Characterizing spike trains with Lempel-Ziv complexity

J. Szczepański[1], J.M. Amigó[2], E. Wajnryb[2], M.V. Sanchez-Vives[3]

[1]Institute of Fundamental Technological Research, Warsaw (Poland)
[2]Operations Research Centre, Miguel Hernández University, Elche (Spain)
[3]Instituto de Neurociencias, UNH-CSIC, San Juan de Alicante 03550 (Spain)

Abstract

We present several applications of Lempel-Ziv complexity to the characterization of neural responses. In particular, Lempel-Ziv complexity allows to estimate the entropy of binned spike trains in an alternative way to the usual direct method based on the relative frequencies of words, with the definitive advantage of no requiring very long registers. We also apply this concept to determine generic values of encoding parameters and to evaluate the number of states of neuronal sources.

1. Introduction

Information theory has often been applied to the analysis of the information content in neuronal spike trains. The concept of complexity provides, in most of its many variants, a quantitative assessment of the structure of symbol sequences and, thus, is also related to the information content of such sequences. One of them, Lempel-Ziv complexity (1976), measures the generation rate of new patterns along a digital sequence and, furthermore, is closely related to such important information-theoretic properties as entropy, compression ratio and redundancy. In our group we have been studying for some time (J.M. Amigó *et al.*, 2001) the characterization of spike trains by means of Lempel-Ziv complexity and other related properties. One of the main interests of this method is that complexity is a fast convergent estimator of the entropy of digital signals (as can be shown by numerical simulation), what speaks for its use especially in those cases in which the registers are too short or the non-stationarity of the source precludes long sequences from being considered.

We will present in this communication three applications of Lempel-Ziv complexity (not to be mistaken for the complexity measure of the same name used for lossless data compression, which is posterior in time) to the analysis of spike trains: (*i*) estimation of the entropy, (*ii*) discrimination of neural responses via complexity curves and (*iii*) discrimination of neural responses via the number of states of the corresponding neuronal sources. Here we have applied this method for the analysis of spike trains intracellularly recorded in the *in vivo* and *in vitro* visual cortex as a response to different stimuli: sinusoidal current injection, visual stimulation and random curent injection. Let us just mention at this point that both complexity curves and the number of states of the corresponding neuronal sources not only discriminate neural responses to different stimuli but also allow to conclude, for instance, that neural discharges *in vivo* carry more information than *in vitro* for the same kind of stimuli.

2. Lempel-Ziv complexity and entropy

Let S be a source that generates words $x_1^n := x_1 x_2 ... x_n$ of length n whose letters x_i ($1 \leq i \leq n$) belong to a set $\mathcal{A} = \{a_1, ..., a_\alpha\}$ of size $|\mathcal{A}| = \alpha < \infty$, called the source *alphabet*. Given the

word x_1^n, a *block* of length l ($1 \le l \le n$) is just a segment of x_1^n of length l. Intuitively speaking (rather than giving a rigorous definition), the *complexity* of the word x_1^n, $C_\alpha(x_1^n)$, counts the number of different patterns it contains. The first symbol on the left of the word defines the first block. From there one moves rightward letter by letter, until the string of symbols beginning just after the previous block and ending at the current position happens not to have appeared before. At this point, a new block is defined. The procedure is illustrated by the following example. The decomposition of the binary word $x_1^{19} = 0101101010001101110010$ into minimal blocks of new patterns is

$$0|\ 1|\ 011|\ 0100|\ 011011|\ 1001|\ 0$$

where the vertical lines separate the blocks. Therefore, the complexity of x_1^{19} is 7.

