Maximally informative dimensions: Analyzing neural responses to natural signals.

Tatyana Sharpee[1]**, Nicole C. Rust**[2]**, and William Bialek**[1,3]
[1] Sloan–Swartz Center for Theoretical Neurobiology, Department of Physiology
University of California at San Francisco, San Francisco, California 94143–0444
[2] Center for Neural Science, New York University, New York, NY 10003
[3] Department of Physics, Princeton University, Princeton, New Jersey 08544
sharpee@phy.ucsf.edu, rust@cns.nyu.edu, wbialek@princeton.edu

We propose a method that allows for a rigorous statistical analysis of neural responses to natural stimuli, which are non-Gaussian and exhibit strong correlations. We have in mind a model in which neurons are selective for a small number of stimulus dimensions out of the high dimensional stimulus space, but within this subspace the responses can be arbitrarily nonlinear. Therefore we maximize the mutual information between the sequence of elicited neural responses and an ensemble of stimuli that has been projected on trial directions in the stimulus space. The procedure can be done iteratively by increasing the number of directions with respect to which information is maximized. Those directions that allow the recovery of all of the information between spikes and the full unprojected stimuli describe the relevant subspace. If the dimensionality of the relevant subspace indeed is much smaller than that of the overall stimulus space, it may become experimentally feasible to map out the neuron's input-output function even under fully natural stimulus conditions. This contrasts with methods based on correlations functions (reverse correlation, spike-triggered covariance, ...) which all require simplified stimulus statistics if we are to use them rigorously.

1   Introduction

From olfaction to vision and audition, there is an increasing need, and a growing number of experiments [1]-[8] that study responses of sensory neurons to natural stimuli. Natural stimuli have specific statistical properties [9, 10], and therefore sample only a subspace of all possible spatial and temporal frequencies explored during stimulation with white noise. Observing the full dynamic range of neural responses may require using stimulus ensembles which approximate those occurring in nature, and it is an attractive hypothesis that the neural representation of these natural signals may be optimized in some way. Finally, some neuron responses are strongly nonlinear and adaptive, and may not be predicted from a combination of responses to simple stimuli. It has also been shown that the variability in neural response decreases substantially when dynamical, rather than static, stimuli are used [11, 12]. For all these reasons, it would be attractive to have a rigorous method of analyzing neural responses to complex, naturalistic inputs.

The stimuli analyzed by sensory neurons are intrinsically high-dimensional, with dimen-

sions $d \sim 10^2 - 10^3$. For example, in the case of visual neurons, input is specified as light intensity on a grid of at least 10×10 pixels. The dimensionality increases further if the time dependence is to be explored as well. Full exploration of such a large parameter space is beyond the constraints of experimental data collection. However, progress can be made provided we make certain assumptions about how the response has been generated. In the simplest model, the probability of response can be described by one receptive field (RF) [13]. The receptive field can be thought of as a special direction $\hat{v}$ in the stimulus space such that the neuron's response depends only on a projection of a given stimulus $\mathbf{s}$ onto $\hat{v}$. This special direction $\hat{v}$ is the one found by the reverse correlation method [13, 14]. In a more general case, the probability of the response depends on projections $s^{(\mu)} = \hat{e}^{(\mu)} \cdot \mathbf{s}$, $\mu = 1, ..., K$, of the stimulus $\mathbf{s}$ on a set of vectors $\{\hat{e}^{(1)}, \hat{e}^{(2)}, ... \hat{e}^{(K)}\}$:

$$P(\text{spike}|\mathbf{s}) = P(\text{spike}) f(s^{(1)}, s^{(2)}, ... s^{(K)}), \tag{1}$$

where $P(\text{spike}|\mathbf{s})$ is the probability of a spike given a stimulus $\mathbf{s}$ and $P(\text{spike})$ is the average firing rate. In what follows we will call the subspace spanned by the set of vectors $\{\hat{e}^{(\mu)}\}$ the relevant subspace (RS). Even though the ideas developed below can be used to analyze input-output functions with respect to different neural responses, we settle on a single spike as the response of interest.

