Enabling Naturalistic, Long-Duration and Continual Experimentation

Joaquin Rapela

February 11, 2025

Contents

1 Vision **3**

2 Data acquisition, management and quality control **4**
- 2.1 Deliverables . . . 4
- 2.2 Previous work . . . 4
- 2.3 Future work . . . 5

3 Data sharing **6**
- 3.1 Deliverables . . . 6
- 3.2 Previous work . . . 6
- 3.3 Future work . . . 6

4 Data visualisation **7**
- 4.1 Need for cloud-based visualisation . . . 7
- 4.2 Deliverables . . . 7
- 4.3 Questions . . . 7
- 4.4 Previous work . . . 8
- 4.5 Future work . . . 8

5 Spike sorting **9**
- 5.1 Deliverables . . . 9
- 5.2 Previous work . . . 9

6 Data analysis **10**
6.1 Deliverables 10
6.2 Previous work 10

1 Vision

At the Sainsbury Wellcome Centre for Neural Circuts and Behavior (SWC) and at the Gatsby Computational Neuroscience Unit (GCNU) we are performing a radically new type of experimentation, that is Naturalistic, Long Duration, and Continual (NaLoDuCo). We are recording bahavioral and electrophsiological data continuosly for weeks to months, while mice forage, indidually or socially, in large naturalistic arenas.

The Allen Institute for Neural Dynamics (AIND) is also creating software and hadware technology, and doing research, on foraging along two parallel paths. First, they are investigating the neural basis of foraging behavior in shorter virtual reality experiments on head fixed mice. Second, they are studying how freely moving mice learn while they naturally explore odors in long-duration experiments.

The SWC/GCNU and the AIND have complementary expertise in foraging research. The SWC/GCNU are experienced in NaLoDuCo experimentation. Since 2021 they have been creating software and hardware infrastructure to enable NaLoDuCo foraging research, and have successfully used this technology to perform simultaneous behavioral and neural recordings in mice foraging in large arenas for weeks to months. The GCNU is a leader in developing advanced neural data analysis methods, that will be essential to understand data generated by NaLoDuCo experiments. The AIND is also experienced in foraging research, and is a pioneer in generating and openly disseminating high-quality neural datasets at scale.

Here we propose to join forces to co-develop NaLoDuCo experimentation technology and disseminate it openly to facilitate the adoption of NaLoDuCo experimentation by research groups around the world.

2 Data acquisition, management and quality control

2.1 Deliverables

1. hardware specifications for long-duration recordings of behaviour and neural activity (foraging arenas, video cameras, ultrasound microphones, weight scales, Neuropixels probes, commutator, ONIX)
2. software for managing long-duration recordings (data storage, data indexing)
3. software for online/offline quality control
4. software for online data visualisation
5. software for online data analysis
6. software for intelligent close-loop control of experiments

2.2 Previous work

- at the SWC we have performed foraging experiments
  - lasting xx weeks and recording behaviour only
  - lasting yy weeks and recording behaviour and electrophysiology
  - data is stored in files and in a MySQL database
- items 1–3: above have been completed for the SWC foraging experiments
- item 4: we have developed some online data visualisation tools in Bonsai
- items 5 and 6: funded by BBSRC, we have integrated into Bonsai tools for online data analysis, to be used for the close-loop control of experiments
  - estimate kinematics of mice

  - estimate kinematic states of mice using Hidden Markov Models
  - clusterless point-process decoder of mice position and replay from spikes

- disseminated detailed documentation on software used at the SWC to control NaLoDuCo experiments (see repo)
- disseminated detailed documentation on machine learning methods integrated into Bonsai for analyzing behavioral and neural tie series in real time (see repo)

2.3 Future work

- item 3: at the SWC and at the AIND we have developed several tools for offline quality control. We next need to build online versions of them.
- item 4: develop more software for data visualisation.
- items 5 and 6: develop more software for online data analysis.
  - online estimate of latent variables from Neuropixels recordings.
  - online estimate of RL models.
- help Dr. Carl Schoonover (AIND) use hardware and software developed at the SWC/GCNU to create his olfaction NaLoDuCo experiments.

3 Data sharing

3.1 Deliverables

1. dashboard (or examples of how) to convert data collected in NaLoDuCo experiments to the Zarr format, and to upload the Zarr files to DANDI.

2. dashboard to stream data collected from NaLoDuCo experiments to DANDI

3.2 Previous work

- the AIND is experienced on sharing their recordings on DANDI. However, these recordings are not as large as those in NaLoDuCo experiments and they are not continual.

3.3 Future work

- develop dashboard to convert data collected in NaLoDuCo experiments to the Zarr format

- develop dashboard to stream data collected from NaLoDuCo experiments to DANDI

- test dashboards on data from:

    1. freely-moving foraging mice (SWC)
    2. head-fixed foraging mice (AIND)
    3. freely-moving odour exploration mice (AIND)

- DANDI is typically used to store neurophysiological datasets much smaller than those generated in NaLoDuCo experiments. Conventional methods to access data in DANDI may not be fast enough to allow performant data visualisation and/or data analysis. We may need to explore parallel computing and/or resource efficient cloud configurations (i.e., optimising cloud configurations to improve runtime performance).

4 Data visualisation

4.1 Need for cloud-based visualisation

A unique feature of the NaLoDuCo recordings collected at the SWC and the AIND, is that they are long-duration and continual. Greatest insights will come from investigating these recordings as a whole, and not by analysing separately its parts. For example, the analysis of shorter duration recordings will not be able to capture long-term temporal dependencies in neural activity, that could be critical to understand infradian modulations of behaviour. Thus, we need software infrastructure to browse and visualise week- to month-long experimental recordings on the order of hundreds of terabytes. It is not feasible to download these huge datasets in order to visualise them. Hence, **offline data visualisation needs to be done on the cloud**, as in Neurosift.

