Discrimination Measure of Correlations in a Population of Neurons by Using the Jensen-Shannon Divergence.

F. Montani*, O. A. Rosso† and S. R. Schultz*

*Department of Bioengineering, Imperial College London, South Kensington London SW7 2AZ, UK.
†Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.
Pabellon II, 1428 Ciudad Universitaria, Buenos Aires, Argentina.

Abstract. The significance of synchronized spikes fired by nearby neurons for perception is still unclear. To evaluate how reliably one can decide if a given response on the population coding of sensory information comes from the full distribution, or from the product of independent distributions from each cell, we used recorded responses of pairs of single neurons in primary visual cortex of macaque monkey (V1) to stimuli of varying orientation. Both trial-to-trial variability and synchrony were found to depend stimulus orientation and contrast in this data set (A. Kohn, and M. A Smith, J. Neurosci. 25 (2005) 3661). We used the Jensen-Shannon Divergence for fixed stimuli as a measure of discrimination between a pairs of correlated cells VI. The Jensen-Shannon divergence, can be consider as a measure distance between the corresponding probability distribution function associated with each spikes fired observed patterns. The Nemenman-Shafee-Bialek estimator was used in our entropy estimation in order to remove all possible bias deviation from our calculations. We found that the relative Jensen-Shannon Divergence (measured in relation to case in which all cell fired completely independently) decreases with respect to the difference in orientation preference between the receptive field from each pair of cells. Our finding indicates that the Jensen-Shannon Divergence may be used for characterizing the effective circuitry network in a population of neurons.

Keywords: Neuron populations, discrimination measures, correlations, visual cortex, Jensen-Shannon divergence.
PACS: 87.19.La, 87.19.Dd, 87.10.+e

INTRODUCTION

One of the main challenges in neuroscience during the last 30 years was to demonstrate that distributed neural populations in the visual cortex process information in a cooperative way. The visual cortex is composed of a large number of areas, which contain neurons that are tuned to different visual features. Temporally correlated activity of individual neuronal pairs within the visual cortex has been investigated in many laboratories starting in the early 1980's [1, 2, 3, 5, 4, 6, 7, 8, 9, 10] most often with a motivation to reveal structural coupling between cells. This approach to functional anatomy had been methodologically and conceptually outlined and successfully applied to the invertebrate nervous system [11, 12, 13, 14, 15, 16]. The results of these many different studies taken together clearly demonstrated the presence of temporally precise correlation within the cat and macaque visual cortex. But, these correlations occurred preferentially between cells showing the same feature preference.

Cortical neurons with similar stimulus selectivities are found in close proximity to each other [17, 18, 19]. How neurons in the cerebral cortex represent a stimulus taking into account that the response variability is correlated across neurons is still highly debated [20, 21, 22, 23]. At another level, these correlations when present may have significant effect on the population coding of sensory information. Several pieces of evidence point to correlations between V1 neurons being orientation and contrast dependent [24, 25]. The importance of identifying appropriate measures of correlation, independence and quantify their relation to the stimulus has been pointed out by [26].

