A downward biased estimator of spike timing information

Marcelo A. Montemurro*, Riccardo Senatore, Stefano Panzeri

Faculty of Life Sciences, The University of Manchester, The Mill, P.O. Box 88, Manchester M60 1QD, UK

Available online 11 November 2006

Abstract

We develop a new simple estimator of the spike timing mutual information between a set of static or dynamic stimuli and the elicited spike trains. Unlike the standard direct procedure (which provides upward-biased information estimation), this new method provides a downward biased (DB) estimator. Therefore, by using this new estimator in conjunction with the direct one it is possible to bound from both above and below the true asymptotic value of the mutual information. The downward bias property of the new method is useful in neurophysiological studies of neural codes because a finding of significant extra information in spike timing obtained with this new method will ensure that this additional spike timing information is genuine and not an artefact due to sampling problems.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Information theory; Spike timing; Neural code; Bias

1. Introduction

A fundamental question in Neuroscience is to understand what features of the neural response carry information about external sensory stimuli. In particular, the determination of the temporal precision at which neurons encode information is crucial for constraining the speed and the nature of sensory computation. The most traditional hypothesis about the time structure of neural code is that neurons represent stimuli by the total number of spikes fired over a relatively long post-stimulus time window [1]. However, there is increasing evidence that precise spike times, down to the sub-millisecond time scale, convey substantial extra information that is not available in spike counts alone [7,5].

A rigorous and principled approach to analyze the problem of neural coding is given by Information Theory. Shannon's Mutual information (called Information in the following) quantifies the maximum amount of knowledge that can be obtained about the stimulus by observing the neural response. However, using Information Theory to probe the nervous system to fine time scales and long

observation windows poses a serious technical challenge since applications to experimental data are often hampered by severe sampling problems. Information depends on the stimulus-response probabilities, which in practice must be estimated from a limited number of experimental trials per stimulus, N. This can lead to a significant upward bias in the Information [6]. This is an important problem when comparing spike count and spike timing information because, as we shall see later, information estimates for the spike timing code are invariably much more biased upward than information estimates for the spike count code. Thus, unless very extensive data samples are obtained, it is possible that the spike timing information estimates are artificially inflated upwards, thereby leading to incorrectly support the spike timing hypothesis even in cases when spike timing may not be crucial. To alleviate this problem here we develop a new spike timing Information estimator which is less biased than previously developed ones, and in addition is biased downward rather than upward. This is an important step because a report of significant extra Information in spike timing when using this new downward biased (DB) method would reliably indicate that the additional spike timing information reflects a genuine neuronal behavior and is not an artefact of the analysis.

---

*Corresponding author. Tel.: +44 161 3063883.

E-mail address: m.montemurro@manchester.ac.uk (M.A. Montemurro).

2. A downward biased (DB) estimator of the spike timing information

We consider a time period T, associated with a dynamic or static sensory stimulus s, which is chosen with probability $P(s)$ from a stimulus set $\mathscr{S}$ with S elements. The time window T can be interpreted as the window used by the neuron to encode information (the "encoding window" in the definition given in Ref. [11]), during which the activity of the neuron is recorded. Under the assumption that the shape of the spikes does not convey information, all the information contained in the neural response is preserved by a representation that keeps the precise timing of individual spikes. In practice though, to compute their probabilities, we have to "digitize" spike sequences using a certain time resolution Δt and transform into a sequence of spike counts in each time bin. The parameter Δt represents the imprecision used in registering the spike times and it can be varied parametrically during the analysis to determine the minimal temporal resolution necessary to preserve all the neural information. The quantity L denotes the number of time-bins in the encoding window (i.e. $T = L\Delta t$). The neuronal response is denoted by a one-dimensional array $\mathbf{r} = \{r_1, r_2, \ldots, r_L\}$, where r_t is the number of spikes emitted by the neuron in the tth time-bin. The maximum number of spikes that can be observed in a single time bin in any trial is denoted by M. If Δt is very short, M is 1 and r_t is binary. We indicate the response space with $\mathscr{R}$, which contains $(M + 1)^L$ elements.

