Evaluating Contrastive Embeddings of Extracellular Dynamics for Cell Typing with ‘Ground Truth’ Data

Statistics for Neuroscience 2023

Aims

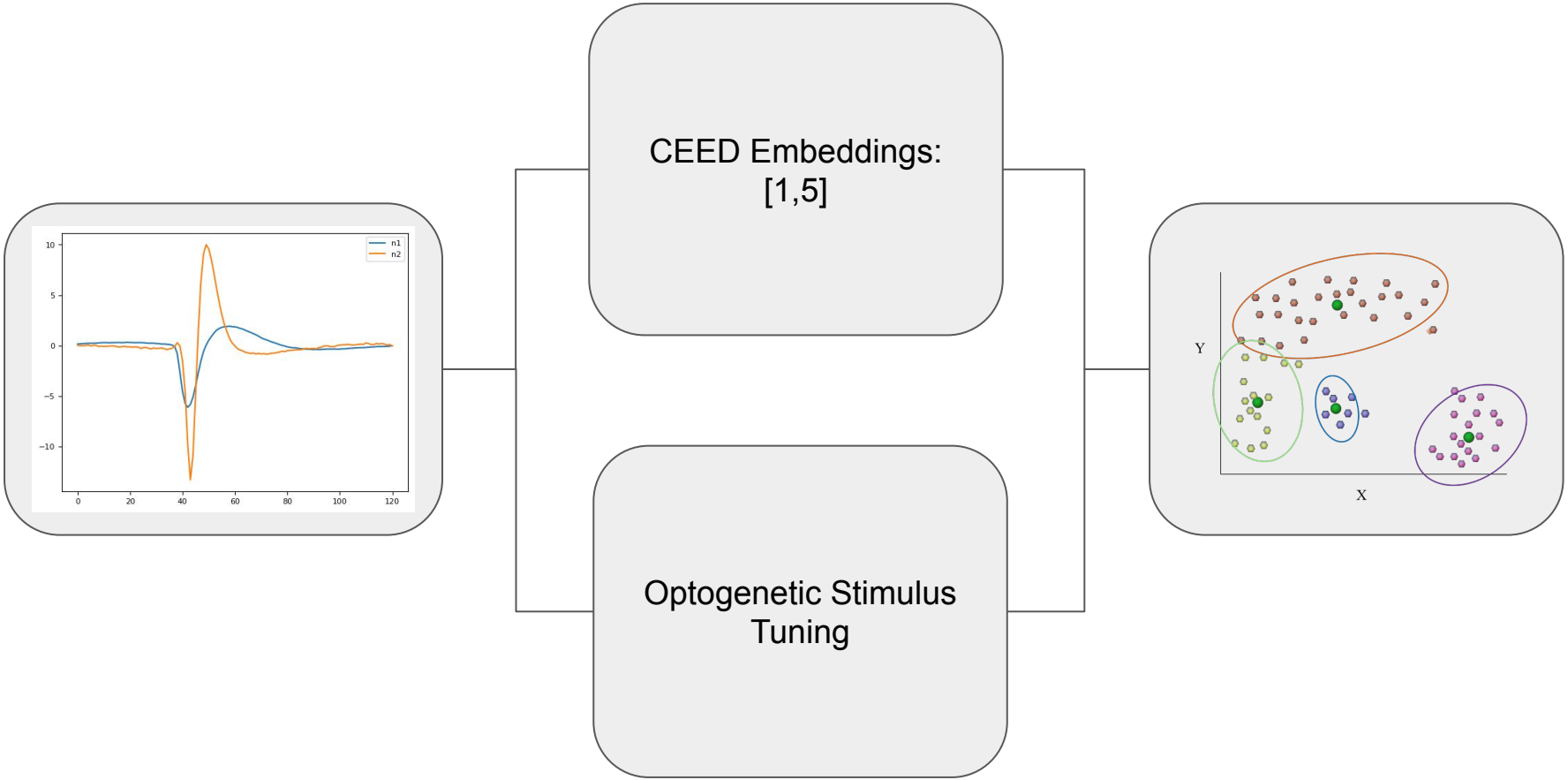
- **Ephys** data from MEAs can record information from 1000s of cells, but with only waveforms and location it can be difficult to cluster individual units and distinguish different cell types.

- Computational techniques can be used to address these problems, however without **ground truth** data they are **difficult to verify**

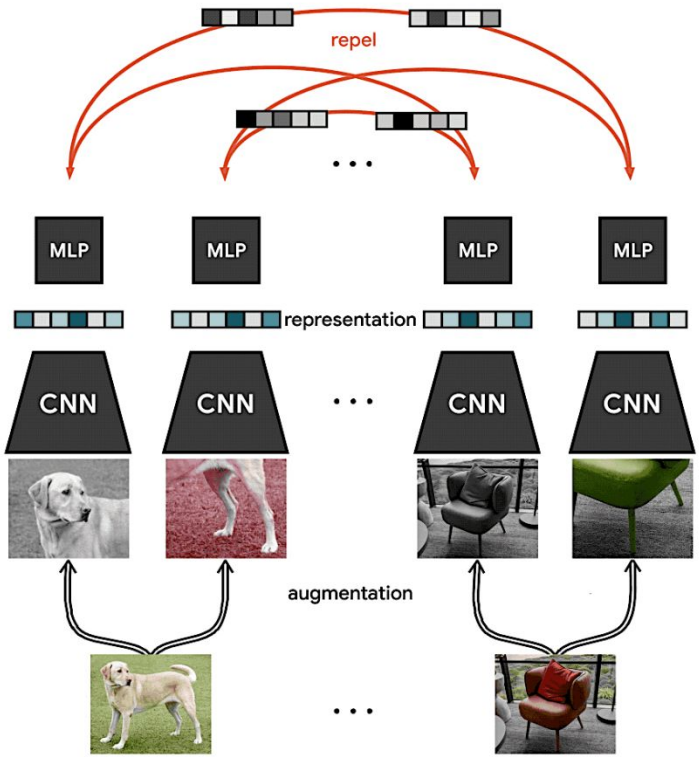
- **Contrastive learning** (CEED) has been proposed as a method to **identify cell types** based on their extracellular dynamics, but are these clusters biologically meaningful?

- Using the allen institute's **optotagging** dataset allows some insight into the genetic properties of cells recorded with NPs

Overview:


n1
n2
10
5
0
−5
−10
0
20
40
60
80
100
120
CEED Embeddings:
[1,5]
Optogenetic Stimulus
Tuning
Y
X

Contrastive Learning



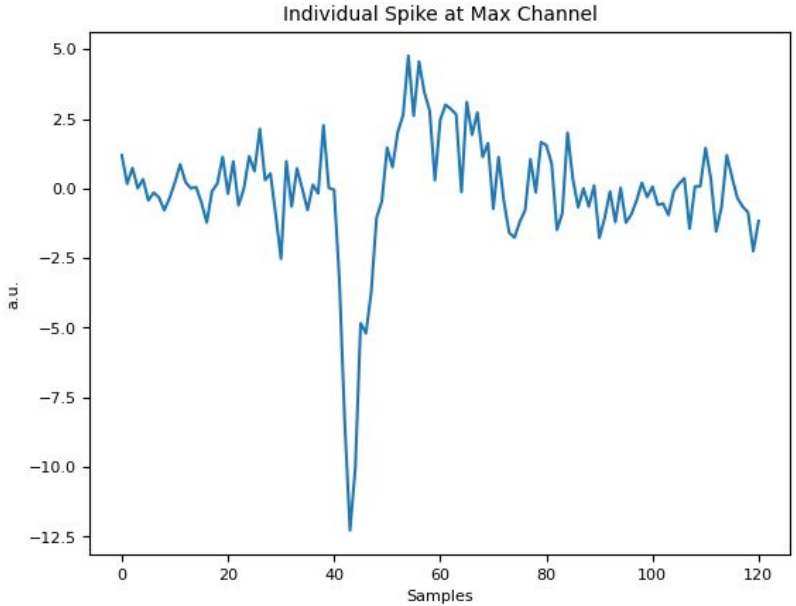
- **Self-supervised** method to generate latent representations of objects
- Relies on **data augmentation** to create multiple views of data
- Minimises distance between representations of similar objects, maximises distance between different objects

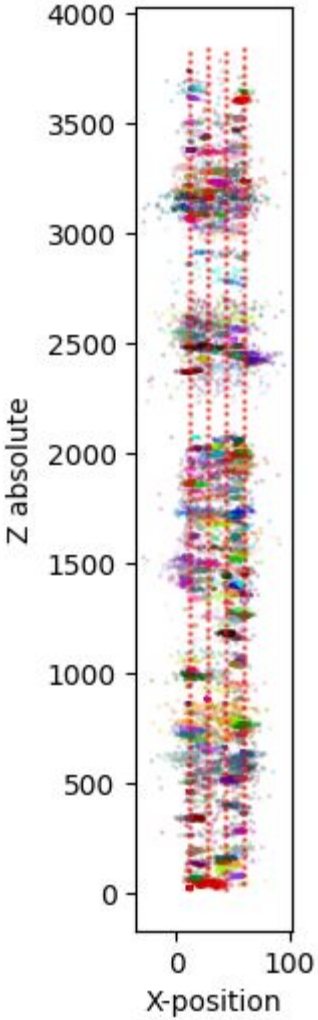
Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *International conference on machine learning*. PMLR, 2020.

Electrophysiology Background

- Ephys Data is **Clustered** into units using spike location, and a descriptor of the spike's waveform
  - E.g. ptp amplitude and maximum channel location
- Other **Dimensionality Reduction** methods can be used to further split or merge these clusters

- There are many **cell-types** throughout the mouse brain, understanding what units correspond to different cell types would be very useful.
- This is **not trivial** with NP data.





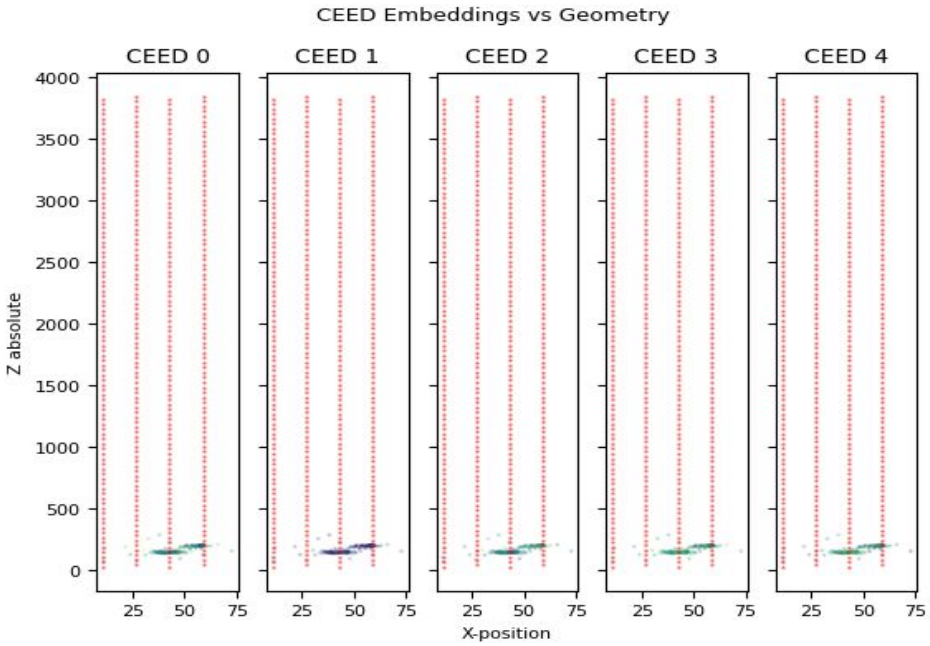
Why CEED embeddings

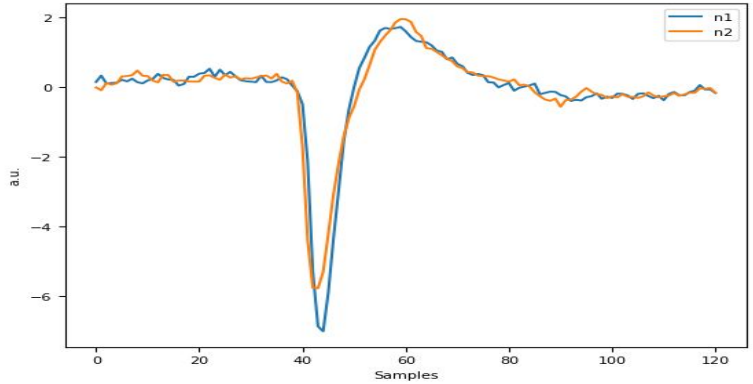
Contrastive Embeddings of Extracellular Dynamics

A representation of waveform shape insensitive to noise, temporal jitter, and other nuisance variables.