Eq. (1) in itself is not yet a simplification if the dimensionality K of the RS is equal to the dimensionality D of the stimulus space. In this paper we will use the idea of dimensionality reduction [15, 16] and assume that $K \ll D$. The input-output function f in Eq. (1) can be strongly nonlinear, but it is presumed to depend only on a small number of projections. This assumption appears to be less stringent than that of approximate linearity which one makes when characterizing neuron's response in terms of Wiener kernels. The most difficult part in reconstructing the input-output function is to find the RS. For $K > 1$, a description in terms of any linear combination of vectors $\{\hat{e}^{(\mu)}\}$ is just as valid, since we did not make any assumptions as to a particular form of nonlinear function f. We might however prefer one coordinate system over another if it, for example, leads to sparser probability distributions or more statistically independent variables.

Once the relevant subspace is known, the probability $P(\text{spike}|\mathbf{s})$ becomes a function of only few parameters, and it becomes feasible to map this function experimentally, inverting the probability distributions according to Bayes' rule:

$$f(\{s^{(\mu)}\}) = \frac{P(\{s^{(\mu)}\}|\text{spike})}{P(\{s^{(\mu)}\})} \tag{2}$$

If stimuli are correlated Gaussian noise, then the neural response can be characterized by the spike-triggered covariance method [15, 16]. It can be shown that the dimensionality of the RS is equal to the number of non-zero eigenvalues of a matrix given by a difference between covariance matrices of all presented stimuli and stimuli conditional on a spike. Moreover, the RS is spanned by the eigenvectors associated with the non-zero eigenvalues multiplied by the inverse of the *a priori* covariance matrix. Compared to the reverse correlation method, we are no longer limited to finding only one of the relevant directions $\hat{e}^{(\mu)}$. However because of the necessity to probe a two-point correlation function, the spike-triggered covariance method requires better sampling of distributions of inputs conditional on a spike.

In this paper we investigate whether it is possible to lift the requirement for stimuli to be Gaussian. When using natural stimuli, which are certainly non-Gaussian, the RS cannot be found by the spike-triggered covariance method. Similarly, the reverse correlation method does not give the correct RF, even in the simplest case where the input-output function (1) depends only on one projection. However, vectors that span the RS are clearly special directions in the stimulus space. This notion can be quantified by Shannon information, and

an optimization problem can be formulated to find the RS. Therefore the current implementation of the dimensionality reduction idea is complimentary to the clustering of stimuli done in the information bottleneck method [17]; see also Ref. [18]. Non–information based measures of similarity between probability distributions $P(\mathbf{s})$ and $P(\mathbf{s}|\text{spike})$ have also been proposed [19]. We illustrate how the optimization scheme of maximizing information as function of direction in the stimulus space works with natural stimuli for model orientation sensitive cells with one and two relevant directions, much like simple and complex cells found in primary visual cortex. It is also possible to estimate average errors in the reconstruction. The advantage of this optimization scheme is that it does not rely on any specific statistical properties of the stimulus ensemble, and can be used with natural stimuli.

2   Information as an objective function

When analyzing neural responses, we compare the *a priori* probability distribution of all presented stimuli with the probability distribution of stimuli which lead to a spike. For Gaussian signals, the probability distribution can be characterized by its second moment, the covariance matrix. However, an ensemble of natural stimuli is not Gaussian, so that neither second nor any other finite number of moments is sufficient to describe the probability distribution. In this situation, the Shannon information provides a convenient way of comparing two probability distributions. The average information carried by the arrival time of one spike is given by [20]

$$I_{\text{spike}} = \sum_{\mathbf{s}} P(\mathbf{s}|\text{spike}) \log_2 \left[P(\mathbf{s}|\text{spike})/P(\mathbf{s}) \right] . \tag{3}$$

