From *SIAM News*, Volume 40, Number 2, March 2007

Signal Processing and Statistical Challenges in Neuroscience Data Analysis

By Emery N. Brown

Neuroscience, like many other areas of science, is experiencing a data explosion, propelled both by improvements in existing recording technologies, such as electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI), and by the advent of new and relatively new measurement and recording technologies, such as multiple electrode arrays, diffuse optical tomography (DOT), and diffusion tensor imaging (DTI). These technologies have had a significant impact on basic and clinical neuroscience research. An analysis bottleneck is inevitable as the collection of data using these techniques now outpaces the development of new methods appropriate for analysis of the data. In each case, resolution of significant signal processing and statistical questions is crucial if researchers and clinicians are to make the best inferences from the data. This article briefly describes three challenging data analysis questions in neuroscience research.

Spike Sorting: Identification and Classification of Spike Events

Multiple-electrode arrays capable of recording the simultaneous spiking activity of many (more than 20) neurons make it possible to study how groups of neurons act in concert to define the function of a given region of the brain [14]. Sorting of extracellular spike events is an obligatory first step in analyzing multiple-electrode recordings, because the potentials measured from multiple electrodes are the spiking and dendritic voltages of many simultaneously recorded neurons [8]. The extraction or sorting of the spike events from these signals, usually an off-line process, is accomplished in three steps: spike-event detection, identification of the number of neurons, and spike-event classification [1].

In the first step, a combination of filtering algorithms and thresholding criteria is used within a specified time window to detect spike events in the extracellular potentials. Each spike event is summarized as a feature vector, which is a sample of the voltage trace in a short time interval around the event. In the classification step, the characteristics of the feature vector are used to assign each spike event to a specific neuron. This can be done manually, by inspecting plots of low-dimensional projections of the feature vectors, or by applying automated spike-classification techniques that use one of several clustering algorithms from machine learning or statistics to place similar feature vectors in the same group. All spike events of one group are taken as coming from the same neuron.

There is no consensus on the optimal approach to spike sorting. Because of the complexity of the spike-sorting problem, different methods applied to the same data set can yield different results. Several factors contribute to the complexity. First, clusters often violate the frequently made assumption of Gaussian errors in model-based parametric analyses; the clusters can also change in time as a result of drifts in experimental conditions [3]. Second, identifying the number of neurons to which the spike events will be assigned is a challenge [11]. Third, simultaneous intracellular–extracellular recordings have shown that spike sorting for large numbers of neurons can have a non-zero error rate because the spike shapes from different neurons can overlap [6]. Fourth, multiple-electrode arrays with different geometries and numbers of electrodes require different sorting algorithms. Finally, in neural prosthetic and brain–machine interface applications, spike sorting has to be done on-line and in real time if the devices are to be practical [7]. Defining optimal strategies for spike sorting is an important question for neuroscience research: All subsequent analyses and conclusions from multiple-electrode recordings depend on the solution obtained in this first step.

Multimodal Data Fusion

A central focus of current functional neuroimaging research is the development of tools for multimodal image fusion, for varying combinations of fMRI, DOT, EEG, and MEG. This means conducting experiments in which imaging is performed in two or more of these modalities, simultaneously or in sequence, so that the information from the different sources can be combined [9]. Use of two or more imaging modalities simultaneously to study brain function provides information about the same sources, yet on different spatial and time scales. The combination of EEG measurements with simultaneously recorded fMRI, for example, takes advantage of both the high spatial resolution of fMRI and the high temporal resolution of EEG. Other combinations being actively explored include EEG with MEG, fMRI with DOT, and DOT with both EEG and MEG.

Optimal fusion of the information from different modalities is a nontrivial task, requiring computed solutions to a sequence of high-dimensional, ill-posed inverse problems. Solving an ill-posed inverse problem means determining the best estimates of activity in specific brain regions from the multimodal measurements [4,5,10]. The problem is ill-posed because without additional constraints the solution is not unique. The constraints imposed should be chosen not arbitrarily but rather on the basis of known biophysical properties of the imaging modalities and the physiological and anatomical features of the brain and the particular experimental question under study [12,13]. How to do this is an open problem whose solution has enormous implications for the optimal use of functional neuroimaging technology. Moreover, the concept, if not the specific techniques, can be used to combine neurophysiological information from other important sources, such as behavioral recordings and genomic data.

