Neuroinformatic Resources for Single- and Multi-neuron Spike Train Analysis

David H. Goldberg[1], Jonathan D. Victor[2], and Daniel Gardner[1,2]

[1]Lab of Neuroinformatics, Dept. of Physiology, [2]Dept. of Neurology & Neuroscience, Weill Cornell Med. Coll., NYC, NY

Enhanced Computational Methods for the Analysis of Neural Coding

Computational neuroinformatics synthesizes computational neuroscience—analyses of neural representation and information processing—and neuroinformatics—standards-based methods for archiving, classifying, and exchanging neuroscience data.

We are assembling a computational neuroinformatic resource that will aid the investigation of neural coding. This resource consists of neurodatabase.org, a web-based neurophysiology data archive, and a collection of information-theoretic spike train analysis tools. By providing experimenters with advanced spike train analysis techniques and theoreticians with broad collections of datasets, we seek to advance collaborations toward understanding neural codes.

Neural Codes and Information Theory

A profound question in neuroscience addresses neural coding: How do neurons use spike trains to represent information? Different neural systems are likely to use different representations. Therefore, examination of a broad collection of preparations is needed to elucidate the general principles of neural coding. The neural coding problem can be approached in a rigorous, quantitative manner. Techniques based in information theory are particularly well-suited to address neural coding because they can quantify the amount of information a neuronal response conveys about which of several discrete stimulus categories it belongs to.

Spike Train Analysis Toolkit

Toolkit Design and Implementation

The goal of this resource is to make information theoretic spike train analysis techniques available to a wide audience. The toolkit is available at http://cortex.med.cornell.edu/toolkit/

- Open-source, runs in Windows, Linux and Mac OS.
- Implemented in C, includes a Matlab interface.
- Uses a simple, platform-independent, human-readable format.

The Toolkit Implements Several Algorithms For Computing Mutual Information.

Not all methods are applicable to all data sets.

The applicability of a particular method to a specific data set depends upon:

- The amount of experimental data
- Assumptions about the topology of the response space
- Assumptions about the nature of the neural code

Different methods provide different insights.

The dependence of information on method parameters provides insight about the nature of the neural code.

Acknowledgements:

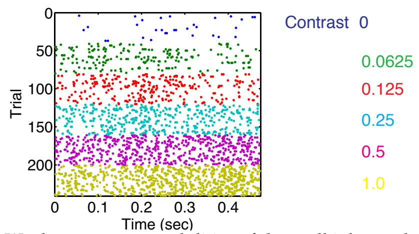
We acknowledge with thanks our collaborators, the originators of the information-theoretic methods we have implemented, and the lab of Esther P. Gardner for valuable cortical spike train datasets.

Human Brain Project / Neuroinformatics research in the Laboratory of Neuroinformatics is funded via MH68012 from NIMH, NINDS, NIBIB, NIA & NSF, and MH57153 from NIMH & NINDS.

LNI Laboratory of Neuroinformatics
Weill Medical College of Cornell University

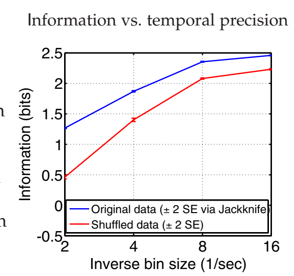
The Spike Train Analysis Toolkit Brings Together Different Information–Theoretic Methods to Address Neural Coding

Dataset: V1 Response to Drifting Gratings



We demonstrate capabilities of the toolkit by analysis of data from Reich *et al* (2001), *J. Neurophysiol.* Spike trains were recorded from a V1 neuron in an anesthetized macaque while drifting sinusoidal gratings differing in contrast were presented at the neuron's optimal orientation, spatial frequency, and temporal frequency.

1a. Direct Method (Strong et al., 1998)

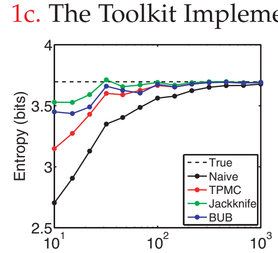


- Spike trains are divided into bins; the number of spikes in each bin is counted.
- Each bin count can be considered a *letter* in a spike train *word*. A histogram of word occurrences is generated and from this histogram entropy can be estimated:
- An experiment yields N words; word i occurs n_i times.
- Entropy is $-\Sigma_i\, p_i \log p_i$, estimated as $-\Sigma_i\, (n_i/N) \log (n_i/N)$
- The mutual information is computed from the difference between the entropy of the responses and the entropy of the responses conditioned on the stimulus.

