CRCNS.org hc-11 data description

Version 0.8 (Dec 7, 2016)

Recordings from hippocampal area CA1, PRE,
during and POST novel spatial learning

Summary

This data set is composed of eight bilateral silicon-probe multi-cellular electrophysiological recordings performed on four male Long-Evans rats in the Buzsáki lab at NYU. These recordings were performed to assess the effect of novel spatial learning on hippocampal CA1 neural firing and LFP patterns in naïve animals. Each session consisted of a long (~4 hour) PRE rest/sleep epoch home-cage recordings performed in a familiar room, followed by a Novel MAZE running epoch (~45 minutes) in which the animals were transferred to a novel room, and water-rewarded to run on a novel maze. These mazes were either A) a wooden 1.6m linear platform, B) a wooden 1m diameter circular platform or C) a 2m metal linear platform. Animals were rewarded either at both ends of the linear platform, or at a predetermined location on the circular platform. The animal was gently encouraged to run unidirectionally on the circular platform. After the MAZE epochs the animals were transferred back to their home-cage in the familiar room where a long (~4 hour) POST rest/sleep was recorded. All eight sessions were concatenated from the PRE, MAZE, and POST recording epochs. In addition to hippocampal electrophysiological recordings, neck EMG and head-mounted accelerometer signals were recorded, and the animal's position during MAZE running epochs was tracked via head-mounted LEDs.

This dataset has previously been published in:

Grosmark, A.D., and Buzsáki, G. (2016). Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. Science *351*, 1440–1443.

Chen, Z., Grosmark, A.D., Penagos, H., and Wilson, M.A. (2016). Uncovering representations of sleep-associated hippocampal ensemble spike activity. Sci. Rep. *6*, 32193.

Conditions for using the data

If you publish any work using the data, please cite the (Grosmark & Buzsáki, 2016) publication above also cite the data set using the following:

Grosmark, A.D., Long J. and Buzsáki, G (2016). Recordings from hippocampal area CA1, PRE, during and POST novel spatial learning. CRCNS.org
http://dx.doi.org/10.6080/K0862DC5

Methods

The methods for gathering and analyzing this data are described in detail in the supplementary methods of Grosmark and Buzsáki, 2016 and are similar to those of dataset hc-3. Note that for the first three recorded animals ('Buddy', 'Gatsby' and 'Achilles') the power source used to power the accelerometer was instead used to power the head-mounted LED's, therefore, accelerometer data is not available during MAZE running for these animals.

Data files organization

There are eight sessions. Data for each session is supplied in three ".tar.gz" files, with names like: "X_eeg.tar.gz", "X_spk.tar.gz" and "X.tar.gz", where "X" is the base name of the session, e.g. "Achilles_10252013". The "X_eeg.tar.gz" file contains the LFP file, the "X_spk.tar.gz" contains the spike waveforms, and the "X.tar.gz" file contains everything else (e.g. cluster assignments, spike times, ...). After downloading and unpacking, all files associated with the same session (e.g. having the same prefix "X") should be in the same directory, which will be named "X". On Unix systems this can be accomplished by executing the following command in the directory containing the downloaded tar.gz files: `find *.tar.gz -exec tar xzf {} \;` (e.g. if X.tar.gz and X_eeg.tar.gz are both downloaded, executing the command will create a folder named X containing the contents of both of the .tar.gz files). The end result of download all the files and running the above command will be eight folders containing data corresponding to each of the eight novelty sessions.

Each folder contains the concatenated pre-processed data associated with each sessions using the same formats as those in dataset hc-3. All wide-band recordings were performed at 20 kHZ, and downsampled to 1250 HZ to produce the '*.eeg' (LFP) files. All spike extraction and clustering was performed on the wideband (20 kHZ) signal. Due to size constraints (number of channels and the long length of the recordings) the raw wideband data is not included in this dataset. All spike sorting was performed within a given silicon probe shank.

