The Gatsby Data Center
Distributing and Applying Advanced Statistical Tools to Understand How Neural Circuits Generate Behavior

Joaquín Rapela

Gatsby Computational Neuroscience Unit
University College London
j.rapela@ucl.ac.uk

May 20, 2020

My Academic History

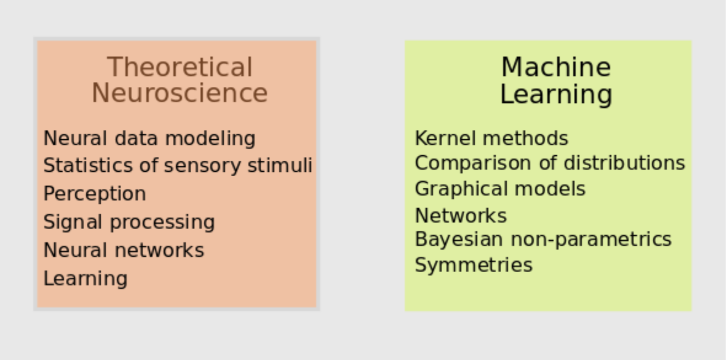
PhD in Electrical Engineering (Signal Processing) – University of Southern California

Postdoc on statistical analysis of EEG recordings – University California San Diego

Postdoc on statistical analysis of recordings from patients with epilepsy– Brown University

Postdoc (and hopefully research engineer soon) building the Gatsby Data Center – Gatsby Computational Neuroscience Unit

Gatsby Computational Neuroscience Unit



- Faculty: 4
- Postdocs: 9
- PhD students: 19

Sainsbury Wellcome Center

We aim to understand how computation in neural circuits gives rise to flexible, complex behaviour.

Sainsbury Wellcome Center

We aim to understand how computation in neural circuits gives rise to flexible, complex behaviour.

- *How does the brain generate and update representations of the world?*
- *How does the brain make decisions and select actions?*
- *How does the brain learn and remember?*

Sainsbury Wellcome Center

We aim to understand how computation in neural circuits gives rise to flexible, complex behaviour.

- *How does the brain generate and update representations of the world?*
- *How does the brain make decisions and select actions?*
- *How does the brain learn and remember?*

Because these questions cannot be answered by any one laboratory, we are developing new models of team science,

Sainsbury Wellcome Center

We aim to understand how computation in neural circuits gives rise to flexible, complex behaviour.

- *How does the brain generate and update representations of the world?*
- *How does the brain make decisions and select actions?*
- *How does the brain learn and remember?*

Because these questions cannot be answered by any one laboratory, we are developing new models of team science, as well as developing exciting new technologies, such as high density Neuropixels probes, which permit recordings from thousands of neurons across multiple brain regions during behaviour.

Sainsbury Wellcome Center

We aim to understand how computation in neural circuits gives rise to flexible, complex behaviour.

- *How does the brain generate and update representations of the world?*
- *How does the brain make decisions and select actions?*
- *How does the brain learn and remember?*

Because these questions cannot be answered by any one laboratory, we are developing new models of team science, as well as developing exciting new technologies, such as high density Neuropixels probes, which permit recordings from thousands of neurons across multiple brain regions during behaviour. These and other emerging technologies will generate entirely new types of datasets at the spatiotemporal scale relevant for distributed computations underlying behaviour . . .

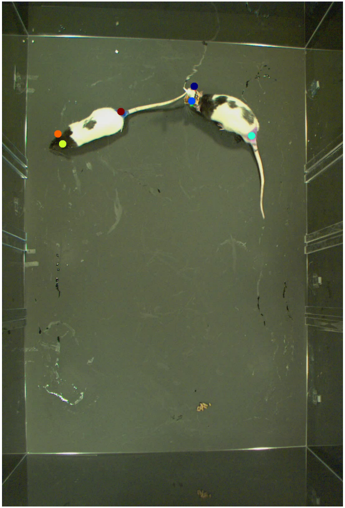
Sainsbury Wellcome Center

We aim to understand how computation in neural circuits gives rise to flexible, complex behaviour.