The information per spike, as written in (3) is difficult to estimate experimentally, since it requires either sampling of the high-dimensional probability distribution $P(\mathbf{s}|\text{spike})$ or a model of how spikes were generated, i.e. the knowledge of low-dimensional RS. However it is possible to calculate I_{spike} in a model-independent way, if stimuli are presented multiple times to estimate the probability distribution $P(\text{spike}|\mathbf{s})$. Then,

$$I_{\text{spike}} = \left\langle \frac{P(\text{spike}|\mathbf{s})}{P(\text{spike})} \log_2 \left[\frac{P(\text{spike}|\mathbf{s})}{P(\text{spike})} \right] \right\rangle_{\mathbf{s}}, \tag{4}$$

where the average is taken over all presented stimuli. Note that for a finite dataset of N repetitions, the obtained value $I_{\text{spike}}(N)$ will be on average larger than $I_{\text{spike}}(\infty)$, with difference $\sim N_{\text{stimuli}}/(N_{\text{spike}} 2 \ln 2)$, where N_{stimuli} is the number of different stimuli, and N_{spike} is the number of elicited spikes [21] across all of the repetitions. The true value I_{spike} can also be found by extrapolating to $N \to \infty$ [22]. The knowledge of the total information per spike will characterize the quality of the reconstruction of the neuron's input-output relation.

Having in mind a model in which spikes are generated according to projection onto a low-dimensional subspace, we start by projecting all of the presented stimuli on a particular direction v in the stimulus space, and form probability distributions $P_v(x|\text{spike}) = \langle \delta(x - \mathbf{s} \cdot v)|\text{spike}\rangle_{\mathbf{s}}$, $P_v(x) = \langle \delta(x - \mathbf{s} \cdot v)\rangle_{\mathbf{s}}$. The information

$$I(v) = \int dx\, P_v(x|\text{spike}) \log_2 \left[P_v(x|\text{spike})/P_v(x) \right], \tag{5}$$

provides an invariant measure of how much the occurrence of a spike is determined by projection on the direction $\hat{v}$. It is a function only of direction in the stimulus space and does not change when vector v is multiplied by a constant. This can be seen by noting that for any probability distribution and any constant c, $P_{cv}(x) = c^{-1} P_v(x/c)$. When evaluated along any vector, $I(v) \leq I_{\text{spike}}$. The total information I_{spike} can be recovered along one particular direction only if $v = \hat{e}^{(1)}$, and the RS is one-dimensional.

By analogy with (5), one could also calculate information $I(v_1, ..., v_n)$ along a set of several directions $\{v_1, ...v_n\}$ based on the multi-point probability distributions:

$$P_{v_1,...,v_n}(\{x_i\}|\text{spike}) = \langle \prod_{i=1}^{n} \delta(x_i - \mathbf{s} \cdot v_i)|\text{spike}\rangle_{\mathbf{s}}, \quad P_{v_1,...,v_n}(\{x_i\}) = \langle \prod_{i=1}^{n} \delta(x_i - \mathbf{s} \cdot v_i)\rangle_{\mathbf{s}}.$$

If we are successful in finding all of the K directions $\hat{e}^{(\mu)}$ in the input-output relation (1), then the information evaluated along the found set will be equal to the total information I_{spike}. When we calculate information along a set of n vectors that are slightly off from the RS, the answer is, of course, smaller than I_{spike} and is quadratic in deviations δv_i. One can therefore find the RS by maximizing information with respect to k vectors simultaneously. The information does not increase if more vectors outside the RS are included into the calculation. On the other hand, the result of optimization with respect to the number of vectors $k < K$ may deviate from the RS if stimuli are correlated. The deviation is also proportional to a weighted average of $P(\text{spike}|s^{(1)}, ..., s^{(k)}) - P(\text{spike}|s^{(1)}, ..., s^{(K)})$. For uncorrelated stimuli, any vector or a set of vectors that maximizes $I(v)$ belongs to the RS. To find the RS, we first maximize $I(v)$, and compare this maximum with I_{spike}, which is estimated according to (4). If the difference exceeds that expected from finite sampling corrections, we increment the number of directions with respect to which information is simultaneously maximized.

The information $I(v)$ as defined by (5) is a continuous function, whose gradient can be computed

$$\nabla_v I = \int dx P_v(x) \left[\langle \mathbf{s}|x, \text{spike}\rangle - \langle \mathbf{s}|x\rangle\right] \frac{d}{dx} \frac{P_v(x|\text{spike})}{P_v(x)}. \tag{6}$$

Since information does not change with the length of the vector, $\nabla_v I \perp v$ (which can also be seen from (6) directly), unnecessary evaluations of information for multiples of v are avoided by maximizing along the gradient. As an optimization algorithm, we have used a combination of gradient ascent and simulated annealing algorithms: successive line maximizations were done along the direction of the gradient. During line maximizations, a point with a smaller value of information was accepted according to Boltzmann statistics, with probability $\propto \exp[(I(v_{i+1}) - I(v_i))/T]$. The effective temperature T is reduced upon completion of each line maximization.