Each session-folder also contains a '*.mat' file (summarized below) with important information such as the spike time, putative cell identities, and tracked position on the maze. For convenience, these '*.mat' files are included in a separate zipped folder 'NoveltySessInfoMatFiles.tar.gz'. A summary pdf ('HC-11_Recording_Summary.pdf') provides summary information on each session's makeup. The recorded channel orders are separately listed in a 'HC-11_Channel_Orders.pdf' file. Note that for ease of use all files channel orders were rearranged such that they reflect the actual spatial configuration of the sites on the probe (i.e. nearby sites have nearby channel numbers) rather than the pseudo-arbitrary order of the corresponding silicon-probe output pins. All provided files reflect this rearrangement.

Data format

The formats of the data within each session follow the standard format used in the Buzsáki lab, as well as other labs, and are described in dataset hc-3 and (Hazan et al., 2006). More details on the details of these formats can be found in the 'Klusters Handbook' (http://klusters.sourceforge.net/UserManual/) . Briefly, each session-folder contains:

>> *baseName.xml* – xml containing information about the session such as channel layout, spiking groups, and sampling frequency.

>> *baseName.eeg* – binary file containing the downsampled (to 1250 HZ) hippocampal LFP, EMG and accelerometer data.

>> *baseName.N.spk* – (where *N* is a shank-number) binary file containing the spike waveform (32 samples from the 20 kHZ raw recorded signal) for each putative (supra-threshold) detected spike event.

>> *baseName.N.res* - binary file containing spike times (in 20 kHZ sampling frequency) for each putative detected spike event.

>> *baseName.N.fet* – binary file containing PCA features (3 per recording site on the shank) used for automatic and subsequent manual spike clustering.

>> *baseName.N.clu* – binary file containing the cluster ID's assigned to each putative spike in the shank. Cluster '0' corresponds to putative electrical noise, and cluster '1' corresponds to putative none-clustarable multi-units.

In addition to these binary files each session contains a ‘baseName_sessInfo.mat’ file (for convenience, these are also bundled separately in the ‘NoveltySessInfoMatFiles.tar.gz’ file). These '.mat' files were saved with the '-v7.3' flag such that they can be opened using the *Python* function '*scipy.io*'. All times are given in seconds. Each of the ‘.mat’ files is a structure with 3 fields:

```
>> sessInfo =
        Spikes: [1x1 struct]
        Position: [1x1 struct]
        Epochs: [1x1 struct]
```

sessInfo.spikes – *discrete spike times, cluster ID’s and putative cell classifications*

```
>> sessInfo.Spikes

        ans =

           SpikeTimes: [5775679x1 double]
           SpikeIDs: [5775679x1 double]
           PyrIDs: [41x1 double]
           IntIDs: [10x1 double]
```

>> *sessInfo.Spikes.SpikeTimes:* spike time (in seconds) for each non-noise putative spike.

>> *sessInfo.Spikes.SpikeIDs:* the spike cluster to which each spike time belongs to. The ‘hundreds’ place of the SpikeIDs corresponds to the number of the shank on which that cluster was recorded such that:

shankNum = floor(ID/100);
cluster_Number_In_Shank = mod(ID, 100);

>> *sessInfo.Spikes.PyrIDs:* a vector of cluster ID’s corresponding to putative pyramidal cells (as determined by CCG’s, spike-width, and AHP).

>> *sessInfo.Spikes.IntIDs:* a vector of cluster ID’s corresponding to putative interneurons

sessInfo.Position – *position information of the animal during the MAZE epoch*

```
>> sessInfo.Position
```

```
ans =

  TwoDLocation: [105180x2 double]
  OneDLocation: [105180x1 double]
  TimeStamps: [1x105180 double]
  MazeType: 'Circular Maze'
```

>> *sessInfo.Position.TwoDLocation:* two-dimensional position of the rat (in meters) on the maze. NaN values corresponds to times when the animal was momentarily outside the field of view of the cameras or when the position tracking failed.

>> *sessInfo.Position.OneDLocation:* One dimensional 'linearization' of the track-running. This is useful for constructing one-dimensional place fields as was done in (Grosmark and Buzsáki, 2016) and many other papers. Note that on the circular maze, the animal received its water reward at position 0.

