Fisher vs Shannon information in Populations of Neurons

Edward A. L. Challis, Stuart Yarrow, Peggy Series

FISHER VS SHANNON INFORMATION IN POPULATIONS OF NEURONS

Edward A. L. Challis, Stuart Yarrow and Peggy Seriès
Institute for Adaptive and Neural Computation (IANC)
University of Edinburgh, 10 Crichton Street,
U.K. EH8 9AB.
email: pseries@inf.ed.ac.uk

ABSTRACT
The accuracy of the neural code is commonly investigated using two different measures: (i) Shannon mutual information and derived quantities when investigating very small populations of neurons and (ii) Fisher information when studying large populations. How these measures compare in finite size populations has not been systematically explored. We here aim at filling this gap. We are particularly interested in understanding which stimuli are best encoded by a given neuron in a population and how this depends on the chosen measure. In models of independent neurons, we find that the predictions of Fisher information and of a stimulus-specific decomposition of Shannon information (the SSI) agree very well, even for relatively small population sizes. According to both measures, the stimuli that are best encoded are then those falling at the flanks of the neuron's tuning curve.

KEY WORDS
Fisher Information, Information theory, Neural Coding.

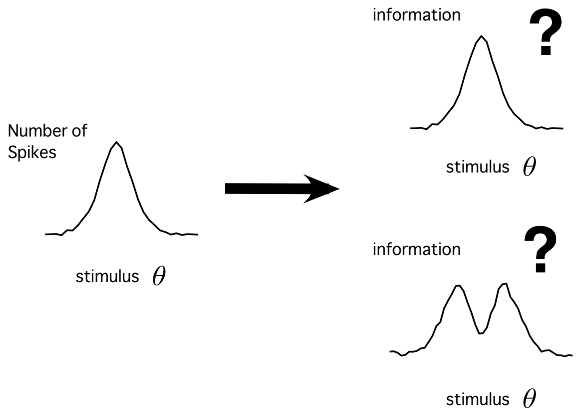

Figure 1. Neurons are commonly characterized by their response tuning curves, which describe the relationship between the stimulus and the spike count. In addition to this description, information measures can also be used to quantify how well each stimulus is encoded by a particular neuron. How does the information tuning curve look like? Is the information maximal where the response is maximal (i.e. at the peak of the tuning curve), or where the slope of the tuning curve is maximal? Does the answer of this question depend on the chosen measure? Does it depend on the population size?

1 Introduction

Understanding how information is represented and transformed by populations of neurons is a major goal of neuroscience research. The accuracy with which information is encoded is classically quantified using two types of measures: (i) Shannon mutual information (MI) and derived quantities and (ii) Fisher information (FI). Shannon information is commonly used to quantify the accuracy of single neurons or pairs of neurons (see e.g. [1]). When populations of neurons are studied, however, FI is typically chosen, both in theory and in experiments (see e.g. [2, 3, 4]). This is because MI is thought to be computationally intractable for more than a few cells (e.g. [5]). MI and FI belong to different statistical fields: information theory and statistical parameter estimation, respectively. MI can be thought of as a measure of statistical dependency between the stimulus ensemble and the response ensemble and is measured in bits. FI, on the other hand, is related to the performance of an ideal estimator whose task would be to estimate the stimulus based on the neural responses, and is measured in units of the inverse of the variance of the estimator. An asymptotic relation (i.e. in the limit of infinite populations) between these quantities has recently been exposed [6, 7], showing that MI is related to the log of the FI. What remains unclear, however, is how these measures compare in plausible conditions with finite-size populations of neurons, and – in situations where their predictions differ – which measure is more relevant to the understanding of the neural code.

In this work, we aim at clarifying these issues. We compare these two measures to investigate the following questions (Fig. 1): What stimuli are best encoded by a single neuron or a population of neurons? Are neurons best at encoding stimuli that elicit maximal response; or stimuli that fall at the flanks of the tuning curve, where the slope is maximal? How does the answer to these questions depend on the chosen information measure?