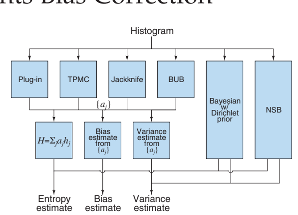
1b. Multineuron Direct Method

- Simultaneously recorded spike trains are assembled into *superwords*. Histograms of superword occurrences are tabulated.
- To reduce the data requirement of this method, the origin of the spikes can be ignored, suggesting a population rate code. Also, permuted spike trains could be considered identical.

1c. The Toolkit Implements Bias Correction



Bias correction performance of each of three methods on a set of Poisson spike trains of known entropy. Method performance depends upon data properties.

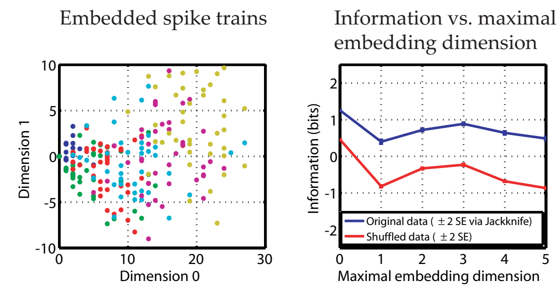


The toolkit will provide Bayesian estimates of the bias and variance for each estimator.

The toolkit includes several methods that correct for bias in the entropy estimate or relax the data requirement:

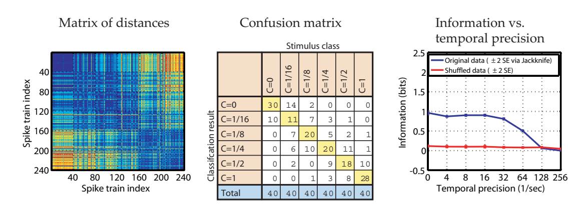
- Classical bias correction techniques: jackknife, Miller-Carlton (see Treves & Panzeri, 1995).
- More sophisticated entropy estimation techniques: best upper bounds (Paninski, 2003), NSB (Nemenman et al., 2002).

2. Binless Method (Victor, 2002)



- Spike trains are embedded in a Euclidean space using Legendre polynomials.
- Mutual information is determined by the extent of the separation of the clouds of points corresponding to each category.

3a. Metric Space Method (Victor & Purpura, 1997)



- Distances between pairs of spike trains are given by cost required to transform one spike train into another. The parameter q gives the cost of shifting a spike per unit time, relative to the cost of insertion/deletion.
- A clustering procedure finds a confusion matrix from the matrix of distances.
- Mutual information is determined by the extent of category-specific clustering in the distances:

3b. Multineuron Metric Space Method (Aronov et al., 2003)

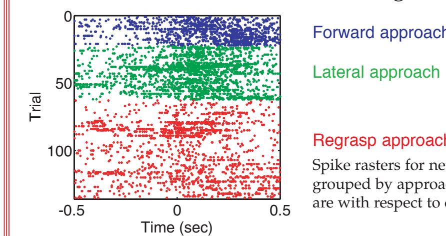
- Simultaneously recorded spike trains are interleaved into a single sequence of labeled events.
- The distance between (interleaved) spike trains is computed. In addition to insertion/deletion or shift costs, a label can be changed from one neuron to another at cost k, which describes the importance of the identity of the neuron that fires a spike.

4. Other Methods Under Development

- Power series (Panzeri & Schultz, 2001)
- Codebook quantization (Dimitrov & Miller, 2001)
- Information bottleneck (Tishby et al., 1999)
- Markov tree method (Kennel et al., 2005)
- Stimulus reconstruction method (Bialek et al., 1991)

The Toolkit Facilitates Analyses of Complex Neuronal Spike Train Datasets

Dataset: Parietal Neurons During Prehension



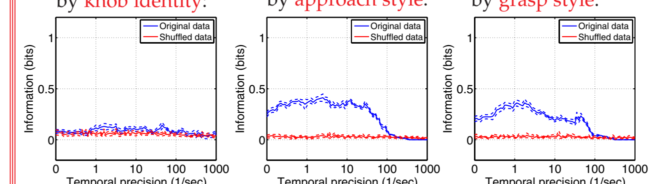
Spike rasters for neuron 70-3 are grouped by approach style; times are with respect to contact time.