4.2 Deliverables

1. web-based dashboard for **online** visualisation of quality control measures.
2. web-based dashboard for **online** data analysis and visualisation of its results.
3. web-based dashboard for **offline** visualisation of NaLoDuCo behavioural and neural recordings on DANDI.
4. web-based dashboard for **offline** visualisation of data analysis results on DANDI.

4.3 Questions

- can I see the visualisation tools from the AIND?
- does the AIND has visualisation tools running on the cloud?
- what visualisation tools do we have at the SWC? developed by Datajoint?
- is the AIND collaborating with Jeremy Magland?

4.4 Previous work

- Neurosift (repo, paper) allows to visualise shorter-duration behavioural and neural recordings in DANDI.
- Dendro (repo) allows to perform analysis on the cloud and visualise the results of such analysis
- offline and precomputed visualisations developed at the SWC, with the help of Datajoint.
- offline and precomputed quality control visualisations developed at the AIND.
- offline and precomputed visualisations developed at the AIND for shorter-duration experiments.
- offline and precomputed visualisations from IBL for short-duration experiments.

4.5 Future work

- Neurosift has been designed to visualise relatively short duration datasets. We will extend it with data pyramids (e.g., `https://github.com/carbonplan/ndpyramid`) to enable it to operate on long-duration recordings.

5 Spike sorting

5.1 Deliverables

1. method (and software implementation) for sorting spikes from Neuropixels probes in long-duration experiments
2. software for curation of results of sorting spikes from Neuropixels probes in long-duration experiments
3. quality control measures for the results of sorting spikes from Neuropixels probes in long-duration experiments

5.2 Previous work

- methods exist for sorting, curating and quality control spikes from short duration experiments
- Dr. Carl Schoonover (AIND) has developed methods to sort spikes from long-duration experiments (find out more about these methods)
- the SWC has managed to sort spikes from a small subset of channels of a Neuropixels probe (how many channels? what duration?)

6 Data analysis

6.1 Deliverables

1. methods to analyse behavioural and electrophysiological recordings from NaLoDuCo experiments that are **online** and **adaptive to non-stationarity** in measurements.

   For behavioural data, we will investigate methods to

   - track multiple body parts of animals
   - infer kinematics of foraging mice
   - segment behaviour into discrete states
   - infer the rules that govern mice behaviour from behavioural observations only] (i.e., policy inference).

   For neural data, we will investigate methods to:

   - track multiple body parts of animals
   - estimate low-dimensional continual representations of neural activity (i.e., latents inference)
   - segment neural activity into discrete states
   - decode environment variables from neural activity

2. integration of these methods into DANDI, as in Dendro, so that users can run them on NaLoDuCo datasets stored in DANDI.

6.2 Previous work

- At the Gatsby we have invented several methods for the characterisation of neural time series (e.g., Yu et al., 2009; Duncker and Sahani, 2018; Rutten et al., 2020; Yu et al., 2024; Buesing et al., 2012b,a; Macke et al., 2015; Soulat et al., 2021; Walker et al., 2023; Turner and Sahani, 2014; O'Shea et al., 2022)

- Dendro allows to perform advanced data analysis on DANDI. It allows to reuse previous analysis.

References

Buesing, L., Macke, J. H., and Sahani, M. (2012a). Learning stable, regularised latent models of neural population dynamics. *Network: Computation in Neural Systems*, 23(1-2):24–47.

Buesing, L., Macke, J. H., and Sahani, M. (2012b). Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. In *Advances in Neural Information Processing Systems 25: 26th Conference on Neural Information Processing Systems (NIPS 2012)*, pages 1691–1699.

Duncker, L. and Sahani, M. (2018). Temporal alignment and latent gaussian process factor inference in population spike trains. In *Advances in Neural Information Processing Systems*, pages 10445–10455.

Macke, J. H., Buesing, L., and Sahani, M. (2015). *Estimating state and parameters in state space models of spike trains*, page 137–159. Cambridge University Press.

O'Shea, D. J., Duncker, L., Goo, W., Sun, X., Vyas, S., Trautmann, E. M., Diester, I., Ramakrishnan, C., Deisseroth, K., Sahani, M., et al. (2022). Direct neural perturbations reveal a dynamical mechanism for robust computation. *bioRxiv*, pages 2022–12.

Rutten, V., Bernacchia, A., Sahani, M., and Hennequin, G. (2020). Non-reversible gaussian processes for identifying latent dynamical structure in neural data. *Advances in neural information processing systems*, 33:9622–9632.

Soulat, H., Keshavarzi, S., Margrie, T. W., and Sahani, M. (2021). Probabilistic tensor decomposition of neural population spiking activity. In *Advances in Neural Information Processing Systems*.

Turner, R. E. and Sahani, M. (2014). Time-frequency analysis as probabilistic inference. *IEEE Transactions on Signal Processing*, 62(23):6171–6183.

Walker, W. I., Soulat, H., Yu, C., and Sahani, M. (2023). Unsupervised representation learning with recognition-parametrised probabilistic models. In *International Conference on Artificial Intelligence and Statistics*, pages 4209–4230. PMLR.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of neurophysiology*, 102(1):614–635.

Yu, C., Sahani, M., and Lengyel, M. (2024). Discovering temporally compositional neural manifolds with switching infinite gpfa. *bioRxiv*, pages 2024–10.