2 Methods

2.1 Model Population

We consider a population of N neurons with tuning curves $\mathbf{f}(\theta) = \{f_1(\theta), f_2(\theta), .., f_N(\theta)\}$ describing the mean firing rate of each neuron as a function of the a unidimensional continuous stimulus, for e.g. the direction θ of a moving bar. These N neurons tile the space of all directions uniformly and the $f_i(\theta)$ are described using the circular normal distribution:

$$f_i(\theta) = f_{max}\text{exp}(\sigma_i^{-1}(\cos(\theta - \theta_i) - 1) \tag{1}$$

The response variability over many presentations of the same stimulus is assumed to be Gaussian with a variance proportional to the mean spike count, and independent between neurons. F denotes the Fano Factor, i.e. the relationship between the variance of the spike count and the mean spike count. τ denotes the temporal window over which spikes are counted. The response of neuron i can be modeled as:

$$r_i = \tau f_i(s) + \eta_i \ , \tag{2}$$

where the η_i's are independent Gaussian random variables. The response statistics of the population response to stimulus θ are described by the probability density $P(\mathbf{r}|\theta)$, where $\mathbf{r} = \{r_1, r_2..., r_N\}$ is the vector of the spike counts of all neurons on each trial.

We used this model to compare systematically the predictions of FI, MI and the Stimulus-Specific Information (SSI).

2.2 Fisher Information

For Gaussian noise, Fisher information can be directly written and computed using [8]:

$$I_F(\theta) = \tau^2 \mathbf{f}'(\theta)^{\mathbf{T}} \mathbf{Q}(\theta)^{-1} \mathbf{f}'(\theta) + \frac{1}{2}\text{Tr}\left(\mathbf{Q}'(\theta)\mathbf{Q}(\theta)^{-1}\mathbf{Q}'(\theta)\mathbf{Q}^{-1}(\theta)\right)$$

where $f'(\theta)$ is the derivative of the tuning curve with respect to θ, and $\mathbf{Q^{-1}}$ and $\mathbf{Q'}$ are the inverse and derivative of the covariance matrix. For independent noise, $\mathbf{Q}(\theta)$ is defined by: $\begin{cases} \mathbf{Q_{ii}}(\theta) = \tau F f_i(\theta) \\ \mathbf{Q_{i \neq j}}(\theta) = 0 \end{cases}$.

2.3 Stimulus-Specific Information (SSI)

SSI is a stimulus-specific decomposition of Shannon information. Based on the specific information (SI) [9], it was shown to be an appropriate measure of the information associated with particular stimuli [10]. The SI is defined as:

$$i_{sp}(r) = H[\Theta] - H[\Theta|\mathbf{r}] \tag{3}$$

i.e. it is the difference between the entropy of the stimulus ensemble $H[\Theta] = -\sum p(\theta)\log_2 p(\theta)$ and that of the stimulus distribution conditional on a particular measurement $\mathbf{r}$, $H[\Theta] = -\sum p(\theta|\mathbf{r})\log_2 p(\theta|\mathbf{r})$. The SSI is the average SI when a particular stimulus θ is present:

$$i_{SSI}(\theta) = \sum_r p(\mathbf{r}|\theta) i_{sp}(\mathbf{r}) \tag{4}$$

The marginal SSI (for a particular neuron in the population) is defined as the difference between the population SSI and the SSI for the population of the remaining neurons after the neuron of interest is removed.

Unlike previous investigations of the SSI [5], which used the method of quadrature, we used Monte-Carlo Integration (see e.g. [11]) to compute the SSI. This means that we approximated the SSI by sampling the multidimensional gaussian $P(\mathbf{r}|\theta)$ n times for each value of θ (giving us $\mathbf{r}_1, .., \mathbf{r}_j, ..\mathbf{r}_n$) then by averaging $i_{sp}(\mathbf{r_j})$ at these values:

$$i_{SSI}(\theta) \simeq \frac{1}{n}\sum_{j=1}^{n} i_{sp}(\mathbf{r_j}) \tag{5}$$

We found that this allowed the SSI to be computed for much larger population sizes than originally investigated ($\simeq 200$ instead of 4) [12]. We systematically compare how individual neurons participate in population codes by computing the FI and the marginal SSI for increasing population sizes, varying time-windows and different noise levels.

