Data description

This data set was used in the study 'Single-trial neural dynamics are dominated by richly varied movements' by Musall, Kaufman et al. in 2019. It includes data from widefield and 2-photon imaging in awake mice, performing a visual/auditory spatial detection task. It also contains data from Neuropixels recordings in two passive, awake mice. The corresponding Matlab code can be found under http://churchlandlab.labsites.cshl.edu/code/. Most of the analysis was done using Matlab 2018b but earlier versions might work as well. If you have problems or questions, feel free to contact Simon Musall at simon.musall@gmail.com.

The different recording modalities are stored in the folders 'Widefield', '2pData' and 'Neuropixels', respectively. Each main folder contains multiple subfolders, containing data for individual mice. For widefield and 2p imaging data, each individual's folder contains multiple recording sessions. Each recording session is saved in a folder, labeled by the day of acquisition.

Behavioral data

Each session folder contains the behavioral data in a .mat file with the format [AnimalName_SpatialDisc_Date_SessionX], where 'AnimalName' is the ID of the mouse, 'Date' is the recording date, and 'X' a single digit number between 1-3 (depending on how many sessions were acquired that day). Each behavioral file contains a variable 'SessionData' that contains behavioral information as produced by the Bpod v0.5 setup. To learn how to interpret the corresponding entries, refer to the 'delayDec_regressModel.m' code that uses behavioral data to create the linear encoding model.

The folder 'BehaviorVideo' contains dimensionality-reduced video data from two cameras, capturing facial and body movements of the mouse. The most important files in that folder are 'SVD_combinedSegments.mat' and 'motionSVD_combinedSegments.mat'. They contain a low-dimensional representation of the video data and its absolute motion energy. In the linear model, we are using the temporal components of this data which are stored in the variable 'vidV'. It's in the format 'frames by components' and contains the same number of frames as 'Vc' (explained below). Check the 'delayDec_regressModel.m' code to see how to use the video data in the linear model.

Widefield data

The hemodynamic-corrected and dimensionality-reduced widefield data is stored in the file 'Vc.mat'. Each file contains the variables 'Vc' and 'U' which are the temporal and spatial components of the imaging data. Vc is of the form 'components by frames by trials' and U is of the form 'Y by X by components' where X and Y are image coordinates.
To re-create the widefield data (Y by X by frames), U and Vc have to be reshaped first. U should be 'pixels by components' and Vc 'components by frames'. In Matlab, this can be done using:

```
[A,B,C] = size(U);
U = reshape(U, [], size(U,3));
Vc = reshape(Vc, size(Vc,1), []);
```

Afterwards, use U*V to recreate images of the form 'pixels by frames'. For example:

```
testFrames = U*Vc(:,1:10);
testFrames = reshape(testFrames, A, B, []);
```

produces a 3D matrix 'testFrames' that contains the first 10 frames of the current data set. 'Vc.mat' also contains the variable 'bTrials' which indicates which trials in the behavioral data were included in the widefield imaging data. This is important because trials where the mouse did not initiate by touching the handles or failed to respond are not included in the widefield dataset. You can use

bhv = selectBehaviorTrials(SessionData,bTrials);

to create a new variable 'bhv' with the same entries as 'SessionData' but only containing trials that are also present in Vc and U.

2-photon data

The 2-photon session folders have a very similar structure as described above but the pre-processed imaging data is stored in the file 'data.mat' instead of 'Vc.mat'. 'data.mat' contains a struct 'data' containing the processed 2-photon data and some additional information. Instead of using the existing variable 'bTrials' as explained above, one should use data.trialNumbers instead. The field data.dFOF contains the de-noised calcium data that is used in the linear model. You can use the same code 'delayDec_regressModel.m' with the input 'dType = 'twoP' to see how it was analyzed in the linear model.

Neuropixels data

The 'Neuropixels' folder contains two recordings from two chronically implanted mice. Refer to the code 'delayDec_NeuroPixels.m' to see how to process and use in the linear model. There are three important files in each recording. First, a video file of the format [AnimalName_XXX_Video_1.mp4] which 'AnimalName' being either NP9 or NP14 and 'XXX' being 006 or 008, correspondingly. Second, the pre-processed spiking data of the format [AnimalName_XXX_spikes.mat] and lastly some temporal/stimulus information in the file [AnimalName_SessionXXX_sync.mat].