3 Discussion

We tested the scheme of looking for the most informative directions on model neurons that respond to stimuli derived from natural scenes. As stimuli we used patches of digitized to 8-bit scale photos, in which no corrections were made for camera's light intensity transformation function. Our goal is to demonstrate that even though spatial correlations present in natural scenes are non-Gaussian, they can be successfully removed from the estimate of vectors defining the RS.

3.1 Simple Cell

Our first example is taken to mimic properties of simple cells found in the primary visual cortex. A model phase and orientation sensitive cell has a single relevant direction $\hat{e}^{(1)}$ shown in Fig. 1(a). A given frame $\mathbf{s}$ leads to a spike if projection $s^{(1)} = \mathbf{s} \cdot \hat{e}^{(1)}$ reaches a threshold value θ in the presence of noise:

$$\frac{P(\text{spike}|\mathbf{s})}{P(\text{spike})} \equiv f(s^{(1)}) = \langle H(s^{(1)} - \theta + \xi)\rangle, \tag{7}$$
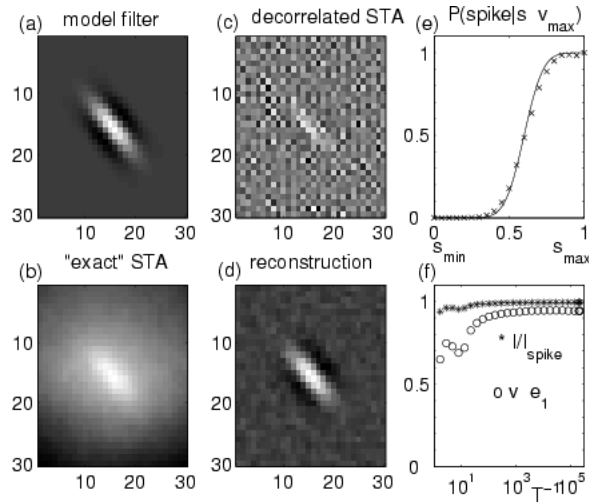
Figure 1: Analysis of a model simple cell with RF shown in (a). The spike-triggered average v_{sta} is shown in (b). Panel (c) shows an attempt to remove correlations according to reverse correlation method, $C^{-1}_{a\,priori}v_{\text{sta}}$; (d) vector v_{max} found by maximizing information; (e) The probability of a spike $P(\text{spike}|\mathbf{s} \cdot \hat{v}_{max})$ (crosses) is compared to $P(\text{spike}|s^{(1)})$ used in generating spikes (solid line). Parameters $\sigma = 0.1(s_{\text{max}} - s_{\text{min}})$ and $\theta = 0.6(s_{\text{max}} - s_{\text{min}})$ [s_{max} and s_{min} are the maximum and minimum values of $s^{(1)}$ over the ensemble of presented stimuli.] (f) Convergence of the algorithm according to information $I(v)$ and projection $v \cdot \hat{e}^{(1)}$ as a function of inverse effective temperature T^{-1}.

where Gaussian random variable ξ of variance σ models additive noise, and function $H(x) = 1$ for $x > 0$, and zero otherwise. Together with the RF $\hat{e}^{(1)}$, the parameters θ for threshold and the noise variance σ determine the input-output function.