3 Results

3.1 Small populations

We first replicated the results of [5]. These authors applied the SSI to very small populations of neurons (N=1-4) and showed that which stimuli are best encoded by a neuron depends on the level of variability. In a 'low noise' regime (Fig 2a), the best encoded stimuli fall at the flanks of the tuning curve. In a 'high noise' regime (Fig 2b), however, the best encoded stimulus corresponds to the peak of the neuron curve. The authors concluded that both intuitions regarding which stimuli are best encoded by individual neurons (Fig 1) are correct, depending on the level of variability. These results are in strong contrast with the predictions of FI, which always favors the tuning curves' flanks.

3.2 Varying the integration time window

The results of [5] can be first extended to the temporal domain. Increasing the temporal window τ over which spikes are counted has the same effect as reducing the variability: the SSI predicts that over time, single neurons dynamically switch from being best at encoding stimuli corresponding to the peak of the tuning curve, to stimuli falling at the flanks of the tuning curve (Fig. 3).

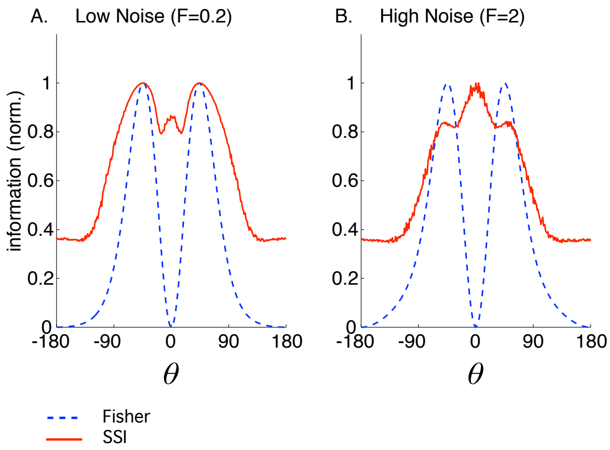

Figure 2. Comparison between the SSI (full lines) and Fisher Information (dashed) for a single neuron. **A.** When the noise level is low (F=0.2), the stimuli that are best encoded according to the SSI fall at the flanks of the tuning curve. **B.** For high noise (F=2), the stimuli that are best encoded correspond to the peak of the tuning curve [5]. On the other hand, Fisher Information always predicts that the stimuli that are best encoded fall at the flanks of the tuning curve. Tuning curve parameters: $f_{max} = 50$ Hz; $\sigma = 33$ deg; T=1 sec. Both measures are normalized to their maximum value.

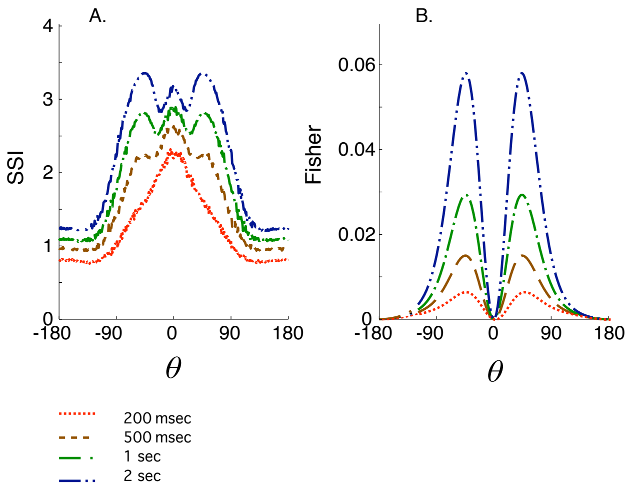

Figure 3. Comparison between the SSI (A) and Fisher Information (B) for a single neuron when the integration time window is increased, from τ=200 msec to $\tau = 2$ sec. **A.** The SSI predicts that single neurons shift dynamically in time from encoding best the stimuli falling at the peak of the tuning curve, to encoding best the stimuli that fall at the regions of maximal slopes of the tuning curve. **B.** The shape of the Fisher Information, on the other hand stays similar with maxima at regions of maximal slope. Same parameters as above; F=1.

3.3 Large populations

The analysis of [5] was then extended to test whether the low and high noise regimes continue to coexist for larger populations of neurons. We computed the SSI for up to 200 neurons for varying temporal windows (20 msec - 1 sec).