The Information transmitted by the neural response about the set of stimuli is defined as [9]

$$I(\mathscr{R}; \mathscr{S}) = H(\mathscr{R}) - H(\mathscr{R}|\mathscr{S}), \tag{1}$$

where $H(\mathscr{R})$ and $H(\mathscr{R}|\mathscr{S})$ are the *response entropy* and the *noise entropy* respectively. They are defined [2] as $H(\mathscr{R}) = -\sum_{\mathbf{r} \in \mathscr{R}} P(\mathbf{r}) \log_2 P(\mathbf{r})$ and $H(\mathscr{R}|\mathscr{S}) = -\sum_{s \in \mathscr{S}} P(s) \sum_{\mathbf{r} \in \mathscr{R}} P(\mathbf{r}|s) \log_2 P(\mathbf{r}|s)$. The summation over $\mathbf{r}$ stands for the sum over all possible neuronal responses. The summation over s indicates a sum over all stimuli s. $P(\mathbf{r}|s)$ is the probability of observing a particular response $\mathbf{r}$ conditional to stimulus s, and $P(\mathbf{r}) = \langle P(\mathbf{r}|s) \rangle_s$ is its average across all stimulus presentations (the angular brackets indicate the average over all possible stimuli). In practice, $P(\mathbf{r}|s)$ is determined experimentally by repeating each stimulus in exactly the same way on many trials, while recording the neuronal responses. The total number of experimental trials across all stimulus conditions is denoted by N. The probability at which stimuli are presented, $P(s)$, is usually chosen by the experimenter.

The length of the encoding window L and the spike timing precision Δt used by the neuron to encode information are not known a priori; however they can be determined empirically by varying systematically these parameters and studying the corresponding changes in information processing. To perform this analysis, it is necessary to be able to measure information up to long encoding time windows (large L) and at fine time scales

(small Δt). In real application to experimental neural data, this is often difficult since there are not enough data available in order to compute accurately the probability distributions, and bias can become important. The leading term in the $1/N$ expansion of the bias is [6]

$$Bias[H(\mathscr{R}|\mathscr{S})] = -\frac{1}{2N \ln(2)} \sum_s (R(s) - 1), \tag{2}$$

where $R(s)$ stands for the number of distinct response words $\mathbf{r}$ with non-zero probability $P(\mathbf{r}|s)$ of being observed in response to stimulus s [6]. In a similar way, we can write the expression for the leading contribution to the bias in the response entropy $H(\mathscr{R})$ as

$$Bias[H(\mathscr{R})] = -\frac{1}{2N \ln(2)}(R - 1), \tag{3}$$

where R stands for the number of distinct response words $\mathbf{r}$ with non-zero probability $P(\mathbf{r})$ of being observed across all stimuli. The resulting bias in the Information will then be given by $Bias[I(\mathscr{R}, \mathscr{S})] = Bias[H(\mathscr{R})] - Bias[H(\mathscr{R}|\mathscr{S})]$. By comparing the expressions in Eqs. (2) and (3) it can be seen that $Bias[H(R)]$ is typically S times smaller than $Bias[H(R|S)]$. Therefore, in situations in which the number of effective stimuli is very large, as e.g. is the case with a continuous dynamic stimulation, then $Bias[H(R|S)] \gg Bias[H(R)]$. Thus, it is possible in those cases to approximate the total bias in the mutual information simply as the positive quantity $-Bias[H(R|S)]$. In what follows, and in order to simplify the discussion, we shall assume that $S \gg 1$ and thus we can neglect the contribution to the bias in the mutual information coming from the response entropy $H(\mathscr{R})$. However, our results would be practically applicable in all cases in which the leading contribution to the bias originates from the noise entropy.

