Neuroanalysis.org: Information-Theoretic and Extended Analyses of Neural Coding

728-Pos

Michael A. Repucci[1], David H. Goldberg[1], Eliza Chan[1], Ajit Jagdale[1], Jonathan D. Victor[2] and Daniel Gardner[1,2]

[1]Lab of Neuroinformatics, Dept. of Physiology & Biophysics and [2]Dept. of Neurology & Neuroscience, Weill Cornell Med. Coll., New York, NY

Analyses of Neural Coding

New and Extended Algorithms are Needed to Elicit Principles of Neural Coding

Neural coding—the representation and processing of information with spike trains—is fundamental to sensation, perception, decision, and action. Our ongoing project is designed to aid new analyses of the significance of experimentally-measured single and multineuron recordings of brain activity, including information-theoretic measures, that complement the many widely-available time-domain and frequency-domain analyses informing contemporary systems neuroscience.

Elucidating principles of neural coding requires multiple analytic methods. This is because specific neural systems are likely to use different codes, or representations. Also, neurophysiology data including the activity patterns of multiple neurons over time are intrinsically complex, and different methods for analyzing such datasets present both conceptual and practical challenges including requirements for specific types and amounts of data.

To facilitate adoption of new and significant methods for the analysis of neural coding, we here report new features, continued adoption, and planned expansions of the Spike Train Analysis Toolkit (STAToolkit) and allied resources. STAToolkit, now version 1.5, is available free, open access, and open source at: http://neuroanalysis.org

For laboratories that record neural data but lack computational resources to fully explore such data, our project offers not only ready-to-use software, but also a 64-processor array for analyses.

Neural Codes, Information Theory, and Beyond…

Shannon's 1948 work in information and communication theory is an appropriate framework to address neural coding and classification, as it informs metrics that quantify the amount of information a neuronal response conveys about which of several stimuli, behaviors, or disease categories it relates to. However, application of information-theoretic principles to experimental neural data is far from straightforward.

Our current methods have addressed the question:
- What information is present in patterns of neural activity?

We now expand this to add such related questions as:
- What observable aspects or features (e.g., firing rate, firing intervals, synchrony) of the activity carry this information?
- What features of synchrony or variability signal neural computation?

and
- In what sense do these features represent or map the space of events, percepts, and actions?

In this work, we continue to develop the new field of computational neuroinformatics, which synthesizes computational neuroscience—analyses of neural representation and information processing—with neuroinformatics, standards-based methods for archiving, classifying, and exchanging neuroscience data.

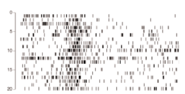
LNI Laboratory of Neuroinformatics
Weill Medical College of Cornell University

Spike Train Analysis Toolkit V1.5, Available Open Source at Neuroanalysis.org

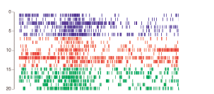
The downloadable open source STAToolkit suite of information-theoretic algorithms implements, documents, and guides application of analytic methods, thus minimizing the effort needed to adopt and use them (Goldberg et al. *Neuroinformatics* 7, 165–178, 2009).

STAToolkit implements multiple information methods:
- Direct (formal and category-specific; Strong et al 1998),
- Metric space (category-specific; Aronov 2003, Victor & Purpura 1997),
- Binless (category-specific; Victor 2002), and
- Context tree Markov chain Monte Carlo (formal; Kennel et al 2005, Shlens et al 2007).

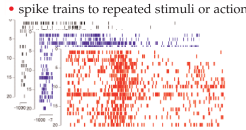
Version 1.5 (released 12 February 2010) includes:
- C/C++ and Matlab code for calculating information, as well as eight methods for estimating entropy and three for entropy variance,
- Modular design, versioned and commented,
- Scripts for Windows, Linux, and MacOS installation,
- Sample data sets, didactic demonstrations, and usage examples,
- Extensive online documentation for installation and use,
- Compatibility with Octave (version 3.0.1 or greater), and
- Continuous data (e.g., LFP) capable binless information analysis.



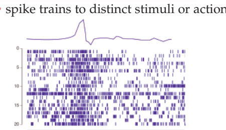
- spike trains to repeated stimuli or actions



- spike trains to distinct stimuli or actions



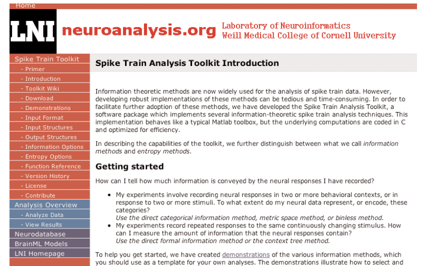
- spike trains from multiple units or areas



- LFPs and spike trains

1. STAToolkit is Designed for Neurophysiologists Recording Current, Informative Data & Protocols

Methods are designed for neurophysiologists recording single- and multi-unit neuronal activity and interactions with population firing, and relating neural data to sensation and motor behavior.