We show that the predictions of the SSI and FI converge very rapidly as a function of the number of neurons in the population. The exact pattern of convergence depends on the parameters of the chosen model. However, we found that in populations of 50 neurons, they are qualitatively identical, even for very small integration windows and high noise (Fig. 4). The stimuli that are best encoded are then always those falling at the flanks of the tuning curves.

This indicates that there is no need to go to extremely large numbers for the SSI and the FI to lead to similar predictions. They actually differ in a very restricted domain (small temporal windows, small populations, high noise) which seems to roughly correspond to the range where Fisher Information 'fails' [12] (i.e. where the Cramér-Rao Bound is not reached by maximum-likelihood or other optimal decoders [13, 14]).

3.4 Extensions

We are working on a number of extensions:

- To assess the generality of our results, a similar analysis can be used to explore different Shannon-based measures, in particular the 'specific surprise' [9].

- We are interested in exploring measures that include explicitly the description of a task (e.g. fine vs broad discrimination as in [5], or detection). The SSI and FI can be interpreted as measures of discrimination performance. Intuitively, information tuning curves related to *detection* performance should look different, with a peak where the response is maximal.

- We are also interested in comparing the SSI and FI in heterogeneous populations. Of particular interest is the situation where coding properties change locally, for eg. when some tuning curves are gain-modulated or sharpened due to adaptation, attention or perceptual learning.

- Preliminary analysis shows that computing the SSI is also possible for large populations (up to 200 neurons) when the noise is correlated [15]. Depending on the structure of the correlations, the convergence between FI and MI may be disrupted.

- Other important extensions include using other models for the response variability, and models where multiple stimulus dimensions are encoded.