When $\Delta t = T$ ($L = 1$), then the neural response is just determined by the total number of spikes fired during that time. In this case the neural response will be denoted as $\mathscr{R}_n$. Then, the Information computed from the response $\mathscr{R}_n$ will be the spike-count information $I_n(\mathscr{R}_n, \mathscr{S})$. Since now the maximum number of different neural responses is M, then the resulting bias in $I_n(\mathscr{R}_n, \mathscr{S})$ will be typically much smaller than the bias in $I(\mathscr{R}, \mathscr{S})$ for $\Delta t < T$. Then, we can decompose the spike timing mutual information in form that will be useful for our derivation of the DB spike-timing information below, as follows:

$$I^{UB}(\mathscr{R}, \mathscr{S}) = I_n(\mathscr{R}_n, \mathscr{S}) + \Delta I_{timing}, \tag{4}$$

where the superscript indicates that the resulting quantity is biased upwards. The first term on the right-hand side of Eq. (4) represents the contribution to the mutual information coming only from the spike counts within the observation time period T. From the data processing inequality [2] $I(\mathscr{R}, \mathscr{S}) \geqslant I_n(\mathscr{R}_n, \mathscr{S})$, and thus $\Delta I_{timing} \geqslant 0$. Therefore, ΔI_{timing}, which is simply estimated as $I(\mathscr{R}, \mathscr{S}) - I_n(\mathscr{R}_n, \mathscr{S})$, gives the extra information gained by looking at the precise spike times-down to the resolution

limit set by Δt. In this form, almost all the bias in the mutual information is contained within ΔI_{timing}.

A key observation is that $I_n(\mathscr{R}_n, \mathscr{S})$ can also be obtained as a spike timing information with resolution Δt computed on a transformed response $\tilde{\mathscr{R}}$ that has the same spike count information about the stimulus as the true response $\mathscr{R}$, but with zero extra information in spike timing. This can be done by generating a surrogate spike response with zero information in spike timing and with the added constraint of having exactly the same spike count information as the original one. The surrogate response set can be generated simply by randomly shuffling all the spikes across the L time bins independently for each response word. This procedure preserves the number of spikes fired in each time window T, but destroys all the spike timing structure up to a time scale given by T. Therefore, we have $\lim_{N\to\infty} I(\tilde{\mathscr{R}}, \mathscr{S}) = I_n(\mathscr{R}_n, S)$. For finite number of trials $I(\tilde{\mathscr{R}}, \mathscr{S})$ will differ from $I_n(\mathscr{R}_n, S)$ because it will be affected by a larger bias.

We can establish a useful relationship between the bias of $I(\tilde{\mathscr{R}}, \mathscr{S})$ and that of the spike time information computed over the original response $I(\mathscr{R}, \mathscr{S})$. The presence of any structure in the firing patterns will tend to limit the number of possible responses, resulting in $R(s)$ being substantially less than $(M+1)^L$. However, after shuffling the response words in order to compute $I(\tilde{\mathscr{R}}, \mathscr{S})$, the resulting number of distinct words $\tilde{R}(s)$ will not be constrained since all the underlying structure of spike patterns is destroyed. Therefore, we will have that $\tilde{R}(s) \geqslant R(s)$ for each stimulus condition. We have noticed above that almost all the upward bias in $I(\mathscr{R}, \mathscr{S})$ is in ΔI_{timing} in Eq. (4), and it can be very well approximated as the bias in the noise entropy $H(\mathscr{R}, \mathscr{S})$ when $S \gg 1$. Now, it is possible to compute another estimator of ΔI_{timing} that uses $I(\tilde{\mathscr{R}}, \mathscr{S})$ instead of $I_n(\mathscr{R}_n, \mathscr{S})$, defined as follows:

$$\tilde{\Delta} I_{\text{timing}} = I(\mathscr{R}, \mathscr{S}) - I(\tilde{\mathscr{R}}, \mathscr{S}). \tag{5}$$