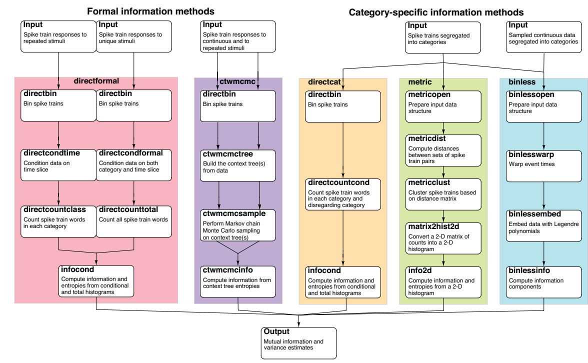


2. STAToolkit Includes Tutorial Documentation

To aid neurophysiologist users, neuroanalysis.org provides fast-track Getting Started (shown) and Demonstrations that use the amount and kind of data available to select appropriate methods, a Primer: What will STAToolkit tell me about my data?, and a collaborative wiki.

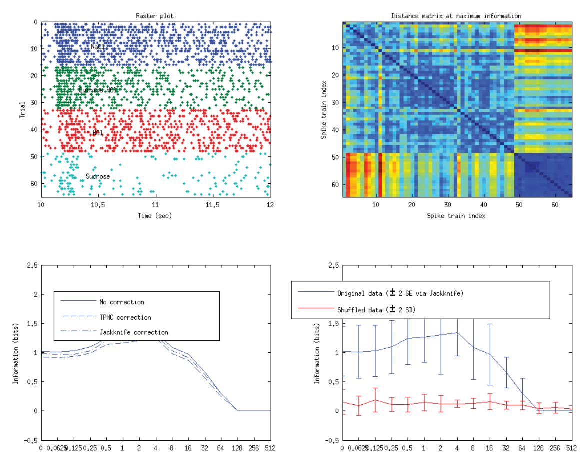
STAToolkit and Neuroanalysis.org: Scope & Usage

STAToolkit is in active use by the neuroscience community, with more than 1,200 downloads applied to many preparations and protocols: retinal coding, prehension in awake behaving primates, mammalian taste discrimination, spike timing in thalamocortical and other networks, EOD analysis in electric fish, and various population codes. STAToolkit has been used in courses and is being adopted by leading neuroinformatic resources in the US, the UK, and Canada.



3. Modular Implementation of Multi-Track Information Calculation Functionality

STAToolkit provides multi-track functionality via five information-calculation methods. Each is easily selectable by a top-level function applied to an input data structure: directformal (formal information analysis via direct method), ctwmcmc (formal via context-tree MCMC), directcat (category-specific analysis via direct method), metric (category-specific via metric space), binless (category-specific via binless method). Each is partitioned into modules providing useful optional intermediate results; a full function list is at http://neuroanalysis.org/neuroanalysis/function/index.html



4. STAToolkit Demos Graphically Verify Compilation, Installation, And Application

Demos include this figure analyzing sample data (Di Lorenzo and Victor 2003) via STAToolkit's metric method. The lower right panel plots information as a function of temporal precision (blue); information for shuffled data (red) serves as a control verifying sufficient data to carry out a valid analysis. The other panels show original rasters, intermediate results, and sample size corrections.

Other demos show how to select and use information and entropy methods: two scripts demonstrate calculation of formal information; three quantify the amount of information that an ensemble of spike trains conveys about a stimulus that can be characterized by discrete categories; two demonstrate all included entropy methods.

Extensions Planned Towards STAToolkit V2.0

Method	Algorithm designer	Current form	Spike train suitability	Multineuron suitability	LFP suitability
Hybrid	I Nelken	Fortran, C code	✓	**	*
Koepsell-Sommer method	F Sommer	Published algorithm	✓	*	***
Anthropic correction	F Theunissen	Published algorithm	✓	***	**
Generalized Information	S Nirenberg, J Victor	Published algorithm, code	✓	***	**
Maximum-entropy models	J Shlens	Published algorithms, code	✓	***	
Multineuronal metrics	C Houghton, M Magnasco, JHG Dauwels	Published algorithm, code	✓	***	
Surrogate data generation	M Bethge, R Brette, S Shoham, M Krumin	Published algorithm, code	✓	***	**
Stochastic event synchrony	JHG Dauwels	Published algorithm, Matlab code	✓	***	***
Event synchrony	R Quian-Quiroga, T Kreuz	Published algorithm, code	✓	***	*
Biophysical variability	B Knight	Published algorithm, code	✓		

