signal during sleep scoring. Numbering based on the first channel being 1 (not 0)
- UPstatechannel - LFP channel used for UP state detection (with units on same probe). Numbering based on the first channel being 1 (not 0).

- voltsperunit = Volts represented by each unit of value on channels in the .lfp/eeg file

basename_ChannelAnatomy.csv: comma separated values indicating the anatomical label for each recorded channel. Only channels from electrodes in cortical regions were used in this analysis. Note, while channel numbers in the .xml file start with 0, this list and all other

basename_GoodSleepInterval.mat: (should be irrelevant) A pair of times in seconds stating the start and end points of baseline homecage/sleep conditions valid for these analyses (ie before behavioral paradigms or other interventions). I used this originally but they should be irrelevant as I believe for this shared version of the dataset I have only included data from times in this window.

basename.xml: same information as Par in _BasicMetaData.mat. Used by neuroscope in viewing .xml files (and was used by process_multi_start and klustakwik for clustering). Contains metainfo about file.

2. Local Field Potential voltage recordings:
basename.eeg: 16bit binary data file, containing 1250Hz voltage measurements for each channel recorded. See basename_ChannelAnatomy.csv for a map of which channels are in which brain regions (details below). Voltage per unit can be found in _BasicMetaData.mat voltsperunit variable. Binary data is formatted according to this website: 2 (8 bit) bytes per measurement (total of 16bits), measurements are ordered as follows for a recording with M time points and N channels: timepoint1channel1,timepoint1channel2,…timepoint1channelN, timepoint2channel1,… timepoint2channelN,… timepointMchannelN (end). Can be read with LoadBinary_bw.m (included).

3. Spiking data: Note this contains both spiking metadata and spiking raw data, also included are both spike trains from unstable and stable units (separated).

basename_SAll: Spike trains of all cells recorded including stable and unstable units and E and I cells unsorted. Two formats used, each includes all spike trains and user may choose which to use.

S_CellFormat –spike train for each unit is listed as a vector in one element of a cell array

S_TsdArrayFormat – TsdArray format of all spike trains (see TStoolbox documentation)

basename_SStable: Spike trains of only stable units including and E and I cells – these are the units used in the publication. SAll was transformed to SStable based on the

information in _ClusterQualityMeasures.mat (see below). Two formats used, each includes all spike trains and user may choose which to use.

S_CellFormat –spike train for each unit is listed as a vector in one element of a cell array

S_TsdArrayFormat – TsdArray format of all spike trains (see TStoolbox documentation)

basename_ClusterQualityMeasures.mat – Struct with many fields

- .allbadcells – cells removed for instability or manually-determined poor quality
- .autobadcells – cells automatically removed for >=30% increase or decrease in line of linear fit for metrics; either 1) mahalanobis distance from own cluster or 2) spike energy
- .manualbadcells – cells manually declared problematic, as seen in _ClusteringNotes.csv
- .SelfMahalDistances/Cell – tsdArray/cell array of per-spike total mahalanobis distances of each spike from it's own cluster centroid
- .StartDistances/.EndDistances – First/Last distances used for calculation of comparison vs 30% threshold
- .SelfMahalDistanceChanges – Directly comparable against +- 30% threshold
- .SpikeEnergies/Cell – tsdArray/cell array of per-spike total principal component features of first principal compnonents of all spikes on each channel
- .StartEnergies/.EndEnergies – First/Last distances used for calculation of comparison vs 30% threshold
- .SpikeEnergyChanges – Directly comparable against +- 30% threshold
- .LRatios, .IsoDistances, .ISIIndices – unused, but calculated on quintiles of the recording

basename_ClusteringNotes.csv – Table with vertical list of cells and 1 in "unstable" column if should be listed as manualbadcell

basename_SBurstFiltered – Same as SStable but with any spikes < =6ms after a previous spike removed – used for functional synapse finding. Two formats used, each includes all spike trains and user may choose which to use.

S_CellFormat –spike train for each unit is listed as a vector in one element of a cell array

S_TsdArrayFormat – TsdArray format of all spike trains (see TStoolbox documentation)

basename_CellIDs.mat – putative excitatory (pE) versus putative inhibitory (pI) status of each cell in SStable. Multiple fields:

.EAll – all pE cells (used in paper)

.IAll – all pI cells (used in paper)

.EDefinite – cells with wide waveforms and excitatory effects on CCGs with other cells at 1-4ms lags (see basename_funcsynapsesMoreStringent.mat)

.IDefinite – cells with narrow waveforms and inhibitory effects on CCGs with other cells at 1-4ms lags (see basename_funcsynapsesMoreStringent.mat)

.ELike – cells with wide waveforms but no excitatory effects on CCGs

.IDefinite – cells with narrow waveforms but no inhibotory effects on CCGs

basename_CellClassificationOutput.mat – Struct array of data leading to the data found in basename_CellIDs.mat. 2 fields:

.CellClassOutput – 5 column array

- Column 1: cell numbers in order
- Column 2: trough-peak lag in ms of average waveform for each unit
- Column 3: wavelet-based spike width of average waveform for each unit
- Column 4: neurons classified as having I-like waveforms
- Column 5: neurons classified as having E-like waveforms

.PyrBoundary – polygon surrounding coordinates of wide-waveform cells manually confirmed by user, axes are colum2, column1 (x,y)

basename_SSubtypes – separate spike train arrays for pE and pE cells. Each is denoted SXXX (below) and is in either Cell Array format or tsdArray format (as above in SStable). Some recordings have no pI cells.