In the infinite trials limit both $\tilde{\Delta} I_{\text{timing}}$ and ΔI_{timing} will coincide. However, for finite N, the bias in $\tilde{\Delta} I_{\text{timing}}$ will be negative, that is opposite in sign to the bias in ΔI_{timing}. This can be easily seen by noticing that the resulting bias in $\tilde{\Delta} I_{\text{timing}}$ will be given by the difference in biases between $I(\mathscr{R}, \mathscr{S})$ and $I(\tilde{\mathscr{R}}, \mathscr{S})$. Again, for the case in which $S \gg 1$ this bias can be written up to first order in $1/N$ simply as

$$Bias[\tilde{\Delta} I_{\text{timing}}] = \frac{1}{2N \ln(2)} \sum_s (R(s) - \tilde{R}(s)) \tag{6}$$

which will be negative since $\tilde{R}(s) \geqslant R(s)$.[1] With this we can write an expression for a DB estimator of the mutual information as follows:

$$I^{\text{DB}}(\mathscr{R}, \mathscr{S}) = I_n(\mathscr{R}_n, \mathscr{S}) + \tilde{\Delta} I_{\text{timing}}. \tag{7}$$

---

[1] There are different types of resampling methods that have be used to generate surrogate data sets for the validation of information theoretic calculations [4]. It is important to note that the property of the negativity of the bias holds for the resampling procedure "without replacement" used here to generate the surrogate data, but it does not necessarily hold for other resampling techniques "with replacement".

In order to illustrate and validate the new estimator and to compare it with the direct estimation of the information used in Ref. [8], we applied it to synthetic spike trains in response to a dynamic stimulus simulated as follows. We generated a post-stimulus time histogram (PSTH) as a succession of randomly distributed peaks of typical width a. Then we used the resulting PSTH to generate spike trains as an inhomogeneous Poisson process. The width a of the peaks in the PSTH determines the underlying precision of the spike timing code. In the simulations shown in this summary $a = 1$ ms.

In Fig. 1A–B we show the mean and standard deviation of the estimators $I^{\text{UB}}(\mathscr{R}, \mathscr{S})$ and $I^{\text{DB}}(\mathscr{R}, \mathscr{S})$ as a function of the number of repetitions of the stimulus (trials). The estimator $I^{\text{UB}}(\mathscr{R}, \mathscr{S})$ was obtained using a direct method [8,10] and the bias was reduced by the extrapolation method used in Ref. [10]. The total duration of the simulated spike trains was of 15 s, and the spike times were binned with a time resolution of $\Delta t = 5$ ms and $L = 10$. In panel A we show the mean of the information estimates evaluated across 100 simulations of the spike trains. We see that $I^{\text{DB}}(\mathscr{R}, \mathscr{S})$ shows strong bias for low number of trials. However, as mentioned above, it converges to the same value as $I^{\text{UB}}(\mathscr{R}, \mathscr{S})$, as the number of trials grows. The asymptotic value of the mutual information is denoted by the dot line. The plots show that the joint use of both the upward and downward biased estimators can bracket the true value of the mutual information. Moreover, the estimation given by $I^{\text{DB}}(\mathscr{R}, \mathscr{S})$ is tighter and it yields an accurate estimation of the mutual information for 128–256 trials, whilst $I^{\text{UB}}(\mathscr{R}, \mathscr{S})$ only reaches a similar accuracy for 1024 trials. In panel B we show the standard deviation of $I^{\text{UB}}(\mathscr{R}, \mathscr{S})$ and $I^{\text{DB}}(\mathscr{R}, \mathscr{S})$. For all number of trials $I^{\text{DB}}(\mathscr{R}, \mathscr{S})$ shows a standard deviation which is very close to that of $I^{\text{UB}}(\mathscr{R}, \mathscr{S})$. The low values of the standard deviation were due to the correlated variability between the two information measures that defining $\tilde{\Delta} I_{\text{timing}}$ in Eq. (5).

In Fig. 2 we report the comparison of the results obtained with our new estimator and the standard direct one [10] at different values of the spike timing time resolutions Δt, ranging from 1 to 10 ms. In this analysis, data were simulated as in Fig. 1; the information was always estimated using a sliding time window of length $T = 10$ ms; the number of simulated trials was 128; and the Information values correspond to an average over 100 realization of the spike trains. Fig. 2 shows that the upper and lower bounds provide a tight estimation of the total information over the whole range of Δt analyzed, the bounds becoming less tight at very high resolution. The joint use of I^{DB} and I^{UB} leaves no doubt that there is a genuine increase of the transmitted information when using a timing precision as fine as 1 ms, correctly reflecting the procedure used to simulate the data.