The Jensen-Shannon divergence measures how reliably one can decide if a given response comes from the full distribution, $p(r_1 r_2)$, or the product of independent distributions, $p(r_1) \cdot p(r_2)$. In present work we use the Jensen-Shannon Divergence for fixed stimuli as a measure of discrimination between a pairs of correlated cells in the primary visual cortex of anesthetized macaque monkeys [24]. Our findings indicates that the Jensen-Shannon Divergence may be used for characterizing the circuitry network in a population of neurons.
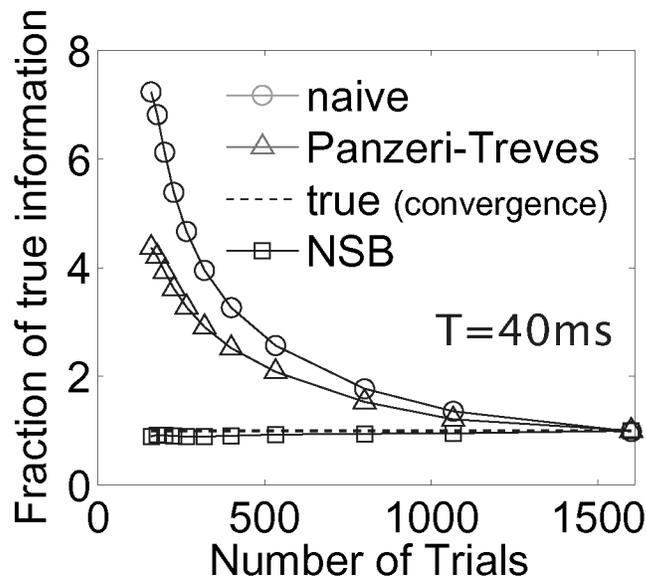
FIGURE 1. Performance comparison of several information estimators. The shows the total information estimated through these different approaches.
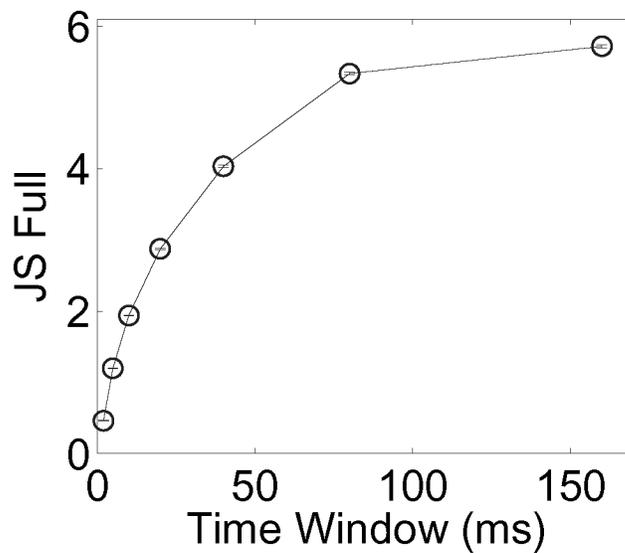

FIGURE 2. For a tipycal pair of cells in V1 the JS was calculated taken into account the full sitmuli orientation dependency at seven different time windows.

METHODS

Entropy estimations

The problem with real measures of entropy is that they depend on a limited number of samples provided by the experiment. It is therefore important to use a theoretical approach that can remove sample size dependent bias from the entropy estimations. But, it is well known that approaches usually performed to estimate the entropy tend to

185

underestimate it since they are biased, and this effect can only be avoided by the use of a perturbative expansion in the asymptotic regime [27]. The idea of calculating entropy by counting coincidences was proposed a long time ago by Ma [28] for physical systems in the micro-canonical ensemble, where a uniform distribution of entropy corresponds to states of fixed energy. The Bayesian prior proposed by Nemenman, Bialek and Steveninck (NSB) [29] extends this idea to an arbitrarily complex distribution. The goal of this method is to construct a Bayesian prior, which generates a nearly uniform distribution of entropies in order to correct sample size dependent bias at its source.

In the following we review the basic ideas of the NSB entropy estimator method [29]. Let consider the problem of estimating the Shannon entropy for a given probability distribution $\mathbf{p} \equiv \{p_i\}$,

$$H_{naive} = \sum_{i=1}^{K} f_i \log_2 f_i ,\qquad(1)$$

where the index i runs over K possibilities. Considering N samples (trials) which were obtained from a given experiment, where each possibility i occurred n_i times. If N is much bigger than K we can approximate $p_i \approx f_i = n_i/N$, and $\mathbf{p}$ might be the distribution of spike counts observed to be fired by a neuron.

The present approach tends to underestimate the entropy and attempts to solve this problem were made to asymptotic bias corrections by adding a term of order $\mathcal{O}(K/N)$ [30, 31]. This approach was developed further by [27] who made use of a Bayesian prior for the number of relevant bins, and iteratively re-estimated it.