Dynamic Signal Processing Methods

Much neuroscience research focuses on understanding how the dynamic properties of synapses, individual neurons, and neural circuits lead to the functional properties of specific brain regions. Proper estimation of the dynamic properties of neural systems, whether from imaging, local

field potentials, or neural spike-train recordings, requires dynamic signal processing methods. Despite the wide array of dynamic signal processing algorithms available for continuous-valued measurements, most current methods for neural spike-train data analysis are static rather than dynamic. Hence, there is a significant need for dynamic statistical methods explicitly designed to analyze neural spike-train data recorded from multiple-electrode arrays [1].

Dynamical systems and neural network models have long provided essential quantitative characterizations of neural systems [2]. Research on dynamic data-analysis methodology, however, should be conducted more in concert with modeling research. With experiments directly related to computational modeling, experimental findings could be put to improved use: in quantifying better predictions from more complex models, in refining model formulation, and, as a consequence, in designing better experiments. Similarly, computational models could suggest formulations of the statistical methods that would make them more successful in extracting salient features from experimental data.

As illustrated by these three examples, solutions to key signal processing and statistics problems in neuroscience revolve around the need for appropriate new methods that can correctly characterize the dynamic, multivariate nature of neuroscience data. In some cases, it will be possible to borrow from existing paradigms in statistics, engineering, computer science, and physics to design these methods. Indeed, use of existing methods should be the first step. The unique features of neuroscience experiments, however, are leading rapidly to a need for signal processing and statistical methods developed from scratch specifically for these studies.

Acknowledgments

Support for this work was provided in part by NIH grants MH66410, MH59733, EB00522, and DA015664.

References

[1] E.N. Brown, P.P. Mitra, and R.E. Kass, *Multiple neural spike train data analysis: State-of-the-art and future challenges*, Nat. Neurosci., 7:5 (2004), 45661.

[2] P. Dayan and L.F. Abbott, *Theoretical Neuroscience*, MIT Press, Cambridge, Massachusetts, 2001.

[3] M.S. Fee, P.P. Mitra, and D. Kleinfeld, *Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability*, J. Neurosci. Meth., 69 (1996), 175–188.

[4] A. Galka, O. Yamashita, T. Ozaki, R. Biscay, and P. Valdes-Sosa, *A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering*, NeuroImage, 23 (2004), 435–453.

[5] M.S. Hamalainen, R. Hari, R.J. Ilmoniemi, J. Knuutila, and O.V. Lounasmaa, *Magnetoencephalography—Theory, instrumentation, and application to noninvasive studies of the working human brain*, Rev. Modern Phys., 65 (1993), 413–497.

[6] K.D. Harris, D.A. Henze, J. Csicsvari, H. Hirase, and G. Buzsaki, *Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements*, J. Neurophysiol., 84 (2000), 401–414.

[7] L.R. Hochberg, M.D. Serruya, G.M. Friehs, J.A. Mukand, M. Saleh, A.H. Caplan, A. Branner, D. Chen, R.D. Penn, and J.P. Donoghue, *Neuronal ensemble control of prosthetic devices by a human with tetraplegia*, Nature, 442 (2006), 164–171.

[8] M.S. Lewicki, *A review of methods for spike sorting: The detection and classification of neural action potentials*, Comput. Neural Syst., 9 (1998), R53–R78.

[9] A.K. Liu, J.W. Belliveau, and A. Dale, *Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations*, Proc. Natl. Acad. Sci, 95:15 (1998), 8945–8950.

[10] C.L. Long, N. Desai, S. Temereanca, P. Purdon, M.S. Hämäläinen, and E.N. Brown, *A dynamic solution to the ill-conditioned magnetoencephalography (MEG) source localization problem*, Third IEEE International Symposium on Biomedical Imaging, 2006.

[11] D. Nguyen, L.M. Frank, and E.N. Brown, *An application of reversible-jump MCMC to spike classification of multiunit extracellular recordings*, Network: Comput. Neurosci., 14 (2003), 61–82.

[12] J.J. Riera, X. Wan, J.C. Jimenez, and R. Kawashima, *Nonlinear local electrovascular coupling, I: A theoretical model*, Hum. Brain Mapp., 11 (2006), 896–914.

[13] J.J. Riera, X. Wan, J.C. Jimenez, and R. Kawashima, *Nonlinear local electrovascular coupling, II: From data to neuronal masses*, Hum. Brain Mapp.; E-pub, August 24, 2006.

[14] M.A. Wilson and B.L. McNaughton, *Dynamics of the hippocampal ensemble code for space*, Science, 261 (1993), 1055–1058.

Emery Brown is a member of the Department of Brain and Cognitive Sciences at MIT and the Department of Anesthesia and Critical Care, at Massachusetts General Hospital and Harvard Medical School.