Finding the invariant feature space of a neural computation

Adrienne Fairhall[1] and Blaise Agüera y Arcas[2]
[1]Dept. of Molecular Biology and [2]Dept. of Applied Mathematics
Princeton University, Princeton, NJ 08544
{*fairhall,blaisea*}*@princeton.edu*

February 24, 2003

Abstract

The spike-triggered average is a widely used experimental tool. We consider an important artifact of the spike-triggered average: a generic dependence on the variance of the white noise stimulus. Several biological systems share this property, and this has been interpreted as enhancing the system's information processing capabilities. We disentangle the generic part of this variance dependence from potentially more interesting biological adaptation, using as a model the integrate and fire neuron, for which the unique relevant filter is known. We apply novel methods to recover the known variance-independent filter. Our methods apply without assumptions to experimental data.

The computation performed by a neuron can be formulated as the projection of the stimulus onto a low-dimensional subspace– dimensional reduction or feature selection– followed by a nonlinear decision function on that projection. This model of neural function embodies the intuition that the system should be sensitive to only a small subset of the many possible stimulus features it receives. The same intuition underlies a description in terms of receptive fields or tuning curves. One of the major experimental challenges of neuroscience is to determine the relevant feature subspace that triggers neural activity. To date, the best solution to this problem has been reverse correlation [1], where one stimulates with white noise and computes the average stimulus history preceding a spike. More recent work has advanced methods of recovering multiple dimensions in stimulus space using the covariance matrix of the stimulus history [2, 3].

It has often been observed experimentally that the feature (or features) obtained using reverse correlation depend on the variance of the white noise

stimulus. It is very likely that biological systems rely on such adaptation to tune the relevant feature encoded to the local statistics of the environment. In the LMC cells of the fly, the details of this feature adaptation match what one would predict by optimising information transmission through the system [4]. However, it is pertinent to ask to what extent variance dependence is a property that is truly tuned by the system, and to what extent it is a necessary artifact of the methods of reverse correlation. Is it possible to recover from data a genuinely stimulus independent feature?

Recent work [5, 6, 3, 7] has pointed out that serial correlations among spikes render the spike-triggered average an inaccurate measure of the "true" feature triggering spikes. This is due simply to the fact that the probability of generating a spike depends both on the stimulus and on the presence of previous spikes. (Note that interspike interaction is also responsible for an apparent "adaptation" of the decision function itself [8]; the distribution of interspike intervals is also a function of the stimulus variance.) Further, as we will discuss, the spike-triggered average is inevitably contaminated by the fact that the decision function producing spiking acts sequentially in time. In [5], a method was introduced to overcome the problem of interspike interaction– retaining only "isolated" spikes, those spikes whose firing was well separated from previous spikes so that their generation is purely stimulus, not prior activity, dependent. Developments of this work, outlined partly in [6], show that this allows one to recover the true exponential filter of the leaky integrate and fire neuron, in contrast to simple reverse correlation which produces a highly peaked function. The price that one pays for this is the explicit requirement of extended silences before each spike, which renders the prior stimulus ensemble nonGaussian, leading to apparent "silence contributions" to the relevant feature space. It was shown that these "silence" modes can be identified and separated from features which are *causally* linked to spike generation.

Stated more generally, the spiking of a neuron depends *causally* on some event in the stimulus, and *non-causally* on certain constraints imposed by the fact that spikes are produced sequentially. By isolating spikes, those constraints include the lack of previous spikes for some time in the past. We show how to use second order methods to recover features which are both causal and non-causal. The non-causal (or "silence") modes emerge as an orderly sequence of Fourier-like components which we are able to identify and decouple from the causal features. Note that both causal and non-causal features may be relevant for the stimulus reconstruction problem; however, for predicting a spike train from input data only the causal features are relevant.

The form of the silence modes is necessarily variance dependent. If one considers the time-dependent distribution of allowed inputs preceding a spike, the silence modes describe the deviation of this distribution from the prior. The timescale for the evolution of that distribution is governed by the stimulus variance. The spike-triggered average for isolated spikes (Fig.1) lies in the space spanned by both silence-related (non-causal) and spike-related (causal) features. The variance dependence of the non-causal part leads to a counter-intuitive *widening* of the apparent filter with increasing variance. This occurs as the time preceding a spike in which the input must be constrained increases with stimulus variance (Fig.2).

We have presented a very general method for recovering the invariant relevant feature space of a neural computation from experimental data. Our method avoids artifacts embodied in the spike generation process that contaminate the usual spike-triggered average. We make no assumptions about the dimensionality of the relevant feature space or the form of the spike interaction (e.g. the reset following a spike). Thus our approach applies even in the most general case of multiple causal features and a nontrivial interspike interaction.