>> *sessInfo.Position.TimeStamps:* time-stamps (in seconds) corresponding to the position tracking. This video tracking was performed at ~39 HZ. Note that the MAZE epoch followed the PRE epoch and consequently this timestamp vector does not start at 0.

>> *sessInfo.Position.MazeType:* the type of novel maze to which the animal was exposed for this novelty session.

sessInfo.Epochs – *durations of the PRE, MAZE and POST epochs*

>> *sessInfo.Epochs.PREEpoch:* [start, end] vector indicating the beginning and end of the PRE home-cage rest/sleep epoch in seconds.

>> *sessInfo.Epochs.MazeEpoch:* [start, end] vector indicating the beginning and end of the novel MAZE epoch in seconds.

>> *sessInfo.Epochs.POSTEpoch:* [start, end] vector indicating the beginning and end of the POST home-cage rest/sleep epoch in seconds.

>> *sessInfo.Epochs.sessDuration:* total concatenated duration of the recording session in seconds.

In addition '***sessInfo.Epochs***' also contains five [start, end] two-column matrices corresponding to the five behavioral states which were scored based on pyramidal-layer LFP wavelet characteristics (particularly the theta (5-10 Hz) to delta (1 to 4 Hz) ratio) as well movement information (as derived from EMG, accelerometer or position tracking). All state scoring was performed using TheStateEditor developed by Andres Grosmark in the Buzsaki lab (https://github.com/buzsakilab/buzcoderough/tree/master/BehavioralStateDetection).

>> *sessInfo.Epochs.Wake:* [start, end] matrix for active waking epochs, which are characterized by high theta-delta ratio and active movement/EMG-activity.

>> *sessInfo.Epochs.Drowsy:* [start, end] matrix for drowsy-wake/light-sleep epochs which are characterized by low overall spectral power, low movement/EMG-activity. The drowsy-wake/light-sleep state is often observed at the wake-NREM transition, within NREM epochs as well as at the REM-NREM transition.

>> *sessInfo.Epochs.NREM:* [start, end] matrix for nonREM sleep epochs which are characterized by high delta/theta ratio, and very low movement/EMG-activity.

>> *sessInfo.Epochs.Intermediate:* [start, end] matrix for 'intermediate' sleep epochs which are short (~30 second) states which occur at the NREM to REM transition and are characterized by highly-elevated pyramidal layer spinde (12 to 20Hz) power and low movement/EMG-activity. Note that the definition of 'intermediate' sleep used here is different from that used by Gottesmann (1992) – in particular the current definition of 'intermediate' sleep corresponds more closely to the period 'frontal spindling' described by Gottesman, and so is better understood as a particular sub-type of 'NREM' sleep observed around the NREM to REM transition, than a truly 'intermediate' stage.

>> *sessInfo.Epochs.REM:* [start, end] matrix for REM sleep epochs, which are characterized by high theta/delta ratio and very low movement/EMG-activity occurring after NREM episodes.

How to get started

To get started with the LFP data I recommend downloading one or more of the session-folders, and opening them with 'Neuroscope' (http://neurosuite.sourceforge.net/). The spike data can be viewed with 'Klusters' available on this same repository. Alternatively, these files may be loaded into Mathwork's Matlab using '.m' code files associated with dataset hc-3. In order to quickly work with the discrete data, 'NoveltySessInfoMatFiles.tar.gz' file may be downloaded, unzipped and the files therein directly loaded and manipulated in Matlab.

How to get help

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

References

Chen, Z., Grosmark, A.D., Penagos, H., and Wilson, M.A. (2016). Uncovering representations of sleep-associated hippocampal ensemble spike activity. Sci. Rep. *6*, 32193.

Gottesmann, C. (1992). Detection of seven sleep-waking stages in the rat. Neurosci. Biobehav. Rev. *16*, 31–38.

Grosmark, A.D., and Buzsáki, G. (2016). Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. Science *351*, 1440–1443.

Hazan, L., Zugaro, M., and Buzsáki, G. (2006). Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. J. Neurosci. Methods *155*, 207–216.