Bayes' theorem states,

$$P(\mathbf{p}|n_i) = \frac{P(n_i|\mathbf{p}) \cdot P(\mathbf{p})}{P(n_i)} ,\qquad(2)$$

where $P(\mathbf{p})$ is the prior distribution and N is the total number of experimental trials, $N = \sum_i^K n_i$, from i observed events which occurred n_i times

In order to eliminate the bias deviation it would be desirable to find a prior $P(\mathbf{p})$ does not depend upon the number of trials. We can express $P(\mathbf{p})$ in terms of the Dirichlet family of priors [29], which allow us to construct a prior which does not depend on inverse powers of N.

$$P_\beta(\mathbf{p}) = \frac{1}{Z} \delta\left(1 - \sum_i^K p_i\right) \left(\prod_i^K p_i^{\beta-1}\right) ,\qquad(3)$$

where Z is a function which enforces the normalization, and the entropy is specified fixing a particular value of β [32]. Ideally, we would like to compute the whole a priori distribution of entropies

$$P_\beta(H) = \int dp_1 \cdots dp_K \, P_\beta(\mathbf{p}) \, \delta\left(H - \sum_i^K p_i \log_2 p_i\right) .\qquad(4)$$

Defining

$$P_\beta(H) = 1 = \int d\xi \, \delta(H - \xi) ,\qquad(5)$$

where ξ is the expected entropy, one can fix a flat prior distribution of entropy $P(H)$. The idea of the NSB method is to build a family of prior $P_\beta(\mathbf{p})$ which result in δ functions over H, by changing β over the whole range of entropies uniformly, removing bias at its source.

The Bayesian prior proposed by Nemenman, Bialek and Steveninck [29],

$$P_\beta(\mathbf{p}) = \frac{1}{Z} \delta\left(1 - \sum_i^K p_i\right) \left(\prod_i^K p_i^{\beta-1} \frac{d\xi(\beta)}{d\beta} P(\beta)\right)\qquad(6)$$

reach the requirements, generating a nearly uniform distribution of entropies in order to avoid bias at its origin (Z is a normalizing factor and $\frac{d\xi(\beta)}{d\beta}$ ensures the uniformity for a priori expected entropy ξ).

We performed calculations of the total mutual information conveyed by a single pair of cells in V1 [24] using the naive estimator, using the Panzeri-Treves bias correction, and using the NSB method. Figure 1 shows the effectiveness of the entropy estimator developed by Nemenman, Bialek and Steveninck [29], for which sampling was obtained at only 100 trials.
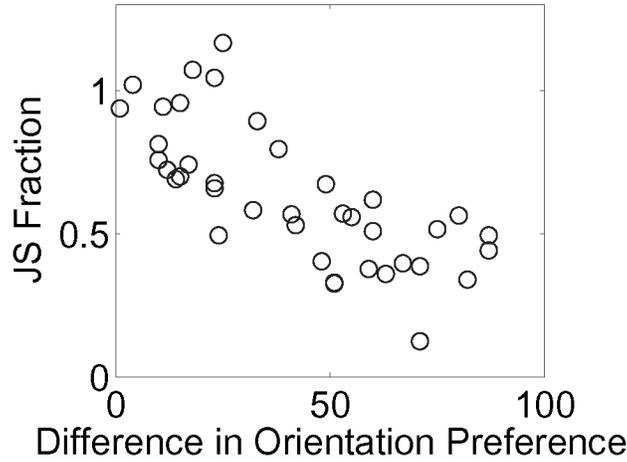
186

FIGURE 3. $T = 5\ ms$, The relative Jensen-Shannon Divergence as a discrimination measure of correlations.

Jensen-Shannon divergence

The Jensen-Shannon divergence (JSD) is defined by [33, 34],

$$JS(\mathrm{P},\mathrm{Q}) = \frac{1}{2}\left[D_{KL}\left(\mathrm{P}\Big|\frac{\mathrm{P}+\mathrm{Q}}{2}\right) + D_{KL}\left(\mathrm{Q}\Big|\frac{\mathrm{P}+\mathrm{Q}}{2}\right) \right], \qquad (7)$$

where D_{KL} is the Kullback-Leibler divergence between the two distributions.

