WORKSHOP PAPER

Speed, noise, information and the graded nature of neuronal responses*

S Panzeri†, G Biella‡, E T Rolls§, W E Skaggs‖ and A Treves†

† SISSA, Biophysics and Cognitive Neuroscience, via Beirut 2, 34013 Trieste, Italy
‡ CNR, Ist Neurosc e Bioimm, Milano, Italy
§ Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK
‖ ARL NSMA, University of Arizona, Tucson, AZ, USA

Abstract. How the firing rate of a neuron carries information depends on the time over which rates are measured. For very short times, the amount of information conveyed depends, in a universal way, on the mean rates only (trial-to-trial variability is irrelevant) and the cell response can be taken to be binary (although an ideal binary response would convey more). For longer times, noise as well as the graded nature of the response come into play, with opposite effects. Which times can be considered 'short' varies with the brain area considered and, possibly, with the processing speed it is required to operate at.

1. Introduction

The rate at which it emits spikes is, for a neuron, an important way of coding information, even though more complex codes (involving, e.g., the temporal structure of the train of spikes, or the degree of synchrony with the spiking of other cells) might also be relevant in certain situations. One can measure, from single-cell or multiple single-cell recording experiments, the amount of information, in bits, contained in the distribution of firing rates that characterizes the response of a neuron to different sensory perceptions and internal states (here collectively called 'stimuli'). Such measurements are becoming increasingly important, as the study of information processing in the brain begins to include quantitative analyses, and attempts to utilize theoretical models at a quantitative level.

A most prominent correlate of the average amount of information in the firing rate of a neuron is the sparseness of the distribution of mean rates to each of the stimuli [1]. Sparse firing, with only a small fraction of the stimuli evoking substantial responses (an example being place-related firing in the rat hippocampus), carries little information, whereas a more even use of its own firing range allows the neuron to transmit more information (as, e.g., in the monkey temporal visual cortex [2]). The effects of sparseness have indeed been analysed in several theoretical models, for example in associative memory networks, in which sparse coding, while reducing the information content of each memory, increases the number of memories that can be stored [3, 4].

At fixed sparseness, there are at least two more aspects of the firing which are important in determining how informative it is. The first is how variable, or noisy, are responses to

---

the same stimulus. The second is how close the distribution of mean rates is to being binary or bimodal, or conversely how graded is the response of the cell. Both aspects are often neglected in the construction and analysis of theoretical models, making the correspondence between such models and real neurons less direct. Many simple models, and a lot of common sense intuition, are based on noiseless binary variables. Deviations from this idealized case have opposite effects: noise always reduces information transmission, whereas a more graded response enhances it. The net effect depends on the time the neuron's activity is sampled for. We discuss here how to analyse the firing of real cells in these terms, and provide examples from single cells recorded in the rat somatosensory system and in the primate temporal visual cortex.

We note that naive information estimates require, to be accurate, very large samples of data. In practical situations, especially with mammals, in which the size of the data sample available for each cell is often limited, these estimates are both very imprecise and subject to systematic error. We have recently introduced a procedure based on the subtraction from naive estimates of an analytically calculated correction term, which results in nearly unbiased estimates [5, 6]. This procedure has been employed in the analysis that follows.

2. The information in the full rate distribution

The neuronal responses considered here are the rates, r, recorded from a cell in correspondence with a given 'stimulus', s (which, as noted above, can be intended as any combination of sensory, motor or internal correlates of the recording). Stimuli are taken from a *discrete* set $\mathcal{S}$ of S elements, each occurring with probability $P(s)$. Rates are measured simply by counting the number of spikes in a given time window $[t_0, t_0 + t]$, and hence are just positive integers, (except they are divided by t itself). The probability of events with rate r is denoted as $P(r)$, and the *joint* probability distribution as $P(s, r)$. The specific information about each stimulus, gained by knowing the rate response, is, on average across r,

$$i(s, t) = \sum_r P(r|s) \log_2[P(r|s)/P(r)].$$ (1)

The information for each stimulus is already an interesting quantity to extract from a recording. It is important to remember, though, that it is defined only *relative* to the probability distribution of *all* stimuli considered, i.e. $P(s')$ for each s' (the other stimuli enter equation (1) through $P(r)$, which depends on their probabilities). Within such relativity, it can for example quantify [7, 2] how much more information a face selective neuron gives, with its firing rate, about faces than about non-face stimuli.

By further averaging the specific information across stimuli, according to their *a priori* probabilities, one obtains the mutual information

$$I(t) = \sum_{s \in \mathcal{S}} \sum_r P(s, r) \log_2 \frac{P(s, r)}{P(s)P(r)}$$ (2)

between stimuli and rates. It quantifies how much the firing rate of a cell is potentially valuable in discriminating among members of the set $\mathcal{S}$, i.e. it provides a relative, but objective, measure of its 'worth', within that task.

Ideally, one would like to measure $i(s, t)$ and $I(t)$ by directly applying equations (1) and (2). In practice one has available, instead of $P(s, r)$, the frequency table computed on the basis of N events, $P_N(s, r)$. It is easy to see that if $P_N(s, r)$ is brutally inserted in equations (1), (2) in place of $P(s, r)$, information is grossly overestimated. For example,

if the underlying $P(s, r)$ is close enough to a continuous distribution in r, that the probability of having exactly coincident rates in the frequency table vanishes, then both information estimates would be equal to their upper bounds, $i(s, t) = -\log_2 P(s)$ and $I(t) = -\sum_s P(s) \log_2 P(s)$; any cell would then yield full knowledge about any set of stimuli! We have discussed procedures to avoid this problem [6], which if untreated makes information estimates from mammalian recording meaningless. In general, a regularization of the responses is necessary, following which the finite sampling effects can be evaluated and subtracted out. Since the regularization itself causes an information loss that is difficult to quantify, it should be kept minimal to minimize this loss. A simple form of discretization is the binning into R response bins, in which case the correction term to be subtracted depends solely on the number R_s of bins relevant for each stimulus [6]:

$$C_1 = \frac{1}{2N \ln 2} \left[\sum_s R_s - R - (S - 1) \right]. \tag{3}$$

When rates are computed from a time window short enough that the maximal number of spikes recorded is not too high, the responses are already binned, no regularization is necessary, and, provided finite sampling effects can be controlled, one measures in fact the 'true' underlying information. In particular, as the window shrinks to zero, the number of bins eventually reduces to just two (one spike or none), which implies that (a) responses are binary and (b) naive estimates can be easily corrected even with just a few trials per stimulus, subtracting a small C_1 correction.

3. Time derivatives of the information

To study the initial rate at which information accumulates from time t_0, one can also consider directly its time derivatives at t_0, which are hardly affected at all by the limited sampling problem (although a small systematic error can still be calculated with error propagation and subtracted out).

For t small, $i(s, t)$ can be approximated by the Taylor expansion

$$i(s, t) = t i_t(s) + \frac{t^2}{2} i_{tt}(s) + \cdots \tag{4}$$

where $i_t(s), i_{tt}(s)$ are the first two time derivatives of $i(s)$ calculated at t_0. The first derivative is universal, i.e. independent of firing statistics, while the second takes a very simple expression under the assumption that the firing of the cell is purely Poissonian (note that it can be singular, instead, in other cases). To first order in t, and with the Poisson assumption to second order, the probability $P(n|s)$ of emitting n spikes in the time window is determined solely by the mean rate r_s to each stimulus s. To second order in t we have

$$
\begin{aligned}
P(0|s) &\simeq 1 - t r_s + \tfrac{1}{2}(t r_s)^2 \\
P(1|s) &\simeq t r_s (1 - t r_s) \\
P(2|s) &\simeq \tfrac{1}{2}(t r_s)^2 \\
P(n > 2|s) &\simeq 0.
\end{aligned}
\tag{5}
$$

Denoting with $\bar{r} = \sum_s P(s) r_s$ the grand mean rate to all stimuli, and with $a = \bar{r}^2 / \sum_s P(s) r_s^2$ the sparseness [4] of the rate distribution, we get

$$i_t(s) = r_s \log_2 \frac{r_s}{\bar{r}} + \frac{\bar{r} - r_s}{\ln 2} \tag{6}$$

and for Poisson statistics

$$i_{tt}^{\text{Pois}}(s) = r_s^2 \log_2 a + \frac{\bar{r}(2r_s - \bar{r})(1 - a)}{a \ln 2} . \tag{7}$$

It can be easily seen that $i_t(s) \geqslant 0$, while $i_{tt}^{\text{Pois}}(s)$ can be positive or negative (but $I_{tt}^{\text{Pois}} \equiv (\bar{r}^2/a \ln 2)[\ln a + (1 - a)] < 0$, implying that for Poisson statistics the rate of information transmission always slows down after the first spike).

The simple formulae for $i_t(s)$ and $i_{tt}^{\text{Pois}}(s)$ are remarkable because they require a measure of only the mean rates r_s, and not of the full distribution of rates to each stimulus $P(r|s)$. This translates into a clear advantage for measuring information, when data are scarce.

Several interesting relationships should be appreciated. First of all, $i_t(s)$ itself represents, apart from a rescaling by the overall mean rate, a universal U-shaped curve which gives, whatever the rate distribution, the initial speed of information acquisition as a function of the rate. Dividing by the overall mean rate and taking an average across stimuli one has

$$\Phi = \sum_s P(s)\frac{r_s}{\bar{r}} \log_2 \frac{r_s}{\bar{r}} \tag{8}$$

which has the meaning of mean information per spike. It is easy to show that in general, for any distribution of rates,

$$0 < \Phi < \log_2(1/a). \tag{9}$$

and for the distributions of rates that are close to binary, with one of the peaks at zero, $\Phi \approx \log_2(1/a)$, while if they are nearly uniform, or strongly unimodal, $\Phi \ll \log_2(1/a)$.

Extensive recordings of rat hippocampal and neocortical cell activity ([1], and unpublished observations), indicate that Φ and a (or equivalently $\log_2(1/a)$) could be used almost interchangeably to characterize firing rate distributions: one parameter turns out to be an excellent predictor of the other.

In recordings of primate temporal cortical cells, it was often found that a cell firing at close to its grand mean rate carried little or no specific information [7]. The dependence of $i(s, t)$ on r_s for finite t (50–500 ms) closely reproduced that predicted at $t \to 0$, i.e. that associated with the time derivative $i_t(s)$.

We note that when n stimuli are presented each with equal frequency, the information per spike Φ is simply related to the breadth of tuning [8]

$$H = -\frac{1}{\log_2 n} \sum \frac{r}{n\bar{r}} \log_2 \frac{r}{n\bar{r}} \tag{10}$$

which was used to characterize the distribution of responses, e.g. to gustatory stimuli in the monkey [9]. In fact

$$\Phi = (1 - H) \log_2 n \tag{11}$$

and when the cell responds to only one stimulus, $H = 0$ and $\Phi = \log_2 n$ (extreme selectivity), whereas when it responds equally to all stimuli $H = 1$ (broad tuning) and $\Phi = 0$ (no information).

4. Real responses and their idealization

Figure 1 provides examples of the way the firing rate of real cells, when measured over increasing time windows from approximately the onset of the response to static stimuli, conveys information about those stimuli. The information in the actual rates is compared with the information present in binarized responses, and with that available from an ideal binary unit.
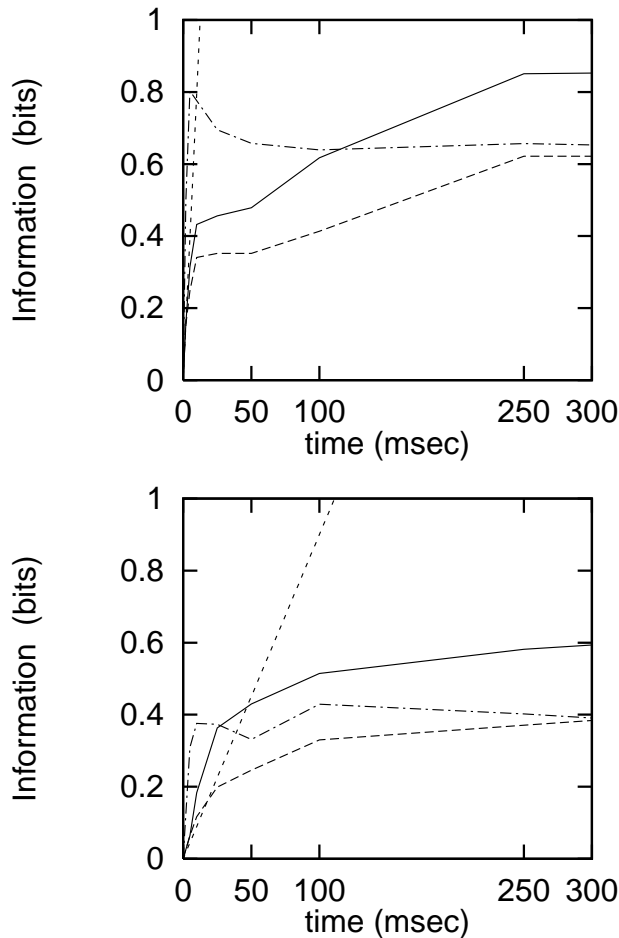
Figure 1. Top: information in the number of spikes emitted in response to four electrical stimuli (——) by a cell in the rat SI cortex, in the binarized responses (– – –) and in the noiseless responses of an ideal binary unit (— · —). The initial slope at t_0 is also indicated (- - - -). Bottom: information in the responses to 20 face stimuli by a cell in the monkey IT cortex, with the same notation.

Responses are binarized by taking as '1' all responses above a certain threshold, and as '0' all others, with the threshold chosen, in each window, to optimize the amount of information transmitted. Note that this binarization preserves at least part of the original trial-to-trial variability, but results in an apparent sparseness different from the true value.

Ideal binary responses are simply those of a unit operating at the same grand mean rate and sparseness as the real unit (in each window), but with zero noise, i.e. mean rates as well as the rates on individual trials are taken to be zero for a fraction $(1 - a)$ of the stimuli, and $\bar{r}/a$ for the remaining fraction a.

In addition, the time derivative I_t is shown, as calculated for actual responses from the shortest window considered (2 ms). For binarized responses the derivative is the same, because for very short windows the responses are already binary; whereas for ideal binary units the derivative is higher, as it is clear from figure 1 and equation (9). Note, though, that for the cell of figure 1 (bottom) the time derivative was almost constant, even when

computed from the distribution of mean rates over longer windows (whereas for that of figure 1 (top) it progressively decreased).

Figure 1 (top) refers to a cell in the rat somatosensory cortex, responding to electrical stimulation of four different intensities. The information in the actual rates rises steeply from t_0 (15 ms post-stimulus onset) and then slows down and saturates when t is of the order of a few hundred milliseconds. Binarized responses convey less informration, but even for long windows only by a factor of about $\frac{3}{4}$. An ideal binary unit with the same sparseness would yield almost instantly all the information it can convey. This levels off around 0.6 bits, which is above the value for the real cell, thanks to the lack of variability, at short times, $t \leqslant 100$ ms; but it is inferior for longer times, when the binary output becomes limiting if contrasted with actual graded output.

Figure 1 (bottom) refers to a cell in the primate visual cortex, responding to 20 face stimuli. t_0 is 100 ms post-stimulus onset, near the peak of the response. The information in the actual rates accelerates almost instantly from a lower initial slope I_t, and slows down only later, having reached values twice those for the binarized responses. In this case the time derivative provides a poorer indication of the information available in the full response. The positive second derivative at t_0 reflects a limited trial-to-trial variability in this cell, much less than for Poisson statistics. This may be related to the operation of recurrent circuits. The ideal binary unit with the same sparseness would again be more informative at short times, but now the time after which the advantages of a graded response take over is shorter, $t \approx 25$ ms.

Analyses of this type are being applied to cells in different brain areas.

References

[1] Skaggs W E and McNaughton B L 1992 Quantification of what it is that hippocampal cell firing encodes *Soc. Neurosci. Abs.* **18** 1216
  Skaggs W E, McNaughton B L, Gothard K and Markus E 1993 An information theoretic approach to deciphering the hippocampal code *Advances in Neural Information Processing Systems* vol 5, ed S J Hanson, J D Cowan and C L Giles (San Mateo, CA: Morgan Kaufmann) pp 1030–7
[2] Rolls E T and Tovee M J 1995 Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex *J. Neurophysiol.* **73** 713–26
[3] Tsodyks M V and Feigel'man M V 1988 The enhanced storage capacity in neural networks with low activity level *Europhys. Lett.* **6** 101–5
[4] Treves A and Rolls E T 1991 What determines the capacity of autoassociative memories in the brain? *Network: Comput. Neural Syst.* **2** 371–397
[5] Treves A and Panzeri S 1995 The upward bias in measures of information derived from limited data samples *Neur. Comput.* **7** 399–407
[6] Panzeri S and Treves A 1996 Analytical estimates of limited sampling biases in different information measures *Network: Comput. Neural Syst.* **7** 87–107
[7] Rolls E T, Tovee M J and Treves A 1995 Information in the neuronal representation of individual stimuli in the primate temporal visual cortex, submitted
[8] Smith D V and Travers J B 1979 A metric for the breadth of tuning of gustatory neurons *Chem. Senses Flavour* **4** 215–9
[9] Rolls E T, Yaxley S and Sienkiewicz 1990 Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey *J. Neurophysiol.* **64** 1055–66