The generation rate of new patterns along x_1^n, a word of length n with letters from an alphabet of size α, is measured by the *normalized complexity* $c_\alpha(x_1^n)$, which is defined by

$$c_\alpha(x_1^n) = \frac{C_\alpha(x_1^n)}{n/\log_\alpha n} = \frac{C_\alpha(x_1^n)}{n} \log_\alpha n \tag{1}$$

Let $H_b(S)$ denote the entropy of the source S,

$$H_b(S) = -\frac{1}{n} \sum_{x_1^n} p(x_1^n) \log_b p(x_1^n)$$

where $p(x_1^n)$ denotes the probability for the word x_1^n to happen and the sum is over all words of length n (α^n in total, though some of them could have zero probability). The subscript b stands for any real constant $b > 1$ and refers to the base of the logarithm. If $b = 2$, the entropy is measured in bits per second (bit/sec). If words can be arbitrarily long, one has to let n go to infinity:

$$H_b(S) = -\lim_{n \to \infty} \frac{1}{n} \sum_{x_1^n} p(x_1^n) \log_b p(x_1^n)$$

provided the limit exists. One can prove that (i) if S is stationary, then

$$\limsup_{n \to \infty} c_\alpha(x_1^n) \le H_\alpha(S) \quad \text{on average} \tag{2}$$

and, moreover, (ii) if S is ergodic, then

$$\limsup_{n \to \infty} c_\alpha(x_1^n) = H_\alpha(S) \quad \text{almost surely} \tag{3}$$

Eqs. (2) and (3) provide ways to bound from below and estimate, respectively, the entropy of a neuronal source (with the corresponding properties) via the Lempel-Ziv complexity of (i) a sample of spike trains, (ii) a typical (*i.e.* randomly chosen) spike train produced by it. It is worth mentioning that the first application of this concept was suggested in cryptography, namely, to assess the quality of random sources.

3. Experimental work

We have studied the complexity of real spike trains. The experimental data were obtained from primary cortex recordings both *in vivo* and in brain slice preparations (*in vitro*). Intracellular recordings *in vivo* were obtained from anesthetized adult cats. For the preparation

of slices, 2-4 month old ferrets of either sex were used. Action potentials were detected with a window discriminator and the time of their occurrence was collected with a 10 μsec resolution. For more details, see (Sanchez-Vives *et al.*, 2000a y b) The resulting time series were used to analyze the neuronal spiking. Concerning the stimuli, they were of three kinds.

a) *Intracellular periodic current injection.* Intracellular sinusoidal currents were injected *in vivo* and *in vitro*. The frequency of the waveform was 2 Hz and the intensity ranged between 0.2 and 1.5 nA. The cell recording ensemble comprised of 8 samples *in vivo* (spike train lengths between 15.56 sec and 47.64 sec) and 8 samples *in vitro* (spike train lengths between 15.87 sec and 23.62 sec).

b) *Visual stimulation with sinusoidal drifting gratings.* The visual stimulus consisted of a sinusoidal drifting grating presented in a circular patch of 3-5 degrees diameter, centered on the receptive field center (*in vivo*). Only simple cells were included in this study. In this case, 8 samples were recorded (spike train lengths between 15.87 sec and 23.62 sec).

c) *Intracellular random current injection.* Random currents with different degrees of correlations were injected during the intracellular recordings from cortical brain slices (*in vitro*). The ensemble consisted of 20 samples (spike train lengths between 16.32 sec and 35.47 sec).

4. Codings

Spike trains can be encoded (or "binned") in many ways. We will consider henceforth only two methods.

a) *Interspike time coding.* Let $\tau_{\min}$ and $\tau_{\max}$ be the minimal and maximal *interspike* times, respectively, in the signal. Divide the interval $[\tau_{\min}, \tau_{\max}]$ into α slots $\Delta\tau_i$ ($1 \leq i \leq \alpha$) of the same length. If τ_j is the interspike time following spike s_j and τ_j belongs to, say, the k-th slot $\Delta\tau_k$, then assign to the spike s_j the k-symbol a_k from a set $A = \{a_1, ..., a_\alpha\}$ of α symbols. In this way, we get an α-nary message whose length equals the number of spikes.

b) *Time bin coding.* Let the first spike of a train occur at time 0 and the last one T time units later. The time interval $[0, T]$ is then split in n bins Δt_i ($1 \leq i \leq n$) of the same length. If there are N_k spikes in the bin Δt_k, then assign the number N_k to Δt_k. The result is a message of length n with no more than n different letters. If, instead, each bin Δt_i is coded by 0 or 1 according to whether it contains no spike (0) or at least one spike (1), the message will be binary.