The JSD allow us to decide how reliably one can decide if a given response comes from two different distributions. However, unlike the Kullback-Leibler divergence, it is symmetric, always well defined and bounded. In particular if we choose $\mathrm{P} \equiv P(r|s)$, the probability of getting response vector r conditional upon the occurrence of stimulus s, and $\mathrm{Q} \equiv P_{ind}(r|s)$, the probability of getting independent population responses,

$$
\begin{aligned}
JS(\mathrm{P},\mathrm{Q}) =\ & \frac{1}{2}\sum_{r_1,r_2}\sum_s P(s)\,P(r_1 r_2|s)\,\log_2 P(r_1 r_2|s) - \frac{1}{2}\sum_{r_1,r_2} P(r_1 r_2)\,\log_2 P(r_1 r_2) \\
& + \frac{1}{2}\sum_{r_1,r_2}\sum_s P(s)\,P_{ind}(r_1 r_2|s)\,\log_2 P_{ind}(r_1 r_2|s) - \frac{1}{2}\sum_{r_1,r_2} P_{ind}(r_1 r_2)\,\log_2 P_{ind}(r_1 r_2) \\
& + \sum_{r_1,r_2}\sum_s P(s)\left[\frac{P_{ind}(r_1 r_2|s)+P(r_1 r_2|s)}{2}\right]\log_2\left[\frac{P_{ind}(r_1 r_2|s)+P(r_1 r_2|s)}{2}\right] \\
& - \sum_{r_1,r_2}\left[\frac{P_{ind}(r_1 r_2)+P(r_1 r_2)}{2}\right]\log_2\left[\frac{P_{ind}(r_1 r_2)+P(r_1 r_2)}{2}\right] \qquad (8)
\end{aligned}
$$

where $P(r_1 r_2) = \sum_s P(s)P(r_1 r_2|s)$ and $P_{ind}(r) = \sum_s P(s)P(r_1|s)P(r_2|s)$ are the average of $P(r|s)$ and $P_{ind}(r|s)$ over all the possible stimuli. By fixing the stimuli we can use the Jensen-Shannon divergence to measure how reliably one can decide if a given response comes from $P(r_1 r_2|s_{fix})$, or , $P_{ind}(r_1 r_2|s_{fix})$. Comparing the JS for pairs of correlated $P(r|s)$ and uncorrelated $P_{ind}(r|s)$ cells at a fixed stimuli, one can ask how much information you can gain about the effective network between cells.

RESULTS

In order to evaluate how reliably one can decide if a given response on the population coding of sensory information comes from the full distribution, $P(r_1 r_2|s_{fix})$, or from the product of independent distributions from each cell,
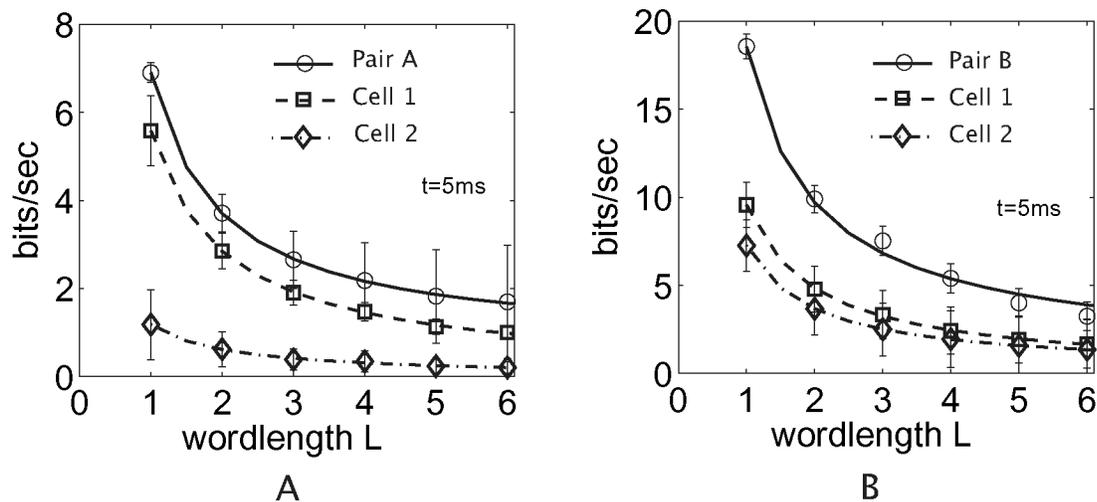
187