- *How does the brain generate and update representations of the world?*
- *How does the brain make decisions and select actions?*
- *How does the brain learn and remember?*

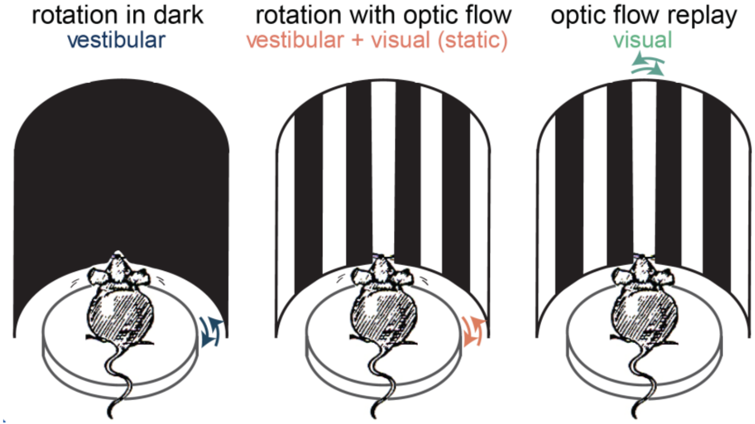
Because these questions cannot be answered by any one laboratory, we are developing new models of team science, as well as developing exciting new technologies, such as high density Neuropixels probes, which permit recordings from thousands of neurons across multiple brain regions during behaviour. These and other emerging technologies will generate entirely new types of datasets at the spatiotemporal scale relevant for distributed computations underlying behaviour . . . Through multi-disciplinary team research, and in close collaboration with the Gatsby Computational Neuroscience Unit (GCNU), scientists at the SWC are generating new theories that relate computational algorithms to identified neural circuits to explain different aspects of cognition and behaviour.

Example experiment (1/3): behavioral and electrophysiological recordings related to social behavior



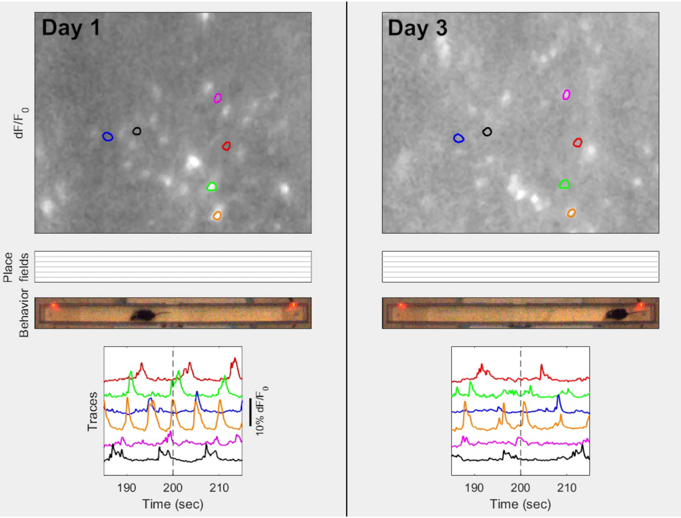
Cristina Mazuski (SWC)

Example experiment (2/3): electrophysiological recordings related to the integration of visual and vestibular information



Sepiedeh Keshavarzi (SWC)

Example experiment (3/3): optical recordings related to spatial navigation

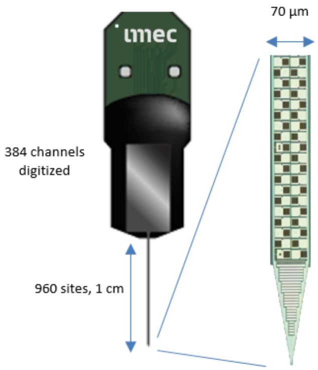


Sheintuch et al., 2017.

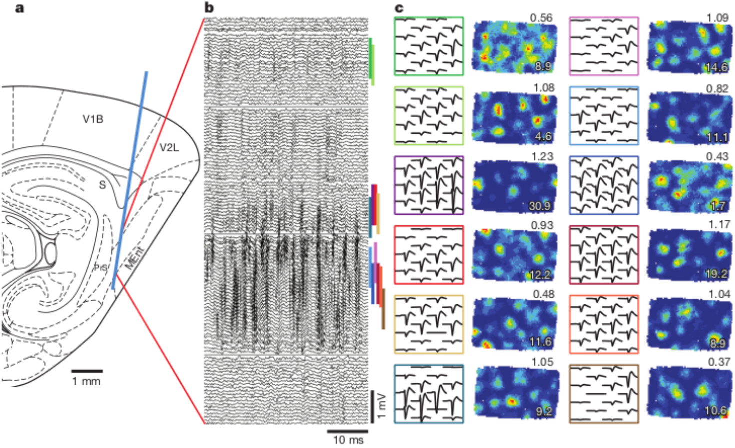
Partial Summary

The SWC is doing sophisticated experiments to understand the relation between neural circuits and behavior.

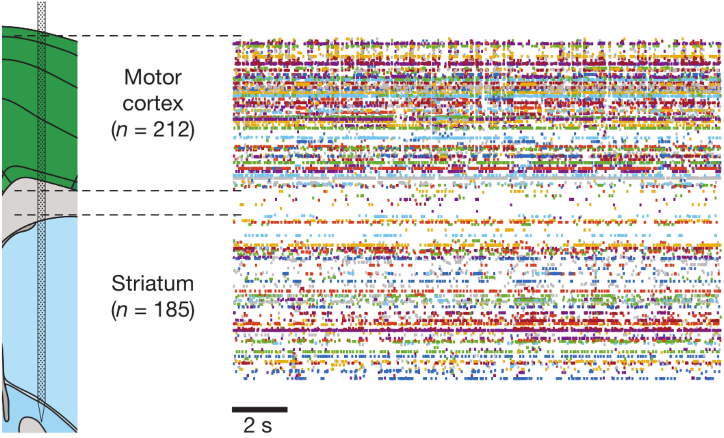
Neuropixels Probes



Neuropixels Waveforms



Neuropixels Spikes



Partial Summary

The SWC is recording (electrically and optically) the activity of large populations of single neurons with state of the technology.

Gatsby Data Center

Mission:

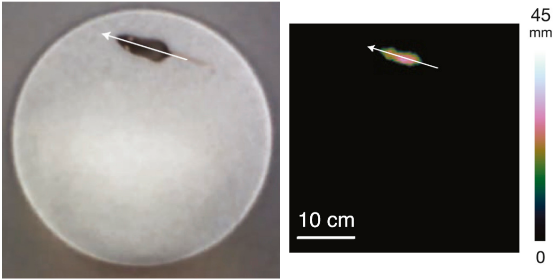
- global: distribute high-quality implementations of advanced statistical methods devised at the unit, and allow neuroscientists without training in statistics to perform sophisticated analysis of their neural recordings,
- local: assist neuroscientists at the SWC in the use of advanced statistical methods to model their recordings.

Overview

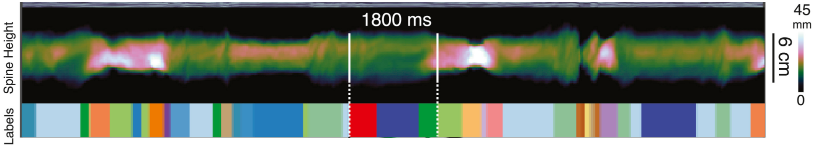
1. Collaboration: discovering the syllables of behavior in mouse social interactions
2. Distribution of the Sparse Variational Gaussian Processes Factor Analysis Python package
3. Collaboration: characterizing neural populations related to the integration of visual and vestibular information
4. Conclusions

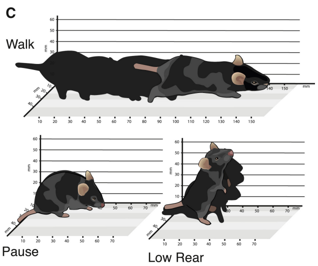
Table of Contents

Behavioral syllables for isolated mice have been reported (1/2)

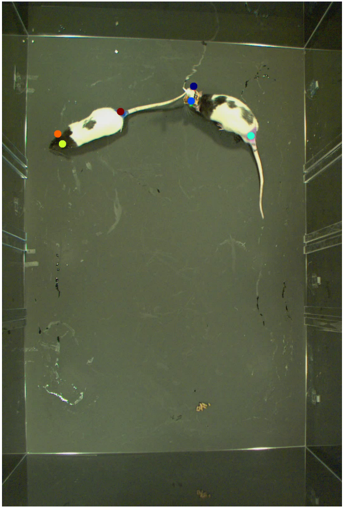


Behavioral syllables for isolated mice have been reported (2/2)





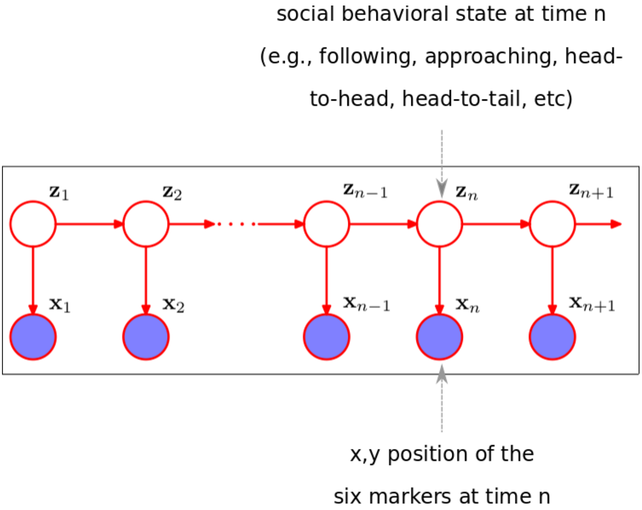
Discovering behavioral syllables in social interactions



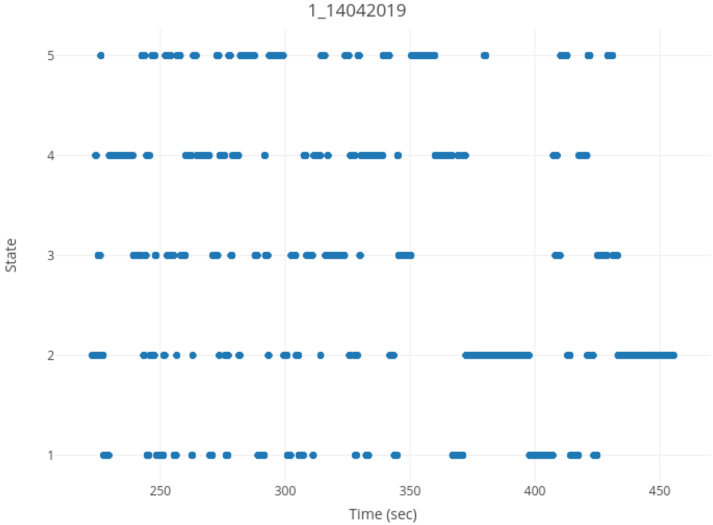
- approach
- head-to-tail
- broadly-social
- following
- no-contact
- conspecific-contact
- head-to-head
- head-to-flank

Cristina Mazuski (SWC)

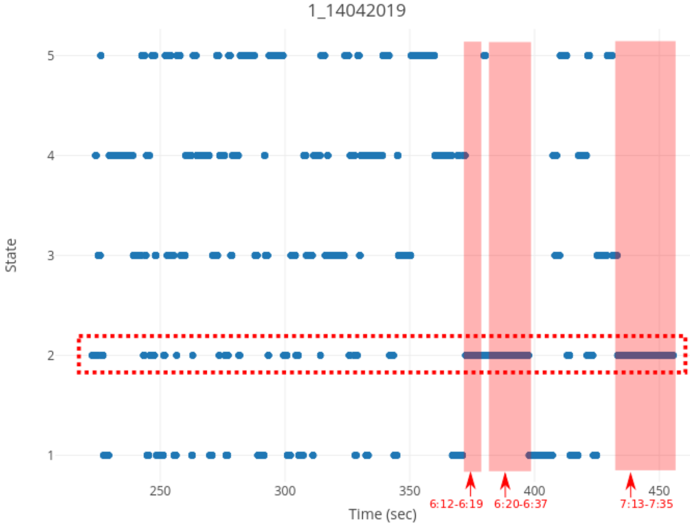
Hidden Markov Model for finding behavioral syllables in social interactions



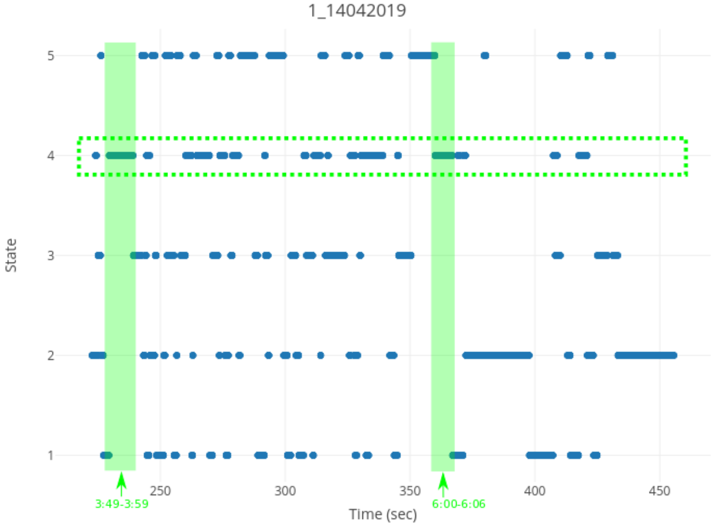
Behavioral syllables for social interactions



Non-Social state



Male-Head Female-Tail state

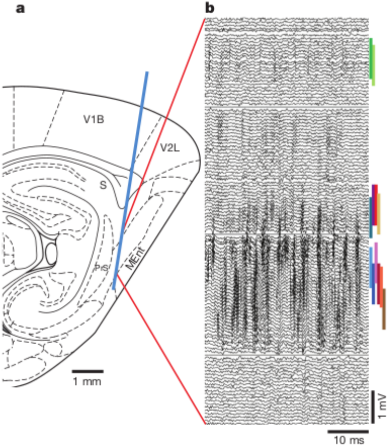


Partial Summary

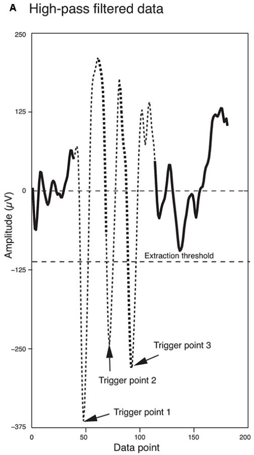
We gave the first steps in the quantitative characterization of social behavior in mice.

Table of Contents

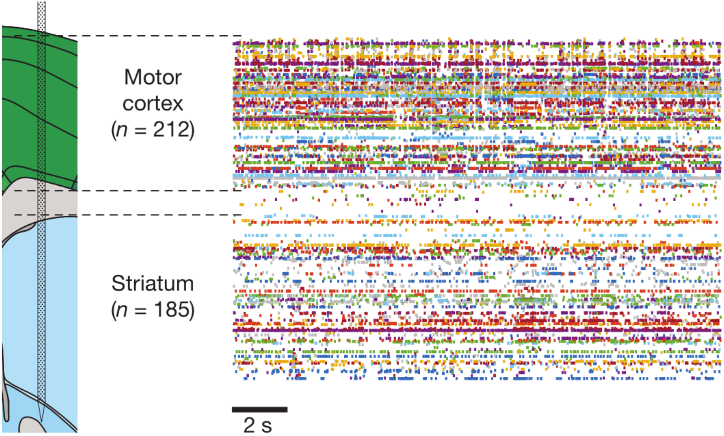
Raw data



Extraction of spikes

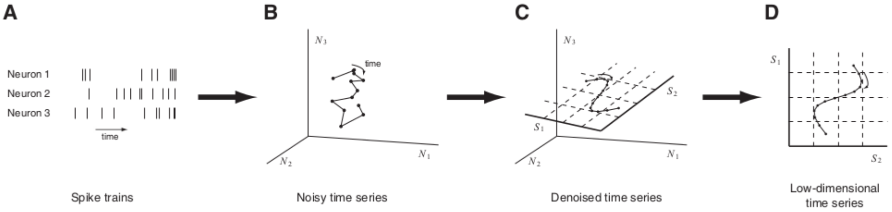


Spikes data



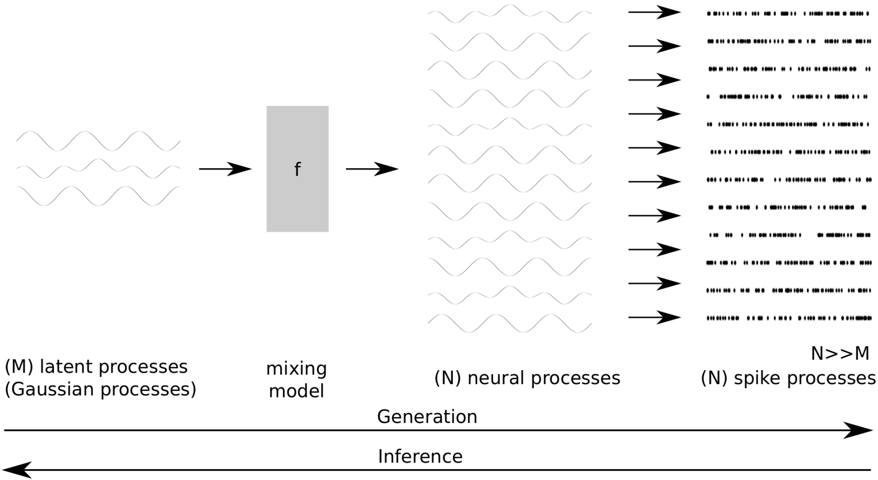
Our task now is to extract meaning from the spikes of populations of neurons.

Modeling neural populations spike recordings with Gaussian Processes Factor Analysis (intuition)



Yu et al., 2009

Modeling neural populations spike recordings with Gaussian Processes Factor Analysis (model sketch)



Modeling neural populations spike recordings with Gaussian Processes Factor Analysis (mathematical expressions)

$$x_k(t) \sim GP(\mu_k(t), k_k(t,t')),\ k = 1, \dots, K$$

$$y_n(t) = \sum_{k=1}^{K} C[n,k] x_k(t) + d[n] + v_n(t),\ v_n(t) \sim N(0, \sigma_n),\ n = 1, \dots N$$

Letting $X(t) = [x_1(t), \dots, x_K(t)],\ Y(t) = [y_1(t), \dots, y_N(t)]$, $d = [d[1], \dots, d[n]]$ and $R = diag(\sigma_1, \dots, \sigma_N)$ the model is:

$$Y(t) \sim N(C\,X(t) + d, R)$$

Parameters: $\theta = \left[C, d, \{\mu_k, \text{hyperparameters of } \{k_k(t,t')\}_{k=1}^{K}\}\right]$

Yu et al., 2009

Modeling neural populations spike recordings with Gaussian Processes Factor Analysis (estimation)

Estimation using an approximate variational method with inducing points.

Variational method: approximates the full posterior $p(\theta|y_1, \ldots, y_n)$.
Inducing points: allow to overcome the $O(N^3)$ complexity in Gaussian Process inference.

Duncker and Sahani, 2018

Goals of the distribution

- open source
- correct
- fast running time
- easy to modify/extend
- appealing to users
  - GUI
  - documentation
  - support/help infrastructure

Tools for the distribution

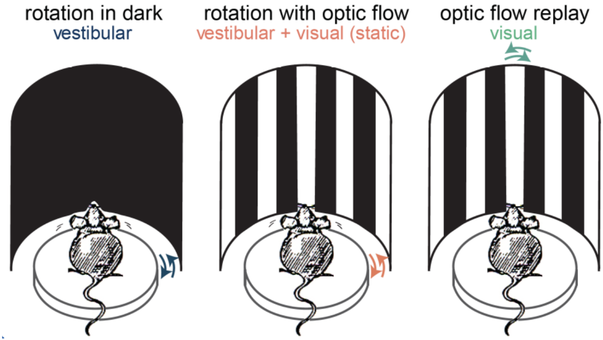
- open source – Github
- correct – Travis CI
- fast running time – GPyTorch
- easy to modify/extend – OO Design: class and interaction diagrams
- appealing to users
  - GUI – Qt
  - documentation – Sphinx/Read the Docs
  - support/help infrastructure – Github Issues, mailing list

Partial Summary

We made good progress in the distribution of the Python package Sparse Variational Gaussian Process Factor Analysis to model spiking activity in populations of neurons.

Table of Contents

Experiment



Sepiedeh Keshavarzi (SWC)

- Recordings from the same cell in the three conditions (visual, vestibular and visual+vestibular stimulation).
- Recordings from multiple brain regions, cortical layers and cell types (inhibitory and excitatory).
- Recordings from large populations of neurons (Neuropixels probe).

Scientific approach

Question: In what cortical areas, layers, and neuron types there is integration of visual and vestibular information?

Hypothesis: The response of neurons that integrate visual and vestibular information should be more correlated to stimuli speed in conditions with visual+vestibular stimulation than in conditions with only visual- or vestibular-only stimulation.

Methodological approach

Fit a state-space model to predict spiking activity of neurons using stimuli speed inputs. Find brain regions, layers and cell types for which predictions of models in the visual+vestibular stimulation condition is significantly larger than predictions of models in the visual- or vestibular-only condition.

Methodological approach

Fit a state-space model to predict spiking activity of neurons using stimuli speed inputs. Find brain regions, layers and cell types for which predictions of models in the visual+vestibular stimulation condition is significantly larger than predictions of models in the visual- or vestibular-only condition.

Requirements:

1. state-space model that can take inputs to generate predictions: Kalman filter estimated with the EM algorithm.
2. method to compare model predictions across cortical regions, cortical layers and cell types: analysis of variance (ANOVA).

Methodological approach

Fit a state-space model to predict spiking activity of neurons using stimuli speed inputs. Find brain regions, layers and cell types for which predictions of models in the visual+vestibular stimulation condition is significantly larger than predictions of models in the visual- or vestibular-only condition.

Requirements:

1. state-space model that can take inputs to generate predictions: Kalman filter estimated with the EM algorithm.
2. method to compare model predictions across cortical regions, cortical layers and cell types: analysis of variance (ANOVA).

Hypothesis: The response of neurons that integrate visual and vestibular information should be more correlated to stimuli speed in conditions with visual+vestibular stimulation than in conditions with visual- or vestibular-only stimulation.

Examples of Kalman filter results

The Kalman filter yields excellent predictions.
We found neurons that integrate visual and vestibular information.

Partial Summary

The Kalman filter appears to be a good model for large population recordings from the visual and vestibular integration experiment. In combination with the ANOVA, it may help us understand how multimodal information is integrated in the brain.

Table of Contents

The Data Center in Five Years

- Distributes high-quality code that allows scientists around the world to easily perform sophisticated neural data analysis and use advanced machine learning algorithms.

The Data Center in Five Years

- Distributes high-quality code that allows scientists around the world to easily perform sophisticated neural data analysis and use advanced machine learning algorithms.
- Educates neuroscientists worldwide on the use of advanced neural data analysis and machine learning methods.

The Data Center in Five Years

- Distributes high-quality code that allows scientists around the world to easily perform sophisticated neural data analysis and use advanced machine learning algorithms.
- Educates neuroscientists worldwide on the use of advanced neural data analysis and machine learning methods.
- Helps scientists from the SWC perform advanced statistical analysis of their neural recordings, and learns from them advanced experimental neuroscience concepts.

The Data Center in Five Years

- Distributes high-quality code that allows scientists around the world to easily perform sophisticated neural data analysis and use advanced machine learning algorithms.
- Educates neuroscientists worldwide on the use of advanced neural data analysis and machine learning methods.
- Helps scientists from the SWC perform advanced statistical analysis of their neural recordings, and learns from them advanced experimental neuroscience concepts.
- Plays a seminal role in key neuroscience findings from the SWC+Gatsby team-science research.

The Data Center in Five Years

- Distributes high-quality code that allows scientists around the world to easily perform sophisticated neural data analysis and use advanced machine learning algorithms.
- Educates neuroscientists worldwide on the use of advanced neural data analysis and machine learning methods.
- Helps scientists from the SWC perform advanced statistical analysis of their neural recordings, and learns from them advanced experimental neuroscience concepts.
- Plays a seminal role in key neuroscience findings from the SWC+Gatsby team-science research.
- Its members are highly skilled in neural data analysis and machine learning methods, and they are excellent software engineers.

The Data Center in Five Years

- Distributes high-quality code that allows scientists around the world to easily perform sophisticated neural data analysis and use advanced machine learning algorithms.
- Educates neuroscientists worldwide on the use of advanced neural data analysis and machine learning methods.
- Helps scientists from the SWC perform advanced statistical analysis of their neural recordings, and learns from them advanced experimental neuroscience concepts.
- Plays a seminal role in key neuroscience findings from the SWC+Gatsby team-science research.
- Its members are highly skilled in neural data analysis and machine learning methods, and they are excellent software engineers.
- The Gatsby+Data Center+SWC research demonstrates that the open-science approach produces excellent results.

Thanks

Cristina Mazuski

Sepiedeh Keshavarzi

Maneesh Sahani