To demonstrate the potential for our resource-catalyzed collaborations, we analyzed data with the Toolkit using only metadata information supplied with Neurodatabase.org datasets. Single- and multi-neuron spike trains from parietal cortex of macaque monkeys recorded during a reach/grasp prehension task (Debowy et al., 2002). Spike trains were labeled with approach style, grasp style, and knob identity and analyzed with the metric space method by Vaknin et al. (2005).

1. Metric Space Analysis Reveals Which Aspects of Task Kinematics Are Encoded by Specific Neurons in Parietal Cortex

1a. Neuron 70-3 Reports Approach & Grasp Styles:

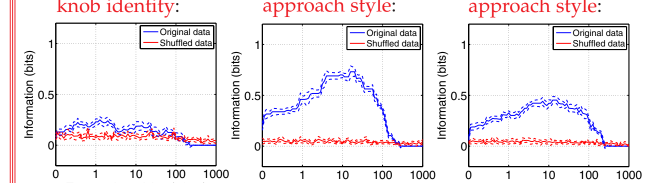
All trials, grouped by knob identity: | All trials, grouped by approach style: | All trials, grouped by grasp style:



- The neuron does not convey information about the knob identity.
- About 0.4 bits of information are conveyed for both approach style and grasp style.
- Information peaks at q=1 s^{-1} (corresponds to a timescale of 2 s).
- No significant difference between sighted and unsighted trials.

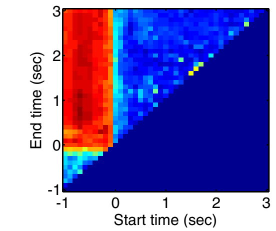
1b. Neuron 131-3.1 Reports Visually-Guided Approach:

Sighted trials, grouped by knob identity: | Sighted trials, grouped by approach style: | Blocked trials, grouped by approach style:



- The neuron does not convey information about the knob identity.
- Significantly more information is conveyed about approach style during sighted (0.7 bits) than during blocked condition (0.4 bits).
- There is a strong peak in information for both sighted and blocked at q=10 s^{-1} (corresponds to a timescale of 200 ms).
- A simultaneously-recorded neuron (131-3.2) displayed qualitatively similar characteristics.

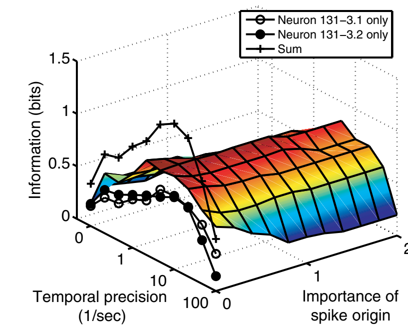
2. Toolkit Exploration of Parameter Space



Neuron 131-3.1, sighted trials, grouped by approach style. Information displayed as a function of analysis window start and end times. Red = 0.8 bit, blue = 0 bit.

The toolkit facilitates exploratory data analysis with comprehensive searches of the parameter space. This analysis reveals that most of the information about approach style is conveyed in a window as small as 100 ms just prior to contact.

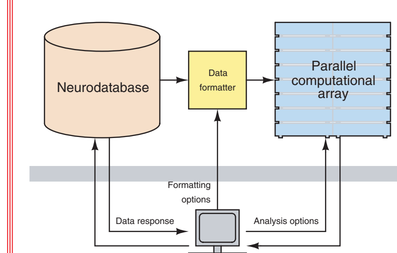
3. Multineuron Analysis Shows Redundancy



Neurons 131-3.1 and 131-3.2, grouped by approach style.

Information conveyed by pairs of nearby neurons is largely redundant, and keeping track of the source of each spike does not increase the amount of information conveyed.

Integrating Computational Resources into Neurodatabase.org



Extending the scope of the toolkit, we are also developing a web-accessible comprehensive resource for computational neuroinformatics. This resource will integrate the data repository and these analysis tools, using a graphical user interface, with:

- A dedicated parallel cluster for spike train analysis

For computationally intensive analyses, we are providing a Web-accessible 64-processor cluster for parallelized algorithms.

- A means for sharing of analytical techniques

Neurodatabase.org will archive and share analytical algorithms as well as data, facilitating exploration of neural signals.