5. A Collaborative Model Extends STAToolkit

Planned extensions to STAToolkit and neurodatabase.org include:
- Enhanced information-theoretic analyses
  - New single-neuron entropy and information methods
  - Analyses of bursting, including alternatives to Poisson
  - Information contributed by oscillatory neurons
- Dimensional reduction
  - Maximum-entropy models of multineuron recordings
  - New measures of similarity of neuronal populations
- Neuronal population variability and synchrony
  - Identifying common and individual activity
  - Distinguishing information from biophysical variation

Beyond new information and entropy methods, we add analyses of multineuronal firing patterns and algorithms for dimensional reduction—essential for analyses of complex patterns of neural activity, exploratory data evaluation, state space reconstruction, and visualization. Other approaches assess synchrony and variability (mathematically and biologically related). Synchrony can inform how a single underlying stimulus, transformed in different ways and recorded at separate locations, is bound to form a percept or plan an action. Variability of responses to repeated stimuli can distinguish informational from random or purely biophysical variation. We will generate surrogate multineuronal datasets with statistical parameters conforming to specific hypotheses.

All methods are suitable for spike trains; asterisks show degree of applicability to multineuron or LFP data.

In a unique collaboration, 13 leaders in computational neuroscience have made available their algorithmic equations, code, and insight, enabling us to develop, standardize, and add to neuroanalysis.org this wide set of recent innovative advances for neural data analysis:
- Matthias Bethge, MPI für biologische Kybernetik
- Romain Brette, École Normale Supérieure
- Justin H.G. Dauwels, MIT
- Conor J. Houghton, Trinity College Dublin
- Bruce Knight, Rockefeller University
- Thomas Kreuz, UC San Diego
- Marcelo O. Magnasco, Rockefeller University
- Israel Nelken, Hebrew University Jerusalem
- Sheila Nirenberg, Weill Cornell
- Rodrigo Quian Quiroga, University of Leicester
- Jonathon Shlens, UC Berkeley and Salk Institute
- Shy Shoham, Technion—Israel Institute of Technology
- Friedrich T. Sommer, Redwood Center; UC Berkeley

New Analyses Enabled by Our Collaborative Model:

Each of these new and in-development methods is designed to answer one or more questions important for understanding neuronal firing, including:
- What neural codes are plausibly represented by spike activity?
- How variable are spike trains to repeated, or a range of, stimuli?
- What aspects of the input LFP influence output firing?
- Are Poisson or Markov models plausible for generating activity?
- Do multineuronal patterns play a role in signaling?
- Do connectivity or common input plausibly account for ensembles?
- How are simultaneously-recorded trains similar in temporal structure? What's common across regions? Is variability common?
- Can biophysics alone characterize variability to repeated stimuli?
- How do data differ from the predictions of a model? Which measures show which models are best?

Community Development Further Extends STAToolkit

Beyond the Laboratory of Neuroinformatics and our collaborators, members of the computational neuroscience community contribute:
- The CARMEN central data storage, pipelining, and analysis project is making STAToolkit available to its many users across the UK.
- SigTOOL, providing GUI-driven MATLAB-centric spike train data input and analysis, is collaborating on adding STAToolkit capability and combining interface development.
- Users have ported STAToolkit to Python and Octave resources.

Conclusions and Projections

These new analytical algorithms, enabling neurophysiologists to test new hypotheses about neural coding, add value to data from model animals—including valuable mammalian species—and advance experimental design by enabling protocols to test new hypotheses.

Tool-enabled examination of neural coding in many preparations will enable testing of five overarching meta-hypotheses:
- Different regions, networks, or modalities within the nervous system utilize different neural codes.
- Individual regions, networks, or modalities utilize different coding at different times, or in different contexts.
- Unified sensory percepts are 'bound' via simultaneous-coded signals in disparate areas.
- Central variability may aid plasticity, sensitivity, or the representation of uncertainty.
- Mechanisms of neural or mental disorders may be better understood by their effects on neural coding.

Acknowledgements

We thank the consultant-collaborators listed at left who have graciously provided algorithms, code, and insight, and the many labs who have provided datasets to neurodatabase.org, as well as the members of our Neuroanalysis Advisory Group:
- Patricia M. Di Lorenzo
- Esther P. Gardner
- Michael E. Goldberg
- Steven S. Hsiao
- Roger Lemon
- Sheila Nirenberg
- Bijan Pesaran
- Charles E. Schroeder

We gratefully acknowledge the support of MH68012 from NIMH, NINDS, NIA, NIBIB, and NSF, and MH57153 from NIMH and NINDS, to D.G. via the Human Brain Project/Neuroinformatics.