5. Applications of Lempel-Ziv complexity

5.1. Estimation of the entropy

The spike trains recorded in the lab are digitalized with *binary time bin coding*. Let $\tilde{p}_i$ be the normalized count of the ith word in the ensemble of words of length l ($= L/\Delta\tau$, $L \leq T$) in a set of observations. Then, the estimation of the entropy rate per second

$$H(\Delta\tau) := \lim_{l\to\infty} H(l, \Delta\tau) = -\lim_{l\to\infty} \frac{1}{l\Delta\tau} \sum \tilde{p}_i \log_2 \tilde{p}_i \qquad (4)$$

and, hence, of the source entropy $H(S) := \lim_{\Delta\tau\to 0} H(\Delta\tau)$, requires words of increasing length l whereas real spike trains are necessarily finite. Now, increasing l when counting different words from spike trains of finite length depletes the word statistics and, therefore, renders the estimations of the relative frequencies $\tilde{p}_i$ of the words of length l less and less reliable. Also, as a result of this statistical depletion (or *undersampling*), $H(l, \Delta\tau)$ gets

3

artificially smaller than $H(\Delta\tau)$ for sufficiently big l's, while $H(1, \Delta\tau) \geq H(2, \Delta\tau) \geq ... \geq H(\Delta\tau)$ should hold.

Consequently, given an ensemble of finite spike trains and a time resolution $\Delta\tau$, rather than letting l become as large as possible, there must be some optimal word length l^* for which the normalized complexity is a good estimator of the entropy. Such an l^* results from a trade-off between the goodness of the approximation $H(l^*, \Delta\tau) \approx H(\Delta\tau)$ and the representativeness of the word statistics for that window length. We propose next a complexity-based method to choose a window length l^* which qualifies for optimality in the sense just explained.

Consider an ensemble of (sufficiently short) spike trains (as to be viewed as) generated by a stationary neuronal source. The ensembles we are thinking of are defined below and consist of neuronal responses to similar stimuli. We calculate then the average normalized complexity $\overline{c}(\Delta\tau)$ of the corresponding binned spike trains as a function of the bin width $\Delta\tau$ and then determined the value of $\Delta\tau$ (call it $\Delta\tau^*$) for which the relative standard deviation of the normalized complexity is minimal. The rationale for this approach is that the neuron responses to similar stimuli are also similar and, therefore, the same must happen to the average information transfer (and, *a fortiori*, to the normalized complexity and to the entropy). Under the proviso that the neuronal source can be considered ergodic, we evaluate now the entropy rate estimate (4) for $\Delta\tau^*$ and fix $l = l^*$ so that $H(l^*, \Delta\tau^*)$ coincides with $\overline{c}(\Delta\tau^*)/\Delta\tau^*$. In doing so, we are replacing $\limsup_{l\to\infty} \overline{c(x_1^l)}/\Delta\tau^*$ by $\overline{c(x_1^{l^*})}/\Delta\tau^*$ in (3).

5.2. Complexity curves

Let x_1^n be the result of encoding a spike train recorded in any of the four experimental settings considered above. In order to gain more insight into the complexity of neuronal responses, we have graphically represented $c_\alpha(x_1^n)$ as a function of the number of letters α (for interspike time coding) and also as a function of the word length n (for time bin coding). Remember that n is fixed (and equal to the number of spikes) for the interspike time coding and $\alpha = 2$ for the binary interval coding while, for the general interval coding, there is a weak dependence of α on n which shows up in the graphs as instabilities. For this reason we limit the discussion of the complexity curves to the interspike time coding and the binary time bin coding. This graphical analysis was repeated with spike trains covering all cases.