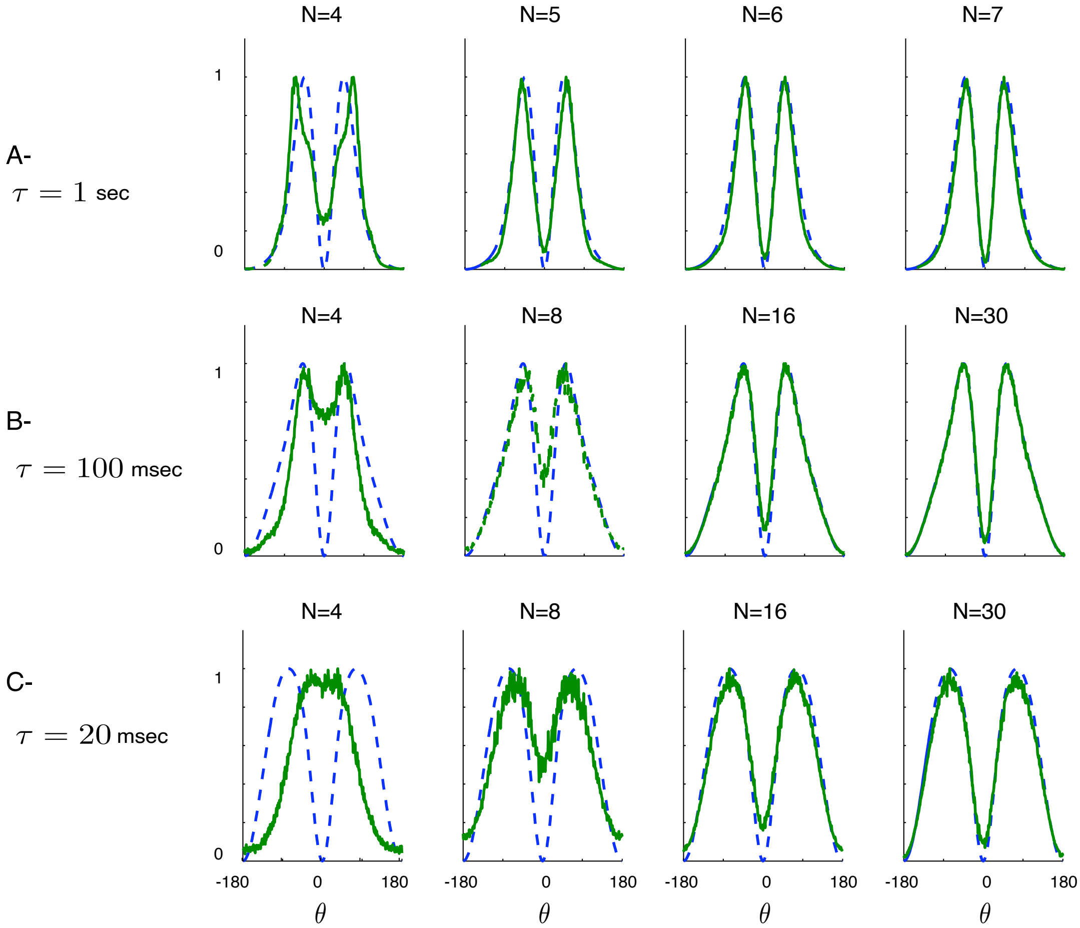

Figure 4. Example comparison between the marginal SSI (plain lines) and Fisher Information (dashed) for a single neuron with a preferred θ of 0 deg, in a population of 4 to 30 neurons, using different time windows: A. τ=1 sec; B. τ=100 msec; C. τ=20 msec. As the population size increases, the predictions of SSI and FI converge. The stimuli that are best encoded are those falling at the flanks of the tuning curve. F=1.

4 Conclusions

We show that Monte-Carlo integration allows a direct comparison between Fisher Information and a Shannon-based measure for small and large populations, thus clarifying the relationship between these two measures in plausible neural settings. Our simulations show that the predictions of FI are compatible with SSI, even for relatively small populations of neurons. This is reassuring, and validates the use of Fisher Information in finite size populations, which is much easier to compute than Shannon-based measures. This also suggests that the 'high noise' regime where the best encoded stimuli are at the peak of the tuning curve (as described by [5]) is in fact very restricted and occurs only for very small numbers of neurons, very small time windows and high response variability. Whether this regime is relevant to the brain is an important question, which will deserve further investigation.

References

[1] A. Borst and F. E. Theunissen. Information theory and neural coding. *Nat Neurosci*, 2(11):947–957, Nov 1999.

[2] P. Seriès, P. E. Latham, and A. Pouget. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nat Neurosci*, 7(10):1129–1135, Oct 2004.

[3] D. A. Gutnisky and V. Dragoi. Adaptive coding of visual information in neural populations. *Nature*, 452(7184):220–224, Mar 2008.

[4] I. Dean, N. S. Harper, and D. McAlpine. Neural population coding of sound level adapts to stimulus statistics. *Nat Neurosci*, 8(12):1684–1689, Dec 2005.

[5] D.A. Butts and M. Goldman. Tuning curves, neuronal variability and sensory coding. *PLOS Biol.*, 4:639–646, 2006.

[6] N. Brunel and J. P. Nadal. Mutual information, Fisher information, and population coding. *Neural Comput*, 10(7):1731–57, 1998.

[7] K. Kang and H. Sompolinsky. Mutual information of population codes and distance measures in probability space. *Phys Rev Lett*, 86(21):4958–4961, May 2001.

[8] L. F. Abbott and P. Dayan. The effect of correlated variability on the accuracy of a population code. *Neural Comput*, 11(1):91–101, 1999.

[9] M. DeWeese and M. Meister. How to measure the information gained from one symbol. *Network: Comput. Neural Syst.*, 10:325–340, 1999.

[10] D. A. Butts. How much information is associated with a particular stimulus? *Network*, 14(2):177–187, May 2003.

[11] W.H. Press, B. P. Flannery, S. A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, 1993.

[12] E. A. L. Challis. Measuring the accuracy of the neural code. Master's thesis, Informatics, University of Edinburgh, 2007.

[13] M. Bethge, D. Rotermund, and K. Pawelsik. Optimal short-term population coding: when Fisher information fails. *Neural comput.*, 14:2317–2351, 2002.

[14] X. Xie. Threshold behaviour of the maximum likelihood method in population decoding. *Network*, 13:447–456, 2002.

[15] S. Yarrow. Accuracy measures for neural population codes with correlated variability. Master's thesis, Informatics, University of Edinburgh, 2008.