3. Discussion

In summary, we have developed a simple estimator $I^{\text{DB}}(\mathscr{R}, \mathscr{S})$ of the Information that is easy to compute, less
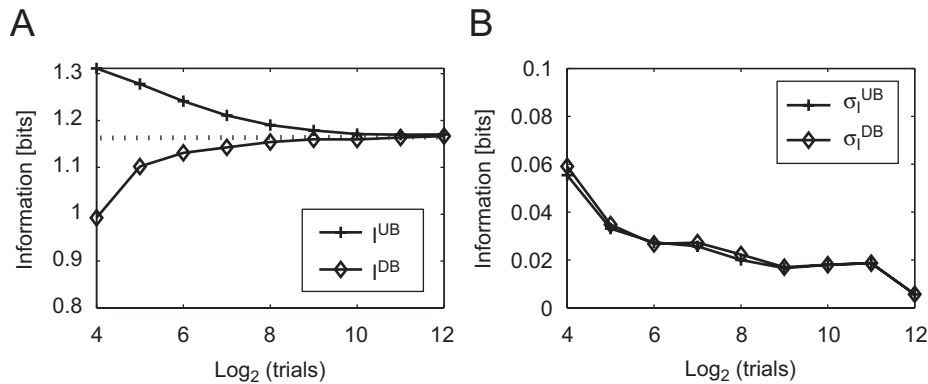
Fig. 1. Information estimates and their standard deviations for simulated data. We simulated spike trains in response to a fast time varying stimulus. First we generated a post-stimulus time histogram (PSTH) as a succession of randomly distributed peaks of typical width $a = 1$ ms. Then we used the resulting PSTH to generate spike trains as an inhomogeneous Poisson process. Finally, for the information analyses the spike times were binned with a time precision of $\Delta t = 5$ ms in $L = 10$ bin words. All the information quantities shown were corrected with the method described in [10]: (A) direct estimation of the mutual information $I^{\mathrm{UB}}(\mathcal{R}, \mathcal{S})$ and the downward bias estimator $I^{\mathrm{DB}}(\mathcal{R}, \mathcal{S})$ as a function of the number of trials. The flat dot line indicates the asymptotic value of the mutual information; (B) standard deviations of $I^{\mathrm{UB}}(\mathcal{R}, \mathcal{S})$ and $I^{\mathrm{DB}}(\mathcal{R}, \mathcal{S})$ as a function of the number of repetitions of the stimulus.
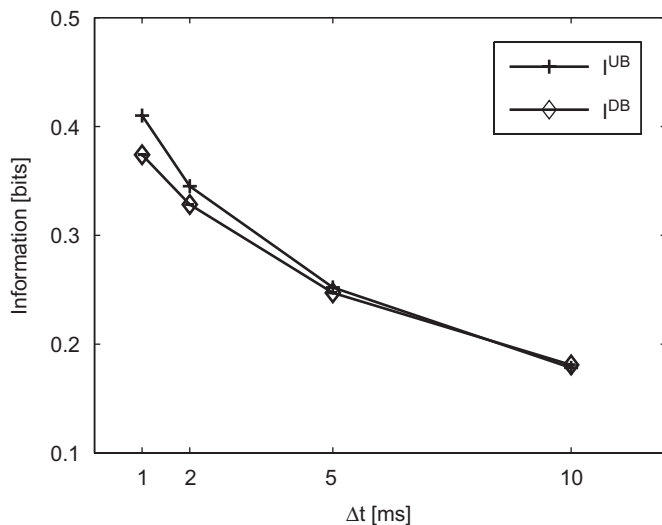