5.3. Neuronal sources and number of states

Once a spike train has been codified into a message, this can be viewed as emitted by an information source, the source comprising everything preceding the message, namely, the stimulus (S), the neuron or neuronal network (N) and, last but not least, the encoding technique (E). Consider now an ergodic source with transition probability $p(x_i|x_{-\infty}^{i-1})$, where $x_{-\infty}^n := ...x_{n-1}x_n$. We say that the source is Markovian of *finite order* if

$$p(x_i|x_{-\infty}^{i-1}) = p(x_i|x_{i-k}^{i-1}), \quad i = 1, 2, ..., n \tag{5}$$

for some integer $k \geq 1$ called the *order* of the source. Eq. (5) means that the probability for the letter x_i at instant i depends directly only on the previous k outcomes: $x_{i-1}, ..., x_{i-k}$. For this reason, k is also called the *lag* of the source.

Next let $H(q_{\mathbf{x}}^k)$ be the *k-th order empirical entropy* of an ergodic Markov process (J.M. Amigó *et al.*, 2003) as measured at the word $\mathbf{x} = x_1^n$. Following Ziv, an *order estimator* is given by

$$k^* = \min\left\{k : H(q_{\mathbf{x}}^k) - c_\alpha(\mathbf{x})\log_2\alpha \leq \lambda\right\} \tag{6}$$

where $c_\alpha(\mathbf{x})$ is the normalized complexity of $\mathbf{x}$ (1). In the calculations we set $\lambda = 0.02$.

Although these numbers are the order estimators of the corresponding sources assumed to be ergodic and finite-order Markovian, we refer to them with the more neutral term of "number of states" because our primary purpose is to use them to distinguish neuronal sources from each other, independently of the concrete meaning of k and the source properties. Important for us is that k^* is a numerical invariant for neuronal sources. In particular, for a given neuron preparation and coding, k^* depends only on the kind of stimulus (i.e. on the experimental subcase considered out of the four experimental cases given in Section 3), but not on individual stimuli. This comes as no surprise since the same is true for the normalized complexity.

6. Results

6.1. Estimation of the entropy

Within each of the experimental cases, we have determined the time resolution $\Delta\tau^*$ for which (according to its definition) the relative standard deviation of the normalized complexity $c = c(\Delta\tau)$ was minimal. The numerical results are shown in Table below, where $\overline{c}^* = \overline{c}(\Delta\tau^*)$. Notice that we have further split the responses to random stimuli into two subsets, according to whether the autocorrelation function of the stimuli decays slowly or fast. Only those records with random stimuli for which this distinction was clear, were considered for evaluation.

	In vivo	In vitro
Periodic current	$\Delta\tau^* = 0.025$ sec, $\overline{c}^* = 0.57$ $\Rightarrow \overline{c}^*/\Delta\tau^* = 22.8$ sec^{-1}	$\Delta\tau^* = 0.025$ sec, $\overline{c}^* = 0.20$ $\Rightarrow \overline{c}^*/\Delta\tau^* = 8.0$ sec^{-1}
Visual stimulus	$\Delta\tau^* = 0.100$ sec, $\overline{c}^* = 0.90$ $\Rightarrow \overline{c}^*/\Delta\tau^* = 9.0$ sec^{-1}	
Random current slow decay		$\Delta\tau^* = 0.100$ sec, $\overline{c}^* = 0.78$ $\Rightarrow \overline{c}^*/\Delta\tau^* = 7.8$ sec^{-1}
Random current fast decay		$\Delta\tau^* = 0.100$ sec, $\overline{c}^* = 0.87$ $\Rightarrow \overline{c}^*/\Delta\tau^* = 8.7$ sec^{-1}