- Se – trains of all ELike+EDefinite cells (see _CellIDs.mat)
- Si – trains of all ILike+IDefinite cells (see _CellIDs.mat)
- SeDef – trains of all EDefinite cells (see _CellIDs.mat)
- SiDef – trains of all IDefinite cells (see _CellIDs.mat)
- SeLike – trains of all ELike cells (see _CellIDs.mat)
- SiLike – trains of all ILike cells (see _CellIDs.mat)

basename _MeanWaveforms.mat – Struct array with mean waveforms of max amplitude channels for each unit in SStable

basename_funcsynapsesMoreStringent.mat – Array of synaptic timescale cross correlations between pairs of neurons. Based on finding statistically significant peaks or troughs in cross-correlograms at 1-4ms lags which do not have other obvious explanations based on visual inspection.

Significance determined by convolution with a sliding window across CCGs of burst-filtered pairs of spike trains with significance determined by assumption of poisson standard deviation equal to sliding mean at each bin (Stark 2009). Alpha set at 0.01 for each bin, bin width 0.5ms, interactions must happen between 1 and 4ms of reference unit spike. Candidate "excitatory" interactions must last at least 2 bins (1ms), inhibitory interacitons must last at least 3 bins (1.5ms). Output of automatic detection is in "OriginalConnectivity" field. Other fields of this array are after all candidate synaptic timescale interactions were manually reviewed and all interactions showing evidence of possibly non-synaptic mechanisms underlying (frequently pairs are eliminated for: 0ms-lag interactions, apparent interactions due to detection artifacts of neurons on same recording shank, synaptic timescale noise in ACGs leading to identical signals in CCGs).

- .fullRawCCGMtx – Full matrix of all pairwise interactions 0.5ms bins from full spike trains (burst filtered version is in .fullCCGMtx). Dimensions are: 1) #bins, 2) ReferenceUnit (spikes define 0ms time), 3) TargetUnit
- .ConnectionsE – list of [pre, post] pairs of neurons determined to be significant and excitatory (increased spiking at synaptic timescale) by automatic+manual review

- .ConnectionsI – list of [pre, post] pairs of neurons determined to be significant and inhibitory (decreased spiking at synaptic timescale) by automatic+manual review
- .ConnectionsEE – subset of ConnectionsE found between pairs of EDef or ELike units (see CellIDs.mat for definitions)
- .ConnectionsEI – subset of ConnectionsE found between EDef or ELike "presynaptic" units and "postsynaptic" IDef or ILike units (see CellIDs.mat for definitions)
- .ConnectionsIE – subset of ConnectionsI found between pairs of IDef or ILike units (see CellIDs.mat for definitions)
- .ConnectionsII – subset of ConnectionsE found between IDef or ILike "presynaptic" units and "postsynaptic" EDef or ELike units (see CellIDs.mat for definitions)

4. Sleep state scoring

basename_WSRestrictedIntervals.mat: A series of time intervals of brain/sleep states specified at 1 second resolution by automatic algorithm and manually edited as necessary. Note, each interval below will actually have two formats: TimePairFormat and IntervalSetFormat

TimePairFormat – n x 2 array of n intervals (say n incidences of REM) with column 1 specifying start times and column 2 specifying end times.

IntervalSetFormat – intervalSet object (in seconds) giving times of start and stop for each interval.

Finally, all interval types are defined in Watson 2016, especially in Table 1 of that paper.

- REMTimePairFormat, REMIntervalSetFormat – REM formats used in final analysis, consolidated
- REMEpisodeTimePairFormat, REMEpisodeIntervalSetFormat – REM periods, unconsolidated and relatively unfiltered from automatic algorithm
- SWSEpisodeTimePairFormat, SWSEpisodeIntervalSetFormat – prolonged periods of nonREM spanning between WAKE/WAKEInterruption or REM (includes both SWSPackets and MAs – both defined below)
- SWSPacketTimePairFormat, SWSPacketIntervalSetFormat – Uninterrupted periods of nonREM/SWS sleep spanning between WAKE/WAKEInterruption/REM/MAs
- MATimePairFormat, MAIntervalSetFormat – Microarousals: up to 40 sec long periods with wake-like lfp between Packets (no theta like REM, no delta like nonREM, but short and between nonREM)
- WakeInterruptionTimePairFormat, WakeInterruptionIntervalSetFormat – Longer than MAs, 40-120second periods of WAKE-like activity during sleep. By definition, if was over 120sec would officially interrupt sleep. Longer than MAs, but not long enough to officially stop sleep. Deemed likely actual wake.
- WakeTimePairFormat, WakeIntervalSetFormat – Periods of at least 7 min of WAKE