FIGURE 4. Calculation of the Information rates. The Figure show the Information rate for two different pais (Pair A and B) and for individual neurons, binwidths $\Delta t = 5ms$

$P_{ind}(r_1 r_2 | s_{fix})$, we used recorded responses of 42 pairs of single neurons in primary visual cortex of macaque monkey (V1) to stimuli of varying orientation. The experimental methods can be found in [24]. They collected data from single pairs of cells in primary visual cortex. For each neuron, they first determined the preferred orientation and direction. This was done quantitatively by measuring the responses to sinusoidal gratings drifting in different directions, centered on the receptive field as determined by initial mapping [24].

Figure 2 shows the Jensen-Shannon divergence for one single pair of cells in V1 [24] using the NSB entropy estimator [29] at different time windows. The Jensen Shannon divergence was calculated taking the full orientation dependency on both probabilities distributions. Plateau like behavior can be appreciated in the Jensen-Shannon as time window becomes bigger.

In Figure 3 we present the Jensen-Shannon Fraction for 42 pairs of single neurons V1 at fixed stimuli orientation as function of the difference in orientation preferences from the tuning curves obtained by each pair of cells (time window fixed at 5 ms). We define the Jensen-Shannon Fraction as the maximum JS divergence obtained by fixing the stimuli for a given set divided by its minimum for a pair of single neurons. Pairs of cells with similar preferred orientations present a higher JS divergence which those that showed less similarities. This result is agreement with the fact that cells which share similar preferred orientation are more likely to have a higher quantity of common inputs than those that present less similarities. Our finding indicates therefore that the Jensen-Shannon Divergence may be used for characterizing the effective circuitry network in a population of neurons.

In Figure 4, we show the calculation of the information rate for a typical pairs of cells, by plotting the information at wordlength L divided by the total time window considered. The asymptote defines the information rate ([35]). The figure compares information versus L behaviour for binwidths of 5 ms, as used in the manuscript ; note that it was not feasible to go beyond $L = 6$ (with pairs of cells, each of which can fire many spikes in an individual time bin), as the dimensionality explodes rapidly. Our use of 5ms is clear: coarser temporal binning leads to a large decrease in the information rate achieved (see [36] for a review). Hence, we use here a 5 ms time windows as a relevant timescale for

188

sensory processing.

CONCLUSION

We performed NSB method on a relative Jensen-Shannon Divergence for pairs of cells in V1 with stimuli dependence in orientation, considering the relevant timescale for sensory information. Our findings show that pairs of cells in V1 with similar preferred orientations present a higher relative JS divergence which those that have less similarities. The use of a relative Jensen-Shannon Divergence at a fixed stimuli provide us a measure of how much information one can gain about the effective network between cells.