6.2. Complexity curves

The results obtained can be summarized as follows.

a) For *interspike time coding*, the curves $c_\alpha(x_1^n)$ vs. α are convex $\cap$, converging sharply with increasing α to flat horizontal profiles. We call these stationary values *saturation levels* and they are about the same for periodic stimuli (*in vivo* and *in vitro*). The saturation levels corresponding to random inputs are more scattered due to their manifold autocorrelation functions but, as one expects, they do not overlap with the non-random ones. The saturation levels were the following. (i) *Periodic current injection in vivo*: The complexity curve increases sharply to its horizontal saturation level with small fluctuations. Saturation levels range from 0.50 to 0.03. (ii) *Periodic visual stimulation in vivo*: Saturation occurs typically at the level 0.50 ± 0.03. (iii) *Periodic current injection in vitro*: Saturation sets in around $\alpha = 600$ at a level $c_\alpha \approx 0.40 \pm 0.03$. (iv) *Random current injection in vitro*: The saturation level changes with the input signal, which hints to a relation between the asymptotic complexity values and the rate of decay of the autocorrelation function of the corresponding stimulus. Typical saturation levels for long correlations are 0.22 ± 0.03 and for short correlations are 0.12 ± 0.03.

b) For *binary interval coding*, the curves $c_2(x_1^n)$ vs. n are not as smooth as with the previous coding. This means that the transfer of information is very sensitive to the changes in the number of intervals used in the encoding process. In this case, the complexity curves do not display plateaus. Rather, from $n \approx 2500\text{-}3500$ time bins on, they decay in a convex $\cup$ way.

6.3. Number of states

The following Table (J.M. Amigó *et al.*, 2003) shows the (rounded-off) values of k^* within the samples in form of intevals, both for the multisymbol and binary time bin coding (with 5000 and 4048 bins, respectively).

	Multisymbol coding		Binary coding	
$k*$	In vivo	In vitro	In vivo	In vitro
Periodic current injection	$2-4$	$9-11$	$4-6$	$9-11$
Visual stimulation	$2-4$	$-$	$3-4$	$-$
Random current injection	$-$	$1-3$	$-$	$2-3$

7. Conclusions

1. Lempel Ziv complexity is a fast convergent estimator of the entropy and, therefore, it can be used to calculate the information content in spike trains even in situations where undersampling renders the standard approach dubious. We conclude that Lempel-Ziv complexity provides a valuable tool for the analysis of spike trains.

2. Apart from the estimation of the entropy, this measure of complexity allows to calculate complexity curves (with several encodings) for a given spike train as well as the number of states of the neuronal source which has produced that spike train.

3. These analytical tools have proved useful to separate time series of spike times that were obtained under different conditions. Thus, we have found significant differences between the spike trains obtained in the cortical neurons *in vivo* versus *in vitro*, such that those *in vivo* had higher information content than those *in vitro*, even when the stimulus was the same (sinusoidal current injection).

4. Finally, we think that the use Lempel Ziv complexity presents some advantages over other commonly used measurements of entropy and, therefore, we consider that it deserves to be better known and more used for spike train analysis.

References

[1] J.M. Amigó, J. Szczepański, E. Wajnryb, M.V. Sanchez-Vives(2001). In Ratta F (editor): *Proceedings of the World Congress on Neuroinformatics*. ARGESIM Report no. 20. ARGESIM/ASIM Verlag, Vienna (ISBN: 3-901608-20-6).

[2] J.M. Amigó, J. Szczepański, E. Wajnryb, M.V. Sanchez-Vives (2003). BioSystems 68:57.

[3] A. Lempel, J. Ziv (1976). IEEE Trans. Inform. Theory vol. IT-22:75.

[4] M.V. Sanchez-Vives, L.G. Nowak, D.A. McCormick (2000 a y b). J. Neurosci.20:4267 and J. Neurosci. 20:4267.