

Mutual information and redundancy in spontaneous communication between cortical neurons

J. Szczepanski · M. Arnold · E. Wajnryb ·
J. M. Amigó · M. V. Sanchez-Vives

Received: 2 September 2010 / Accepted: 31 January 2011 / Published online: 22 February 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract An important question in neural information processing is how neurons cooperate to transmit information. To study this question, we resort to the concept of redundancy in the information transmitted by a group of neurons and, at the same time, we introduce a novel concept for measuring cooperation between pairs of neurons called relative mutual information (RMI). Specifically, we studied these two parameters for spike trains generated by neighboring neurons from the primary visual cortex in the awake, freely moving rat. The spike trains studied here were spontaneously generated in the cortical network, in the absence of visual stimulation. Under these conditions, our analysis revealed that while the value of RMI oscillated slightly around an average value, the redundancy exhibited a behavior character-

ized by a higher variability. We conjecture that this combination of approximately constant RMI and greater variable redundancy makes information transmission more resistant to noise disturbances. Furthermore, the redundancy values suggest that neurons can cooperate in a flexible way during information transmission. This mostly occurs via a leading neuron with higher transmission rate or, less frequently, through the information rate of the whole group being higher than the sum of the individual information rates—in other words in a synergetic manner. The proposed method applies not only to the stationary, but also to locally stationary neural signals.

Keywords Neurons · Shannon information · Entropy · Mutual information · Redundancy · Visual cortex · Spikes train · Spontaneous activity

J. Szczepanski (✉) · E. Wajnryb
Institute of Fundamental Technological Research,
Pawińskiego 5B, 02-106 Warsaw, Poland
e-mail: jszczepa@ippt.gov.pl

E. Wajnryb
e-mail: ewajnryb@ippt.gov.pl

M. Arnold
Instituto de Neurociencias de Alicante, Universidad Miguel
Hernández-CSIC, 03550 San Juan de Alicante, Spain
e-mail: marta.arnold@umh.es

J. M. Amigó
Centro de Investigación Operativa,
Universidad Miguel Hernández, 03202 Elche, Spain
e-mail: jm.amigo@umh.es

M. V. Sanchez-Vives
ICREA-Institut d'Investigacions Biomediques August Pi y Sunyer
(IDIBAPS), 08036 Barcelona, Spain
e-mail: msanche3@clinic.ub.es

J. Szczepanski
Kazimierz Wielki University, Bydgoszcz, