CRCNS.org alm-1 data description

Version 0.91 (May 8, 2017)

Summary

These experiments probe neural dynamics in the anterior motor cortex (ALM) and its relationship to voluntary movement. The data represent extracellular recordings from ALM neurons of adult mice performing a tactile decision behavior. The data is from 19 animals and includes 99 recording sessions. Results from the experiments are described in:

> A motor cortex circuit for motor planning and movement
> Nuo Li, Tsai-Wen Chen, Zengcai V. Guo, Charles R. Gerfen & Karel Svoboda
> *Nature* 519, 51–56 (05 March 2015) doi: 10.1038/nature14178

The format of the data is described starting on page 3 of this document. MatLab scripts included with the data set are described on the last page ("**How to get started**" section).

Animals and surgery

This study is based on 19 mice (age > P60). 9 Sim1_KJ18-cre mice (MMRRC 031742), 3 Rbp4-cre mice (MMRRC 031125), 4 Tlx_PL56-cre mice (MMRRC 036547), 1 Sim1_KJ18-cre crossed to Ai32 (Rosa26-ChR2 reporter line, Jackson laboratory, JAX Stock#012569), 2 Tlx_PL56-cre crossed to Ai32. Mice were implanted with a clear-skull cap and a headpost [1]. A small craniotomy (approximately 1 mm in diameter) was made over left ALM (2.5mm anterior, 1.5 mm lateral of bregma) one day prior to the recording sessions.

Experimental method

Mice measured the location of an object using their whiskers during a sample epoch (1.3 s) [2]. After the sample epoch they must hold their decision about object location in memory for a delay period (1.3 s) [1]. At the end of the delay period, an auditory cue (0.1) instructed the mice to report their decision with directional licking ("lick left"/"lick right"). Mouth movements were monitored using a photodiode and an infra-red laser diode to measure reaction times.

Extracellular spikes were recorded using NeuroNexus silicon probes (Part# A4x8-5mm-100-200-177) [1]. Silicon probes were introduced acutely before each recording session. The 32 channel voltage signals were multiplexed, recorded on a PCI6133 board at 312.5 kHz (National instrument), and digitized at 14 bit. The signals were demultiplexed into the 32 voltage traces at the sampling frequency of 19531.25Hz and stored for offline analyses. 3-8 recordings were made from each craniotomy. Recording depth was inferred based on manipulator readings. To minimize brain movement, a drop of silicone gel (3-4680, Dow Corning, Midland, MI) was applied over the craniotomy after the electrode was in the tissue. The tissue was allowed to settle for several minutes before the recording started.

To optogenetically tag ALM IT and PT neurons, we first infected ALM neurons with Cre-dependent AAV virus expressing ChR2 (AAV2/5.hSyn1.FLEX.hChR2.tdTomato) [3]. Virus was

injected in ALM into 3 Rbp4-cre mice (targeting both layer5 IT and PT neurons), 4 Tlx_PL56-cre mice (layer 5 IT neuron), and 9 Sim1_KJ18-cre mice (PT neurons). 100 nL volumes were injected each at 500 and 800 μm depth. We also used 2 Tlx_PL56-cre x Ai32 transgenic mice for antidromic tagging of IT neurons.

To photoexcite the PT neurons, an optical fiber was implanted into the reticular formation (5.5 mm posterior, 1mm lateral, 5mm deep), ipsilateral to the viral injection site to target both the axon terminals and passing axons. 1-4 months post virus injection and fiber implant, silicon probe recordings were made from the virus infaction site in ALM. Photostimulation through the optical fiber was used to target PT neuron axons; photostimulation through a cranial window over the contra-lateral ALM was used to target IT neuron axons. Pairs of laser pulses (1 ms duration, 47-82 mW peak power) separated by 10 ms were deployed every 500 ms to elicit antidromic responses from ChR+ neurons. Occasionally, a slightly longer pulse duration (3 or 5ms) necessary to evoke antidromic spikes. Antidromic responses were typically detected on only 1 or 2 recording channels in a recording session.

Data analysis

The extracellular recording traces were band-pass filtered (300-6k Hz). Events that exceeded an amplitude threshold (4 standard deviations of the background) were subjected to manual spike sorting to extract clear single units [1]. Spike sorting was performed on data from individual recording sessions. 1408 single units were extracted across 99 recording sessions. For each unit, its spike width was computed as the trough to peak interval in the mean spike waveform. We defined units with spike width <0.35 ms as fast-spiking GABAergic (FS) neurons (124/1408) and units with spike width >0.45 ms as putative pyramidal neurons (1245/1408). Units with intermediate values (0.35 - 0.45 ms, 39/133) were excluded from our analyses.

73 neurons were antidromically activated by photostimulating the PT neuron axons. 134 neurons were antidromically activated by photostimulating the IT neuron axons. These neurons were further tested for collisions in which we looked for absence of antidromic spikes when they were preceded by spontaneous spikes [4]. Neurons that passed the collision test were classified as PT neurons (45/73) or IT neurons (27/134). Neurons that failed the test were classified as PT-coupled neurons (22/73) or IT-coupled neurons (106/134) because they were presumably synaptically-connected to ChR+ PT neurons and IT neurons. Finally, a few cells could not be tested due to an absence of spontaneous activity (6/73 from PT neurons tagging; 1/134 from IT neurons tagging) and these neurons were excluded from further analyses.

1. Guo ZV, Li N, Huber D, Ophir E, Gutnisky DA, et al. (2014) Flow of cortical activity underlying a tactile decision in mice. Neuron 81: 179-194.
2. O'Connor DH, Clack NG, Huber D, Komiyama T, Myers EW, et al. (2010) Vibrissa-based object localization in head-fixed mice. J Neurosci 30: 1947-1967.
3. O'Connor DH, Hires SA, Guo ZV, Li N, Yu J, et al. (2013) Neural coding during active somatosensation revealed using illusory touch. Nature neuroscience 16: 958-965.
4. Swadlow HA, Waxman SG, Rosene DL (1978) Latency variability and the identification of antidromically activated neurons in mammalian brain. Exp Brain Res 32: 439-443.

Data format

An overview of the data format is given in file "Svoboda_lab_data_format_general.pdf ". The data for each of the 99 sessions is stored in three types of .mat files and in a NWB file:

- "data structure" file (one file per session) – contains processed data. This is the main file that most users will want to use for analysis. These are stored inside directory "data_structure_files" in tar.gz files, each of which contains the all the data structure files from a given animal. (e.g. file data_structure_files/data_structure_ANM210861.tar.gz contains the data structure files for all sessions using animal "ANM210861"). Each file "data_structure_XXX.mat" contains the data from one experimental session (e.g. data_structure_ANM210861_20130701.mat contains data from animal 210861 collected on 2013/07/01).

- "meta data" file (one file per session) – contains basic meta data about the session. Size of each file is about 2k. The meta data files for all of the sessions are stored in compressed archive file "meta_data_files.tar.gz". There is one "meta_data_XXX.mat" file for each "data_structure_XXX.mat" file.

- Data that are in the above two files are also provided in the Neurodata Without Borders: Neurophysiology (NWB) format. The files in the NWB format are stored in the "nwb_files" directory. The name of each NWB file is the same as the corresponding "data structure" file described above, but with extension "nwb". The NWB format is described in the documentation (in particular the specification) linked to from: http://nwb.org. The NWB files are based on the HDF5 format is designed to be fairly understandable by looking at the data files using HDFView.

- raw_trace files (one file per trial; about 400 files per session) - contain the recorded raw voltage traces from each trial in each session. These are stored in tar files inside directory voltage_traces. There is one tar file per session. The contents of each tar file are all the raw voltage trace files for that session.

In addition to the above there is a directory named "old" which contains these same types of files for three example sessions that were initially provided before the full 99 sessions were made available. Most users can ignore this directory because a newer version of the data is provided in the above described files. The data in directory "old" is described by the "README.txt" file in that directory.

Details of the three types of files are given below.

Meta Data File (".mat")

See general description from the lab. (File "Svoboda_lab_data_format_general.pdf"). The meta data is stored in a matlab structure with variable name "meta_data".

Data structure file (also called "processed object file").

Each .mat data file contains data from one session. The data is in the format of matlab structure with name "obj". Each structure contains the following fields:

Top level data description

timeUnitIds: A vector of integers, with the following convention: 1--ms; 2--second; 3--minute; 4--hour; 5—day.
timeUnitNames: Description of time units in **timeUnitIds** (e.g. "second").

descrHash: contains the directory and filename of the raw data file.
descrHash. keyNames: type of data (e.g. "original behavioral data").
descrHash. descr: more detailed description of the entries in **descrHash. keyNames.**
descrHash. value: directory and filename of the entries in **descrHash. keyNames.**

Behavior data

trialTimeUnit: specifies the time unit of the data (refers to **timeUnitIds** above).
trialTypeStr: description of the rows in **trialTypeMat** (e.g. "HitR", "HitL", "Photostim").
trialTypeMat: each column describes one trial by the description in **trialTypeMat** (e.g. a "lick left" trial in which the animal correctly reported choice will have a entry of "1" for "Correct lick left" and "0" for "Error lick left". A phostostimulation trial will have "1" for "StimTrial"). The photostimulation waveform is stored in "**timeSeriesArrayHash.value{(1)}.valueMatrix**" (see below).
trialIds: trial number to reference to the trials.
trialStartTimes: start times of the trials. The time is referenced to session start, (i.e. time 0 is the start of the session).
trialPropertiesHash: contains detailed information about trial structures and timing information. It has the following sub fields:
trialPropertiesHash.**keyNames**: {'PoleInTime' 'PoleOutTime' 'CueTime' 'GoodTrials' 'PhotostimulationType'}.

PoleInTime is the start of sample period for each trial, in units of seconds, relative to **trialStartTimes**.

PoleOutTime is the end of the sample period and start of the delay period.
CueTime is the end of the delay period. Time is in units of seconds, relative to the start of the trials.

GoodTrials has values of "0" or "1"; trials with "0" entries should be discarded for analysis, these indicates periods in the session when mice are not performing (e.g. during periods of passive photostimulation). For non-performing trials, the 'PoleInTime', 'PoleOutTime', 'CueTime' 'PhotostimulationType' all have entries of "NaN".

PhotostimulationType has values of "0", "1", "2", or "NaN". Trials with "0" entries are trials with no photostimulation trials; Trials with "1" entries are trials with PT axonal photostimulation trials; Trials with "2" entries are trials with IT axonal photostimulation trials

(see Methods). Trials with "NaN" entries are photostimulation configurations for testing purposes, these trials should not be analyze.

trialPropertiesHash.**descr**: describes entries in **keyNames.**
trialPropertiesHash.**value**: contains the values of the properties in **keyNames** for each trial. The time is referenced to trial start, (i.e. time 0 is the start of the trial).

timeSeriesArrayHash: contains the time series data for behavioral monitoring and photostimulation. It has the following sub fields:
timeSeriesArrayHash. keyNames: {"EphysVars"}, these data are collected in Ephus acquisition software (ephus.org).
timeSeriesArrayHash.**descr**: describes the content of the data: the data contains recordings of tongue movements (see methods) and photostimulation waveforms.
timeSeriesArrayHash.**value**: contains the data for each trial
obj.timeSeriesArrayHash.value{(1)} contains:
id: e.g. [1 2 3], channel numbers in Ephus acquisition software, the number of channels should match the number of columns in **valueMatrix.**
idStr: e.g. {'lick_trace' 'aom_input_trace' 'laser_power'}, description of the time series data from corresponding channels in **id**.