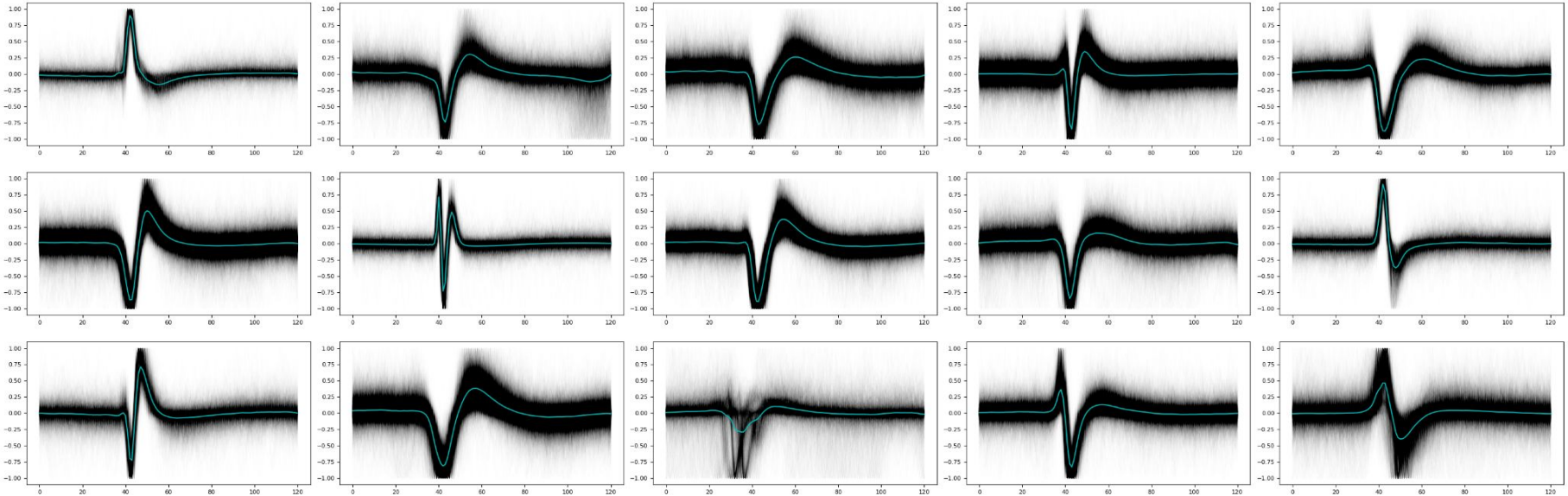
Can allow a more ‘human’ interpretation of a waveform shape

This representation could improve upon clustering methods previously mentioned and possibly allow for waveform based cell-type classification



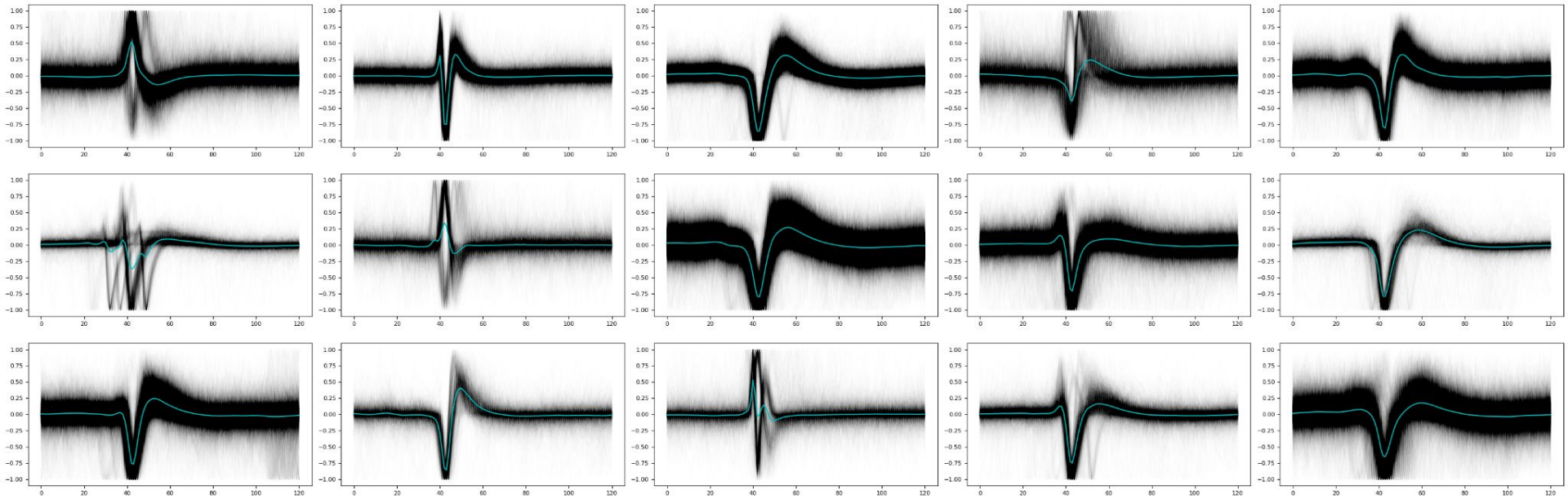


CEED for Cell Typing



GMM Clustered CEED embeddings group similar waveforms well. But, does waveform clustering == Cell-typing?

CEED for Cell Typing

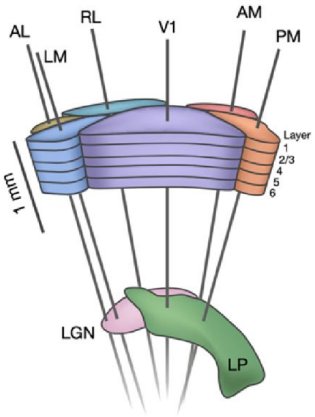


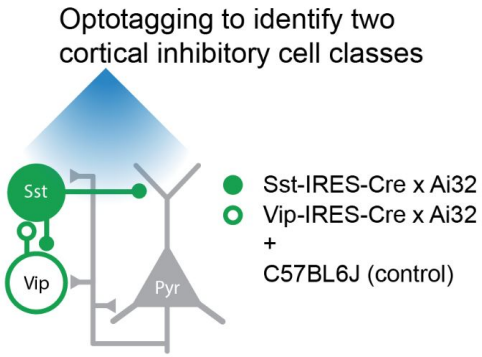
The same method produces seemingly worse clusters when using hand picked features, unsure if it’s better than PCA

Optotagging Dataset

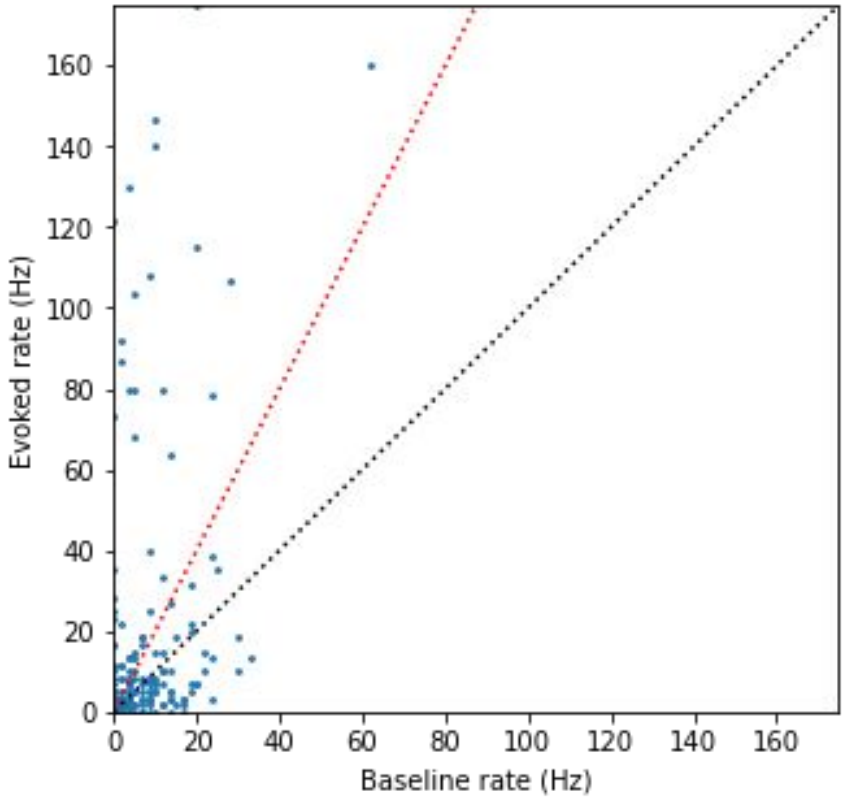
- Neuropixel Data recorded from mouse brain - filtered to only include cortical neurons
- 3 genotypes used: Vip, Sst, Pvalb. Tagged with ChR2, allowing stimulation with blue light
- Allows cell-type identification based on tuning to this stimulus

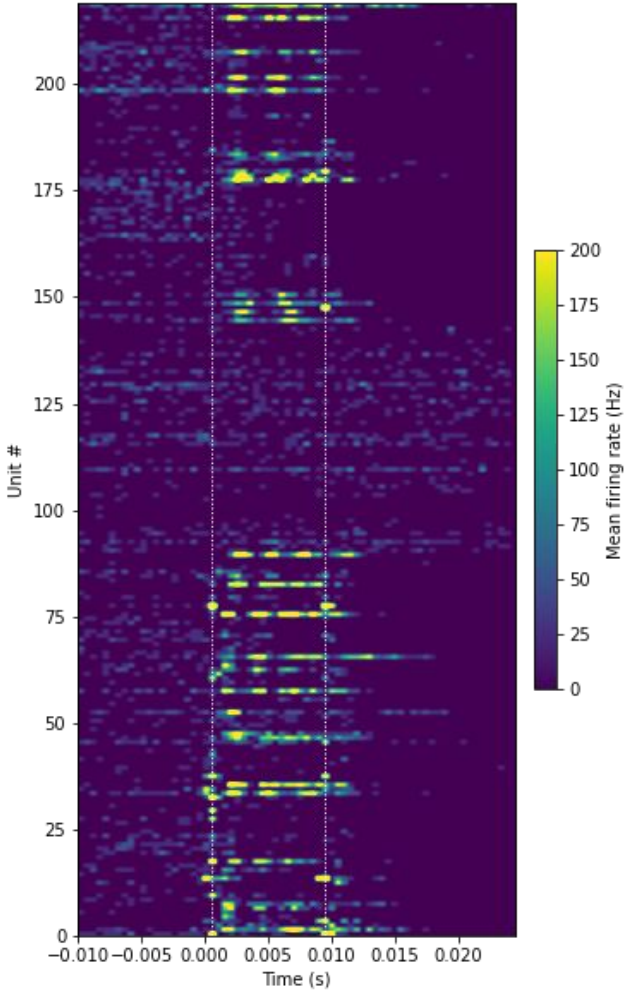
- Only templates - no single spikes
- Allen processing is different than IBL, ‘artifactual waveforms’ are removed