ACKNOWLEDGMENTS

REFERENCES

1. K. Toyama, M. Kimura, and K. Tanaka, *J. Neurophysiol.* **46**, 191–201 (1981).
2. K. Toyama, M. Kimura, and K. Tanaka, *J. Neurophysiol.* **46**, 202–214 (1981).
3. A. Michalski, G. L. Gerstein, J. Zarkowska, and R. Tarnecki, *Exp. Brain Res.* **51**, 97–107 (1983).
4. F. Aiple, and J. Krueger, *Exp. Brain Res.* **72**, 141–149 (1988).
5. D. Y. Ts'o, C. D Gilbert, and T. N. Weisel, *J. Neurosci.* **6**, 1160–1170 (1986).
6. D. Y. Ts'o, C. D. Gilbert, *J. Neurosci.* **8**, 1712–1727 (1988).
7. P. M. Gochin, E. K. Miller, C. G. Gross, and G L. Gerstein, *Exp. Brain Res.* **84**, 505–516 (1991).
8. Y. Hata, T. Tsumoto, H. Sato, K. Hagihara, and H. Tamura, *Nature* **335**, 815–817 (1998).
9. J.M. Alonso, and L.M. Martinez, *Nat. Neurosci.* **1**, 395–403 (1998).
10. S. Molotchnikoff, and S. Shumikhina, *Vis. Res.* **36**, 2037–2046 (1998).
11. D. H. Perkel, G. L. Gerstein, and G. P. Moore, *Biophys. J.* **7**, 419–440 (1967).
12. G. L. Gerstein, D. H. Perke, *Science* **164**, 828–30 (1969).
13. G. P. Moore, J. P. Segundo, D. H. Perkel, and H. Levitan, *Biophys. J.* **10**, 876–900 (1970).
14. P. A. Kirkwood, *J. Neurosci. Methods* **1**, 7152–7161 (1979).
15. A. M. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm, *J. Neurophysiol.* **61**, 900–917 (1989).
16. D. J. Surmier, and R.J. Weinberg, *Brain Res.* **331**, 180–184 (1985).
17. D. H. Hubel, and T. N. Weisel, *J. Physiol. (London)* **160**, 106–54 (1962).
18. V. B. Mountcastle, *J. Neurophysiol.* **20**, 408–34 (1957).
19. G. C. De Angelis, G. M. Ghose, I. Ohzawa, R. D. Freeman, *J. Neuroscience* **19**, 4046–4064 (1999).
20. E. Zohary, M. N. Shadlen, and W. T. Newsome, *Nature* **370**, 140–143 (1994).
21. M. N. Shadlen, W. T. Newsome, *J. Neuroscience* **18** , 3870–3896 (1998).
22. F. Montani, A. Kohn, M. Smith and S. R. Schultz, *Neurocomputing* in press (2006).
23. F. Montani, A. Kohn, M. Smith and S. R. Schultz, *J. Neuroscience* in press (2007).
24. A. Kohn, and M. A. Smith, *J. Neurosci.* **25**, 3661–3673 (2005).
25. J. M. Samonds, and A. B. Bonds, *J. Neurophysiol* **93**, 223–236 (2005).
26. E. Schneidman, W. Bialek, and M. J. Berry, *J. Neurosci.* **23(37)**, 11539–11553 (2003).
27. S. Panzeri, and A. Treves, *Network* **7**, 87–107 (1996).
28. S. K. Ma, *J. Statistical Physics* **26**, 221–240 (1981).
29. I. Nemenman, W. Bialek, and R. Steveninck, *Phys. Rev. E* **69**, 056111–6, (2004).
30. A. G. Carlton, *Pyschological Bulletin*, **71**, 108–109 (1969).
31. G. A. Miller, *Information Theory in Psychology; Problems and Methods II-B*, 95–100 (1955).
32. I. Nemenman, F. Shafee, and W. Bialek, "Entropy and inference, revisited", *In Advance Neural Information Processing Systems 14*, edited by T. .G. Dietterich, S. Becker, and Z. Ghahramani, MIT Presions, Cambridge, MA, 2002.
33. P.W. Lamberti, M. T. Martin, A. Plastino and O. A. Rosso, *Physica A* **334**, 119–131 (2004).
34. O. A. Rosso, M. T. Martin, A. Figliola, K. Keller and A. Plastino, *J. Neuroscience Meth.* **153**, 163–182 (2006).
35. S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, *Physical Review Letters* **80**, 197–200 (1998).
36. S. Panzeri, S. R. Schultz, A. Treves, and E. T. Rolls, *Proc. Royal Soc. London B: Biological Sciences* **266**, 1001–1012 (1999).