References

[1] E. de Boer and P. Kuyper. Triggered correlation. *IEEE Trans. Biomed. Eng.*, 15:169–179, 1968.

[2] W. Bialek and R. R. de Ruyter van Steveninck. Features and dimensions: motion estimation in fly vision. *in preparation*, 2002.

[3] B. Agüera y Arcas, A. L. Fairhall, and W. Bialek. Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Computation, in press*, 2003.

[4] J. H. van Hateren. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *J. Comp. Physiol. A*, 171:157–170, 1992.

[5] B. Agüera y Arcas, W. Bialek, and A. L. Fairhall. What can a single neuron compute? In T.K. Leen, T.G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems 13*, pages 75–81. MIT Press, 2001.

[6] B. Agüera y Arcas and A.L. Fairhall. What causes a neuron to spike? *Neural Computation, in press*, 2003.

[7] J. Pillow and E. P. Simoncelli. Biases in white noise analysis due to non-poisson spike generation. *Neurocomputing, in press*, 2003.

[8] L. Paninski, B. Lau, and A. Reyes. Noise–driven adaptation: *in vitro* and mathematical analysis. *Neurocomputing, in press*, 2003.

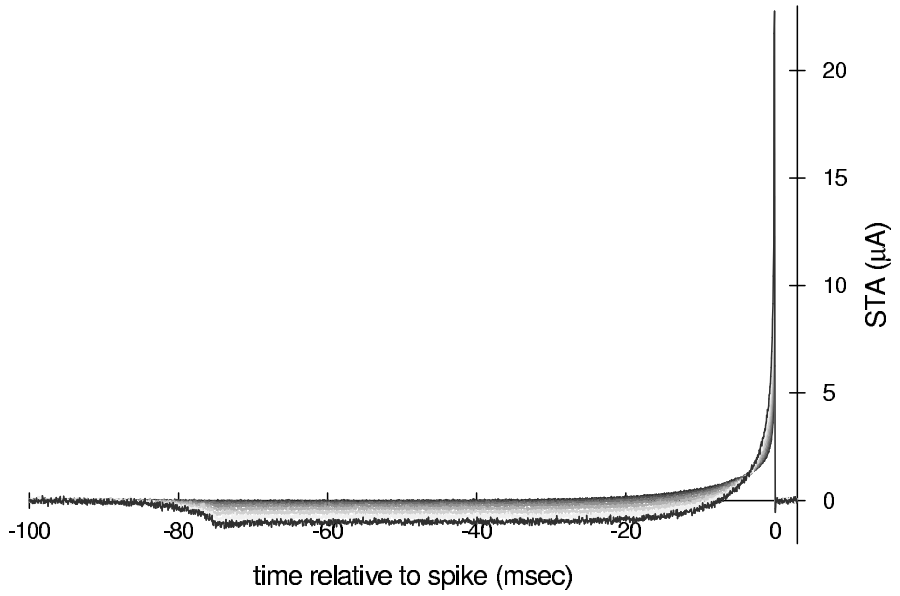

Figure 1: Isolated spike-triggered averages from the leaky integrate and fire neuron for a range of stimulus variances. Note the broadening of the filter for larger variances.

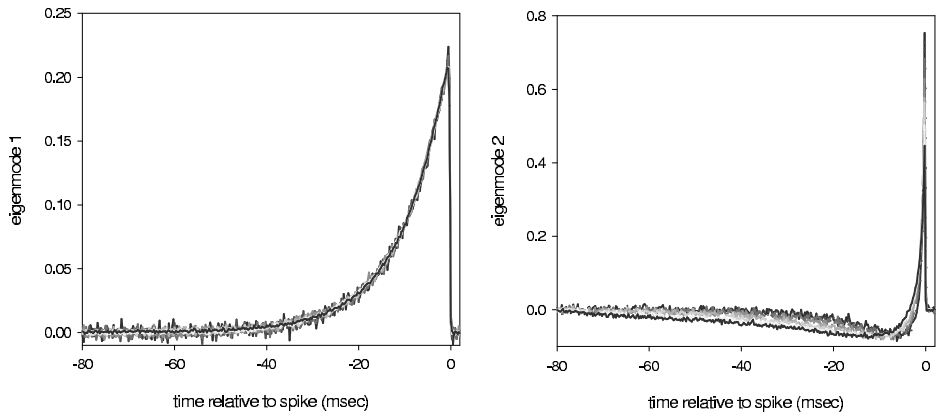

Figure 2: Left, the leading mode of the isolated-spike second-order analysis for 8 values of the stimulus variance, ranging over an order of magnitude. We recover each time the unique exponential filter of the leaky integrate and fire neuron. Right, the leading silence mode, showing variance dependence.