- SleepTimePairFormat, SleepIntervalSetFormat – Periods of 20+ minutes of sleep (including SWSepisodes, SWSpackets, REM, MA, and WAKE Interruptions)
- WakeSleepTimePairFormat, WakeSleepIntervalSetFormat – Paired periods of WAKE (7+min) followed immediately by SLEEP (20+ min), used as basic unit of our initial analysis

basename_EMGCorr.mat: Data conveying correlation (pearson correlation r values) across probe shanks of lfp signal filtered between 300-600Hz. This measures EMG tone (Schomberg 2014). Data is in format of a struct array with two fields

- .EMGCorr = #of secondsx2 x 2 array. (:,1) gives timepoints in seconds of each measure at 2Hz sampling rate. (:,2) gives mean pairwise r value at each point across all pairwise comparisons across shanks
- .ChannelsCompared – list of channels, compared pairwise for correlation scores

basename_Motion.mat: This of variable quality. Per-second motion across entire recording. Comes from a variety of sources across recordings, including accelerometers, motion pads, movie motion. Not reliable to use, though very good for some recordings. Struct array with 3 fields

- .motion – one point per second of original data, zscored motion for that second
- .filttype – idiosyncratic description of filter used to generate thresholding (next)
- .thresholdedsecs – logical 0 or 1 given to each second based on filttype based on whether motion was above or below threshold

5. Detected events

basename_UPDOWNInts.mat

- UPIntsTimePairFormat, UPIntsIntervalSetFormat – Start/Stops of detected UP states (based on DOWNstates, ONs, OFFs and LowGammaPeriods)
- DNIntsTimePairFormat, DNIntsIntervalSetFormat – Start/Stops of detected DOWN states (based on OFFs and LowGammaPeriods)
- ONIntsTimePairFormat, ONIntsIntervalSetFormat – Start/Stops of detected ON states (based on spiking between OFF states, like Vyazovskiy 2009)
- OFFIntsTimePairFormat, OFFIntsIntervalSetFormat – Start/Stops of detected ON states (based on lack of spiking for 50+ms, like Vyazovskiy 2009)
- GammaIntsTimePairFormat, GammaIntsIntervalSetFormat – Start/Stops of detected ON states (periods of decreased gamma power (proxy for delta waves that does not assume layer))
- UPchannel – lfp channel used for detection of Gamma/UP states

basename_Spindles.mat – detected spindle events in the 10-20Hz range (waves with ZScore 2-7 fold power over baseline, 300-3000ms duration), extremely large amplitude events (ZScore 7-12) are discarded as high-voltage spindles. (Detection is based on a modification of DetectRipples.m from FMAToolbox.)

- .normspindles: #spindles x4 array – one row for each spindle. Column 1 is

spindles start, Column2 is spindle max amplitude time, Column3 is spindle end, Column4 power. Based on Hilbert transform used to detect.
- .maps: +-600ms data of raw trace, phase, frequency, amplitude, each in a separate field
- .data: per-spindle frequency (Hz), amplitude (V), duration(s)

How to get started

This dataset assumes the use of MATLAB. Of course this depends on your questions, but I assume for most purposes the following would work: Start with most refined and finalized states, spikes and LFP. (Unrefined spikes, raw motion and EMG are likely only for those with specific sub-interests.) This includes:

Spikes:
basename_SSubtypes or basename_SStable (just load with matlab load)

LFP:
basename.eeg (load with LoadBinary_bw.m, included)

Brain/Sleep States:
basename_WSRestrictedIntervals.mat (load with matlab load)

Detected Events:
basename_UPDOWNIntervals.mat (load with matlab load)
basename_Spindles.mat (load with matlab load)

How to get help

To get help with the data set post any questions on the forum at CRCNS.org.

References

Stark E, Abeles, M. Unbiased estimation of precise temporal correlations between spike trains. J Neurosci Methods. 2009 Apr 30;179(1):90-100.

Schomburg, E.W., Fernández-Ruiz, A., Mizuseki, K., Berényi, A., Anastassiou, C.A., Koch, C., and Buzsáki, G. (2014). Theta phase segregation of input-specific gamma patterns in entorhinal-hippocampal networks. Neuron 84, 470–485.

Vyazovskiy VV, Olcese U, Lazimy YM, Faraguna U, Esser SK, Williams JC, Cirelli C, Tononi G. Cortical firing and sleep homeostasis. Neuron. 2009 Sep 24;63(6):865-78.