The spike-triggered average (STA), shown in Fig. 1(b), is broadened because of spatial correlations present in natural stimuli. If stimuli were drawn from a Gaussian probability distribution, they could be decorrelated by multiplying v_{sta} by the inverse of the *a priori* covariance matrix, according to the reverse correlation method. The procedure is not valid for non-Gaussian stimuli and nonlinear input-output functions (1). The result of such a decorrelation is shown in Fig. 1(c). It is clearly missing the structure of the model filter. However, it is possible to obtain a good estimate of it by maximizing information directly, see panel (d). A typical progress of the simulated annealing algorithm with decreasing temperature T is shown in panel (e). There we plot both the information along the vector, and its projection on $\hat{e}^{(1)}$. The final value of projection depends on the size of the data set, see below. In the example shown in Fig. 1 there were $\approx 50,000$ spikes with average probability of spike ≈ 0.26 per frame. Having reconstructed the RF, one can proceed to sample the nonlinear input-output function. This is done by constructing histograms for $P(\mathbf{s} \cdot \hat{v}_{\text{max}})$ and $P(\mathbf{s} \cdot \hat{v}_{\text{max}}|\text{spike})$ of projections onto vector $\hat{v}_{\text{max}}$ found by maximizing information, and taking their ratio. In Fig. 1(e) we compare $P(\text{spike}|\mathbf{s} \cdot \hat{v}_{\text{max}})$ (crosses) with the probability $P(\text{spike}|s^{(1)})$ used in the model (solid line).

3.2 Estimated deviation from the optimal direction

When information is calculated with respect to a finite data set, the vector $\hat{v}$ which maximizes I will deviate from the true RF $\hat{e}^{(1)}$. The deviation $\delta v = \hat{v} - \hat{e}^{(1)}$ arises because the probability distributions are estimated from experimental histograms and differ from the
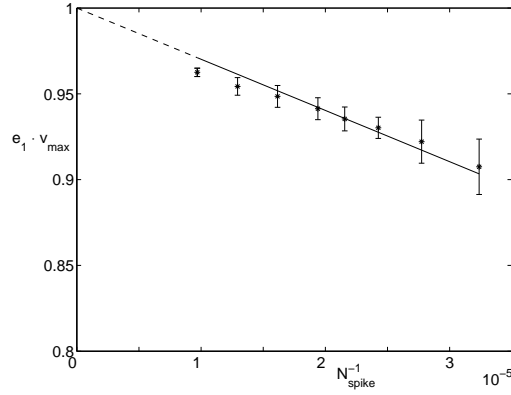
Figure 2: Projection of vector $v_{\max}$ that maximizes information on RF $\hat{e}^{(1)}$ is plotted as a function of the number of spikes to show the linear scaling in $1/N_{\text{spike}}$ (solid line is a fit).

distributions found in the limit on infinite data size. For a simple cell, the quality of reconstruction can be characterized by the projection $\hat{v} \cdot \hat{e}^{(1)} = 1 - \frac{1}{2}\delta v^2$, where both $\hat{v}$ and $\hat{e}^{(1)}$ are normalized, and δv is by definition orthogonal to $\hat{e}^{(1)}$. The deviation $\delta v \sim A^{-1}\nabla I$, where A is the Hessian of information. Its structure is similar to that of a covariance matrix:

$$A_{ij} = \frac{1}{\ln 2} \int dx P(x|\text{spike}) \left(\frac{d}{dx} \ln \frac{P(x|\text{spike})}{P(x)} \right)^2 (\langle s_i s_j | x \rangle - \langle s_i | x \rangle \langle s_j | x \rangle) \quad (8)$$

When averaged over possible outcomes of N trials, the gradient of information is zero for the optimal direction. Here in order to evaluate $\langle \delta v^2 \rangle = \text{Tr}[A^{-1}\langle \nabla I \nabla I^T \rangle A^{-1}]$, we need to know the variance of the gradient of I. By discretizing both the space of stimuli and possible projections x, and assuming that the probability of generating a spike is independent for different bins, one could obtain that $\langle \nabla I_i \nabla I_j \rangle \sim A_{ij}/(N_{\text{spike}} \ln 2)$. Therefore an expected error in the reconstruction of the optimal filter is inversely proportional to the number of spikes and is given by:

$$1 - \hat{v} \cdot \hat{e}^{(1)} \approx \frac{1}{2}\langle \delta v^2 \rangle = \frac{\text{Tr}'[A^{-1}]}{2N_{\text{spike}} \ln 2}, \quad (9)$$

where Tr' means that the trace is taken in the subspace orthogonal to the model filter, since by definition $\delta \mathbf{v} \cdot \hat{e}^{(1)} = 0$. In Fig. 2 we plot the average projection of the normalized reconstructed vector $\hat{v}$ on the RF $\hat{e}^{(1)}$, and show that it scales with the number of spikes.

