Information theoretic measure of stimulus significance is not confounded by stimulus correlations or non-linearities

Daniel A. Butts, Patrick O. Kanold

Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115
dbutts@hms.harvard.edu

Information theory provides tools to analyze neural ensembles that do not depend on assumptions of linearity or correlations in the stimulus ensemble. Here, we use a newly derived information theoretic measure, the stimulus-specific information (SSI), to analyze the important stimuli to simulated visual neurons. Using full-frame flicker stimuli as input to a simulated LGN neuron, we find that the SSI returns the linear kernel of the neuron. For correlated input, linear reconstruction is confounded by correlations in the stimulus. However, we show that the SSI reliably finds the linear kernel of the neuron independent of stimulus correlations. Thus, the SSI is a general tool to discern the important stimuli of neurons, independent of stimulus correlations. It should also be useful for neurons with non-linear response properties.

SUMMARY

The use of information theoretic techniques provides unbiased measures to compare different coding schemes in neural ensembles. Analyzing the neural code using information removes both inherent biases caused by preconceptions of the coding schemes (such as choosing a particular set of stimuli), as well as complications inherent in analyzing neural data such as non-linearities and correlations in the stimuli. To access the code underlying neural behavior, it is useful to know what symbols are important to the neuron, and how those symbols are conveyed by neuronal output.

Typically, neurons are characterized by their linear response properties. For example, in the case of visual neurons, the spike-triggered average stimulus is used to describe its receptive field. While linear responses seem to explain most of the behavior of neurons early in the visual pathway (retinal ganglion cells and thalamocortical neurons), information theory is still necessary to characterize effects such as synchronization and non-classical receptive field effects. Furthermore, such linear techniques are less successful in later stages of visual processing; for example, they completely break down in the case of complex cells.

As a result, techniques that can characterize the response properties of [visual] neurons independent of bias in stimulus ensemble, correlations within this ensemble, and non-linearities in the neural code are necessary for a full characterization of the neural code as well as comparison between different neurons. Here, we develop new techniques that use information theory to fully characterize the neural code of visual system neurons. These new methods are an extension of the work of Liu et al (2001) and Butts (2002). In this work, we demonstrate that these new methods reliably characterize significant features of neuronal responses using a realistic neuronal model. Though we have not yet characterized cells with significant non-linearities, here we demonstrate that these techniques work where linear techniques fail: in the case of correlated input.

SHORT TALK

The mutual information can be decomposed to reveal significant features of the stimulus and response. Despite the symmetry of mutual information, meaningful decompositions of stimuli and of responses differ. DeWeese and Meister (1999) proposed the *specific information* $I_2(r)$ as a measure of information conveyed by particular responses r, given by:

$I_2(r) = H[S] - H[S|\, r]$ (where $H[X]$ is the entropy of the probability distribution of the set X).

Specific information of a response r is the reduction of uncertainty (i.e., the change in entropy) gained by measuring the particular response r. The appropriate information-theoretic measure of stimulus significance is the *stimulus-specific information* (SSI), proposed by Butts (2002). The SSI of a particular stimulus s is defined to be:

$$I_3(s) = \sum_{r \in R} p(r \mid s)\left\{H[S] - H[S \mid r]\right\} = \sum_{r \in R} p(r \mid s) I_2(r),$$

The SSI is the average reduction in uncertainty per measurement for a particular stimulus, meaning that stimuli with a relatively large SSI are well encoded by the neural system. Also, its straight-forward interpretation makes it directly comparable to both the specific information of a given response, and the total mutual information $I[R,S]$, which is the average reduction in uncertainty from one measurement over all possible stimuli. Note that $I[R,S]=\sum p(s)\, I_3(s)=\sum p(r)\, I_2(r)$.

We demonstrate the use of these information theoretic measures in simulated visual system neurons. Keat et al (2001) have produced models of retinal ganglion cells and thalamocortical neurons that reproduce the precise timing of action potentials relative to full-field flicker stimuli, including their variability from trial to trial. This model is essentially a linear kernel with a mechanism to simulate the neural refractory period, with two noise sources. We use this model to produce realistic visual system spike trains, which we then analyze using both linear methods such as reverse correlation and information theoretic measures.

Using uncorrelated full-field-flicker input, we find that the *information receptive field*, representing the most significant stimuli, matches the reverse reconstructed stimulus precisely. Because the model neuron is essentially linear, information theoretic techniques return this linear behavior. These results show that information theoretic measures correctly characterize the cells linear response properties.

Analysis techniques based on linear techniques such as reverse correlations break down when the stimulus contains correlations or when non-linear cells (i.e. cortical complex cells) are probed. However, with temporally correlated full-field flicker input, the SSI yields the same information receptive field of the neuron independent of stimulus correlations.

We also find that information receptive fields can be calculated without an excessive amount of data: 50000-100000 stimulus frames – or 400-800 seconds. This is only slightly more than what is typically used to estimate the linear properties of neurons using reverse correlation methods. Thus using these techniques is experimentally feasible.

To summarize, in this work, we present here a new information theoretic measure and apply it to realistic neurons and spike trains. We show that the *information-receptive field* of the neuron can be characterized independent of stimulus correlations extending the classical analysis of the linear response properties of neurons with uncorrelated stimuli and reverse correlation techniques. Therefore information theoretic measures such as the SSI are a general method to identify the salient features of the neural code.

SHORT TALK

References:

Butts DA (2002). What is the information of stimuli? *In submission*. Presented at CNS 2001.

Butts DA, Rokhsar DS (2001) The information content of spontaneous retinal waves. *J. Neurosci.* **21**: 961-973.

DeWeese MR, Meister M (1999) How to measure the information gained from one symbol. *Network: Comput. Neural Syst.* **10**: 325-340.

Keat J, Reinagel P, Reid RC, Meister M (2001). Predicting every spike: a model for the responses of visual neurons. *Neuron* **30**: 803-817.

Liu R, Tzonev S, Rebrik S, Miller KD (2001) Variability and information in a neural code of the cat lateral geniculate nucleus. *J Neurophysiol.***86**: 2789-806.