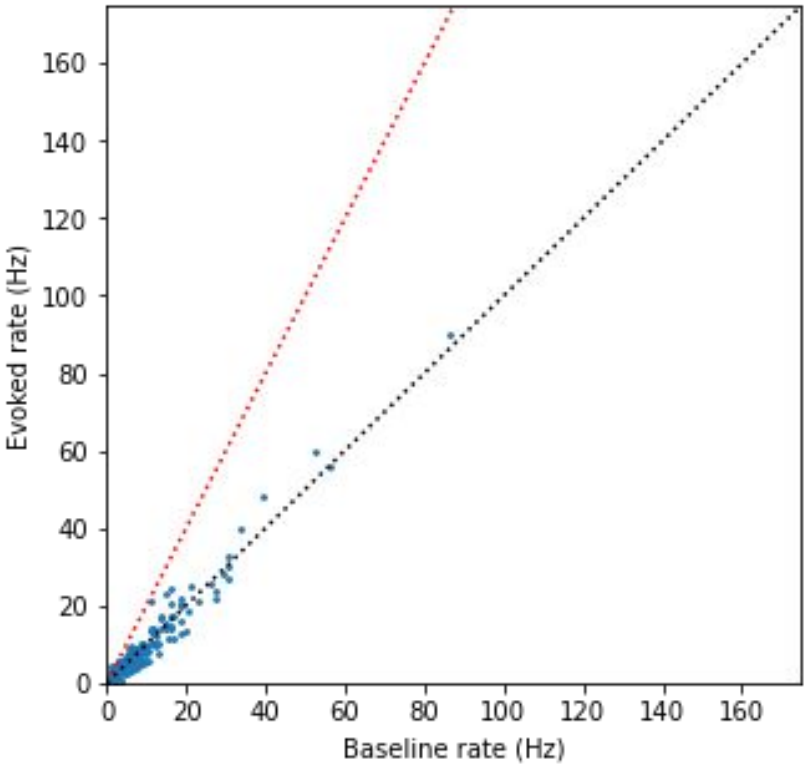


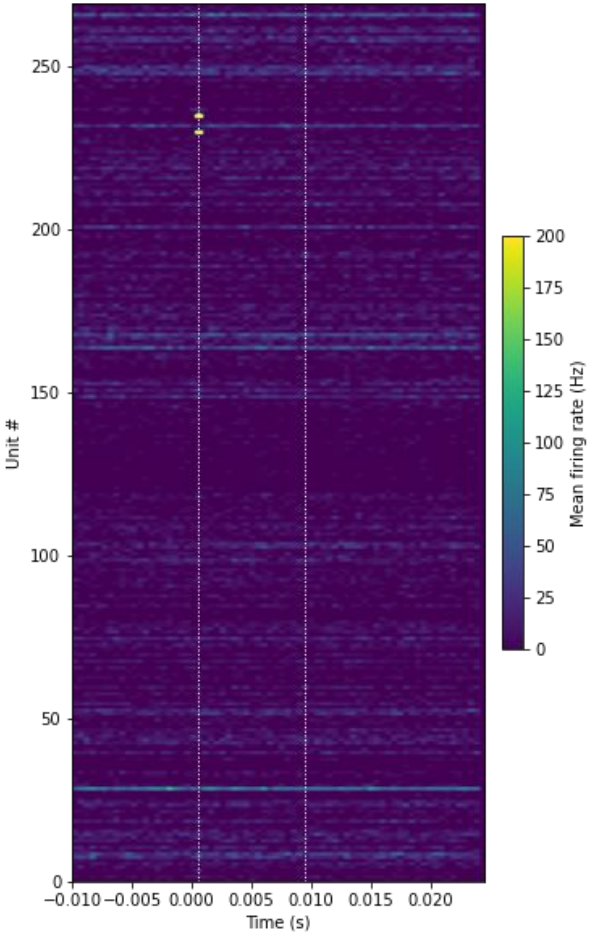
Extracting Tuned Units



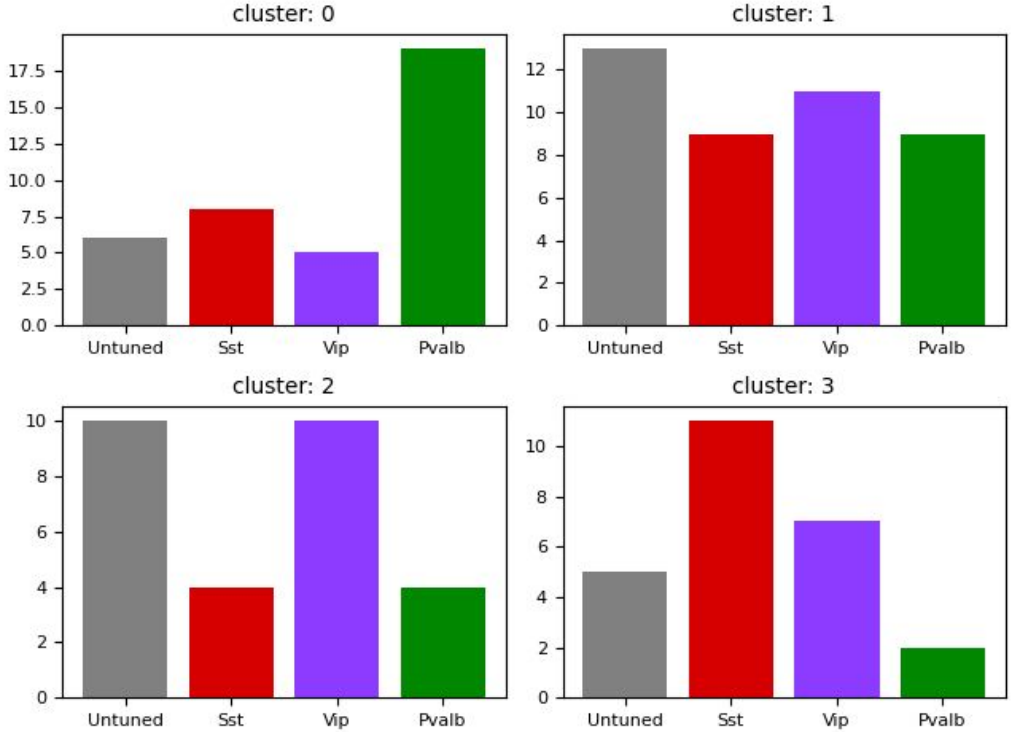


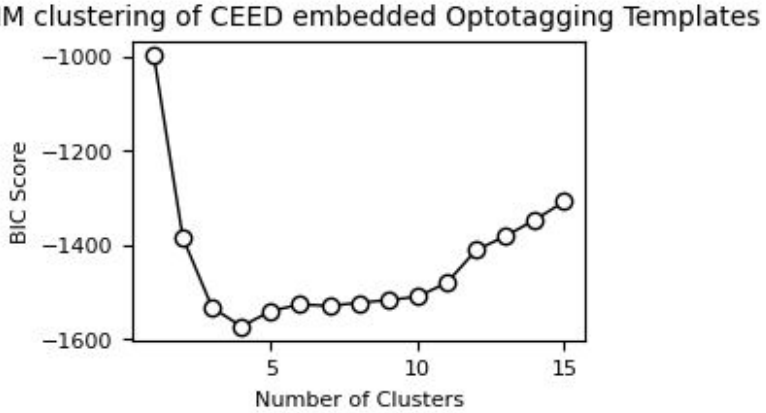
Wild Type Comparison

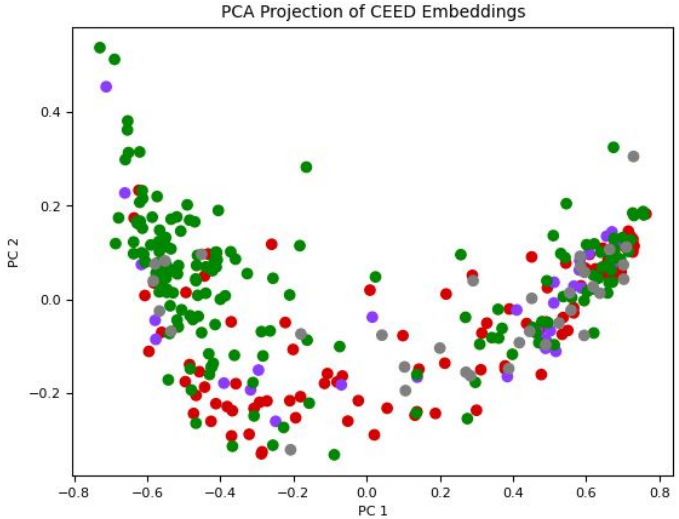




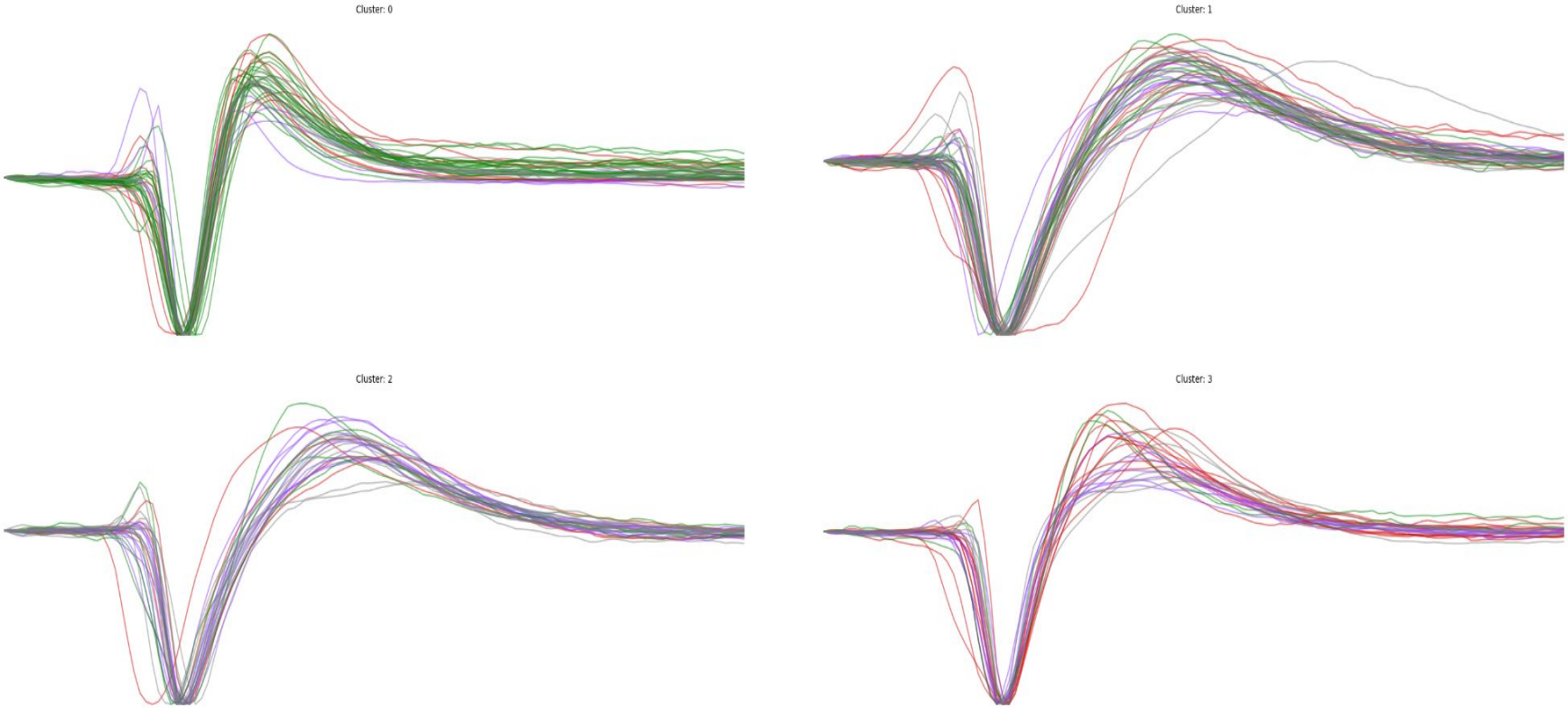
Results


cluster: 0
cluster: 1
cluster: 2
cluster: 3
Untuned
Sst
Vip
Pvalb


GMM clustering of CEED embedded Optotagging Templates
BIC Score
Number of Clusters


PCA Projection of CEED Embeddings
PC 1
PC 2

Results



What’s not perfect

Tuning selection is brute force

Not individual spikes

Different processing makes it hard to compare to IBL data

Questions