3.3 Complex Cell

A sequence of spikes from a model cell with two relevant directions was simulated by projecting each of the stimuli on vectors that differ by $\pi/2$ in their spatial phase, taken to mimic properties of complex cells, see Fig. 3. A particular frame leads to a spike according to a logical OR, that is if either $s^{(1)} = \mathbf{s} \cdot \hat{e}^{(1)}$, $-s^{(1)}$, $s^{(2)} = \mathbf{s} \cdot \hat{e}^{(2)}$, or $-s^{(2)}$ exceeds a threshold value s_t in the presence of noise. Similarly to (7),

$$\frac{P(\text{spike}|\mathbf{s})}{P(\text{spike})} = f(s^{(1)}, s^{(2)}) = \langle H(|s^{(1)}| - \theta - \xi_1) \ \lor \ H(|s^{(2)}| - \theta - \xi_2) \rangle, \quad (10)$$

where ξ_1 and ξ_2 are independent Gaussian variables. The sampling of this input-output function by our particular set of natural stimuli is shown in Fig. 3(c). Some, especially large, combinations of values of $s^{(1)}$ and $s^{(2)}$ are not present in the ensemble.

We start by maximizing information with respect to one direction. Contrary to analysis for a simple cell, one optimal direction recovers only about 60% of the total information per spike. This is significantly different from the total I_{spike} for stimuli drawn from natural scenes, where due to correlations even a random vector has a high probability of explaining 60% of total information per spike. We therefore go on to maximize information with respect to two directions. An example of the reconstruction of input-output function of a complex cell is given in Fig. 3. Vectors v_1 and v_2 that maximize $I(v_1, v_2)$ are not orthogonal, and are also rotated with respect to $\hat{e}^{(1)}$ and $\hat{e}^{(2)}$. However, the quality of reconstruction is independent of a particular choice of basis with the RS. The appropriate measure of similarity between the two planes is the dot product of their normals. In the example of Fig. 3, $\hat{n}_{(\hat{e}^{(1)}, \hat{e}^{(2)})} \cdot \hat{n}_{(\hat{v}_1, \hat{v}_2)} \approx 0.8$.

Maximizing information with respect to two directions requires a significantly slower cooling rate, and consequently longer computational times. However, an expected error in the reconstruction, $1 - \hat{n}_{(\hat{e}^{(1)}, \hat{e}^{(2)})} \cdot \hat{n}_{(\hat{v}_1, \hat{v}_2)}$, follows a N_{spike}^{-1} behavior, similarly to (9), and is roughly twice that for a simple cell given the same number of spikes. In this calculation there were $\approx 10^5$ spikes.
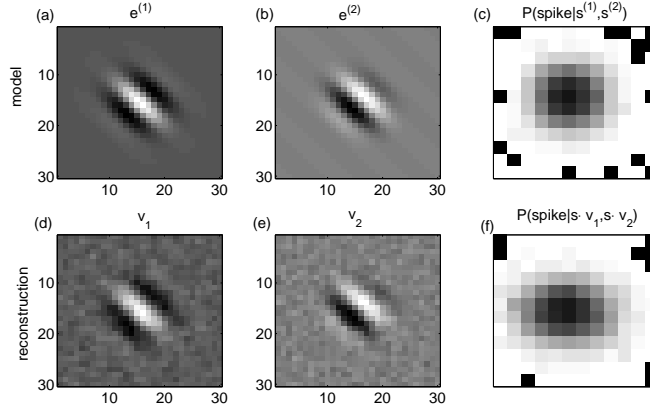

Figure 3: Analysis of a model complex cell with relevant directions $\hat{e}^{(1)}$ and $\hat{e}^{(2)}$ shown in (a) and (b). Spikes are generated according to an "OR" input-output function $f(s^{(1)}, s^{(2)})$ with the threshold $\theta = 0.6(s_{\max} - s_{\min})$ and noise variance $\sigma = 0.05(s_{\max} - s_{\min})$. Panel (c) shows how the input-output function is sampled by our ensemble of stimuli. Dark pixels for large values of $s^{(1)}$ and $s^{(2)}$ correspond to cases where $P(s^{(1)}, s^{(2)}) = 0$. Below, we show vectors v_1 and v_2 found by maximizing information $I(v_1, v_2)$ together with the corresponding input-output function with respect to projections $\mathbf{s} \cdot v_1$ and $\mathbf{s} \cdot v_2$.