Fig. 2. Upper and lower bounds to the information conveyed at different timing resolutions. We simulated neural responses to a dynamic stimulus as in Fig. 1. We then estimated mutual information on the data binned at different time resolutions. In particular, the binning size, Δt, was 1, 2, 5 and 10 ms, and the information was always estimated using a sliding time window of length $T = 10$ ms. The number of trials was 128, and the Information values were averaged over 100 realizations of the simulation. The full line with crosses shows the upper biased estimator obtained with the method described in [10], and the line with diamonds corresponds to the new downward biased estimator. The joint use of both estimators allows to bound precisely the true value of the mutual information and confirms that the simulated spike trains transmit information at the ms time scale.

biased than the direct estimator of Information, and moreover, is biased downward. Like the direct estimator, our new technique can compute the information in the full encoding window, including the contribution of all possible patterns of correlations between the spike times emitted in the encoding window. Therefore, this method complements other approach, such as e.g. that of [3], designed to capture the information contained is specific aspects of the response such as its time dependent firing rate. Although our work was developed here to be used with mutual information quantities, this approach can be extended to be used with other quantities of interest in neural coding, such has the Kullback–Leibler distances described in [4], which are not as dependent on the choice of the stimulus set as mutual information is.

The DB Information estimator is useful because, when used in conjunction with the standard direct estimator of Ref. [8], it can be used to bound both from above and from below the asymptotic value of the mutual information. This can validate and confirm the accuracy of experimental Information estimates. The downward bias property of $I^{\mathrm{DB}}(\mathcal{R}, \mathcal{S})$ is also useful in practical studies of neural codes because a finding of significant extra Information in spike timing obtained with this new method will ensure that this additional spike timing information is genuine and not an artefact due to sampling problems.

References

[1] E.D. Adrian, The impulses produced by sensory nerve endings: Part I, J. Physiol. (Lond.) 61 (1926) 49–72.

[2] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.

[3] A. Hsu, A. Borst, F.E. Theunissen, Quantifying variability in neural responses and its application for the validation of model predictions, Network 15 (2004) 91–109.

[4] D.H. Johnson, C.M. Gruner, K. Baggerly, C. Seshagiri, Information-theoretic analysis of neural coding, J. Comput. Neurosci. 10 (2001) 47–69.

[5] S. Panzeri, R.S. Petersen, S.R. Schultz, M. Lebedev, M.E. Diamond, The role of spike timing in the coding of stimulus location in rat somatosensory cortex, Neuron 29 (2001) 769–777.

[6] S. Panzeri, A. Treves, Analytical estimates of limited sampling biases in different information measures, Network 7 (1996) 87–107.

[7] P. Reinagel, R.C. Reid, Temporal coding of visual information in the thalamus, J. Neurosci. 20 (2001) 5392–5400.

[8] R. de Ruyter van Steveninck, G.D. Lewen, S.P. Strong, R. Koberle, W. Bialek, Reproducibility and variability in neural spike trains, Science 275 (1998) 1805–1808.

[9] C.E. Shannon, A mathematical theory of communication, AT&T Bell Labs. Tech. J. 27 (1948) 379–423.

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

[11] F. Theunissen, J.P. Miller, Temporal encoding in the nervous system: a rigorous definition, J. Comput. Neurosci. 2 (1995) 149–162.

Marcelo A. Montemurro received a Ph.D. in Theoretical Physics from the Universidad Nacional de Cordoba, Argentina. He is currently an HFSP Research Fellow at the University of Manchester. His main research interests are the theoretical modelling of information processing at cortical level, and the development of data analysis methods to accurately estimate the contribution of different possible mechanisms to the neural code.

Riccardo Senatore received a Laurea in Theoretical Physics from Turin University, Italy, with a thesis in the area of Statistical Quantum Mechanics. He is currently a Ph.D. student in Computational Neuroscience at the University of Manchester. His main research focuses on developing efficient data analysis methods to analyze how sensory systems encode information.

Stefano Panzeri received a Laurea in Physics from Turin University and a Ph.D. in Computational Neuroscience from SISSA, Trieste. He is a Senior Lecturer at the University of Manchester, where he directs the Laboratory of Neuroinformatics. His research focuses on understanding the computational principles underlying cortical information processing.