Channel 1 is typically licking traces.
Channel 2 is typically photostimulation waveform.
Channel 3 is laser power delivered into tissue in units of mW

idStrDetailed: more detailed description of idStr.
timeUnit: time unit used in **timeUnitIds**.
time: e.g. [16150000x1 double], time stamps for the time series data, this is in session time, time 0 is session start. To align to trial start, use **trialStartTimes,** or subfield **trial.**
trial: e.g. [16150000x1 double], trial number for each sample in the time series, the trial numbers are according to **trialIds** (see above).
valueMatrix: e.g. [16150000x3 double], time series data. The number of columns should match number of channels in **id**. The number of samples should match time stamps in **time.**

Ephy data
eventSeriesHash: contains the spike times as well as neuron information; each entry is data from one single unit or multi-unit site
eventSeriesHash.keyNames: name of the entry (e.g. "unit1", or "site1").
eventSeriesHash.descr: description of the entry (e.g. "single unit 1").
eventSeriesHash.value: each entry contains the data structure from one neuron.
Each structure has the following fields:
timeUnit: time unit of the data (refers to **timeUnitIds** above).
eventTimes: spike times for all the events. The time is referenced to session start, (i.e. time 0 is the start of the session). To obtain trial aligned spike time, reference to the trial start time in **trialStartTimes.**
eventTrial: this vector indicates which behavior trial in **trialIds** the spike times were from.
waveform: snippets of spike waveform. Each snippets is 29 samples long (sampled at 19531.25Hz).
depth: estimated depth of the neurons, in micrometers.

channel: which channel on the silicon probe the neuron is recorded from. (see "meta data" file for silicon probe site locations and configuration).
cellType: {"pyramidal" or "FS" or "PT" or "IT"**}.** "Pyramidal", putative pyramidal neuron identified by waveform shape; "FS", putative fast spiking neuron identified by waveform shape; "PT", pyramidal-tract neuron identified by antidromic stimulation and collision test; "IT" corpus callosum projecting neuron identified by antidromic stimulation and collision test. A neuron can have more than one entry (e.g. a PT neurons will have entries{"pyramidal", "PT"}). If a cell has no entries, it is unclassified.

Raw voltage trace data (many ".mat" files)

Each .mat data file contains data from one trial of one session. These files are referenced according to trialID in the processed object file (in "**descrHash.values**").

Each voltage trace data file contain the following variables:

Bitcode_allCh – for housekeeping and synchronization.
Trigger_allCh – for housekeeping and synchronization.
allOther_allCh – for housekeeping and synchronization.
ch_MUA – *n* x 32 matrix, containing 32 channel recording data.
TimeStamps – n x 1 vector, containing time stamps for voltage recording data. Time is referenced to trial start (t=0). (refer to "**trialStartTimes**" in the processed object file).

Only "ch_MUA" and "TimeStamps" are useful. Others are for housekeeping purposes.

How to get started

Directory "matlab_scripts" contains scripts to help get started with the data set. They are in two subdirectories. The "demos" subdirectory contains the following demo scripts:.

> **Demo_get_performance(example_session_data_object) –** extract and plot the behavioral performance data in %correct
>
> **Demo_get_trial_aligned_raster_PSTH(example_session_data_object) –** extract and plot the spike times information and plot the trial aligned PSTH and raster.

Directory "analyses_scripts" has analyses scripts that reproduces the figures in "Li, Chen, Guo, Gerfen, Svoboda (2015).". To run a script, (e.g. "analysis_ALM_population_selectivity"):

1) open MATLAB
2) change current folder to the ".\analyses_scripts\" folder
3) type run('analysis_ALM_population_selectivity')

How to cite the data

If you publish any work using the data, please cite the Nuo et al. (2015), publication above and also cite the data set in the following recommended format:

Nuo Li, Charles R Gerfen, Karel Svoboda (2014); Extracellular recordings from anterior lateral motor cortex (ALM) neurons of adult mice performing a tactile decision behavior. CRCNS.org. http://dx.doi.org/10.6080/K0MS3QNT

Change history

Version 0.9 (Nov. 23, 2015) – Original document.
Version 0.91 (May 8, 2017): added description of files in NWB format.