In conclusion, features of the stimulus that are most relevant for generating the response of a neuron can be found by maximizing information between the sequence of responses and the projection of stimuli on trial vectors within the stimulus space. Calculated in this manner, information becomes a function of direction in a stimulus space. Those directions that maximize the information and account for the total information per response of interest span the relevant subspace. This analysis allows the reconstruction of the relevant subspace without assuming a particular form of the input-output function. It can be strongly nonlinear within the relevant subspace, and is to be estimated from experimental histograms. Most importantly, this method can be used with any stimulus ensemble, even those that are strongly non-Gaussian as in the case of natural images.

Acknowledgments

oration began at the Marine Biological Laboratory in a course supported by grants from NIMH and the Howard Hughes Medical Institute.

References

[1] F. Rieke, D. A. Bodnar, and W. Bialek. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. Lond. B*, 262:259–265, (1995).

[2] W. E. Vinje and J. L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287:1273–1276, 2000.

[3] F. E. Theunissen, K. Sen, and A. J. Doupe. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.*, 20:2315–2331, 2000.

[4] G. D. Lewen, W. Bialek, and R. R. de Ruyter van Steveninck. Neural coding of naturalistic motion stimuli. *Network: Comput. Neural Syst.*, 12:317–329, 2001.

[5] N. J. Vickers, T. A. Christensen, T. Baker, and J. G. Hildebrand. Odour-plume dynamics influence the brain's olfactory code. *Nature*, 410:466–470, 2001.

[6] K. Sen, F. E. Theunissen, and A. J. Doupe. Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.*, 86:1445–1458, 2001.

[7] D. L. Ringach, M. J. Hawken, and R. Shapley. Receptive field structure of neurons in monkey visual cortex revealed by stimulation with natural image sequences. *Journal of Vision*, 2:12–24, 2002.

[8] W. E. Vinje and J. L. Gallant. Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J. Neurosci.*, 22:2904–2915, 2002.

[9] D. L. Ruderman and W. Bialek. Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.*, 73:814–817, 1994.

[10] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4:2379–2394, 1987.

[11] P. Kara, P. Reinagel, and R. C. Reid. Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27:635–646, 2000.

[12] R. R. de Ruyter van Steveninck, G. D. Lewen, S. P. Strong, R. Koberle, and W. Bialek. Reproducibility and variability in neural spike trains. *Science*, 275:1805–1808, 1997.

[13] F. Rieke, D. Warland, R. R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the neural code*. MIT Press, Cambridge, 1997.

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

[15] N. Brenner, W. Bialek, and R. R. de Ruyter van Steveninck. Adaptive rescaling maximizes information transmission. *Neuron*, 26:695–702, 2000.

[16] R. R. de Ruyter van Steveninck and W. Bialek. Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc. R. Soc. Lond. B*, 234:379–414, 1988.

[17] N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. In *Proceedings of the 37th Allerton Conference on Communication, Control and Computing*, edited by B. Hajek & R. S. Sreenivas. University of Illinois, 368–377, 1999.

[18] A. G. Dimitrov and J. P. Miller. Neural coding and decoding: communication channels and quantization. *Network: Comput. Neural Syst.*, 12:441–472, 2001.

[19] L. Paninski. Convergence properties of some spike-triggered analysis techniques. In *Advances in Neural Information Processing 15*, edited by 2003.

[20] N. Brenner, S. P. Strong, R. Koberle, W Bialek, and R. R. de Ruyter van Steveninck. Synergy in a neural code. *Neural Comp.*, 12:1531-1552, 2000.

[21] A. Treves and S. Panzeri. The upward bias in measures of information derived from limited data samples. *Neural Comp.*, 7:399, 1995.

[22] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek. Entropy and information in neural spike trains. *Phys. Rev. Lett.*, 80:197–200, 1998.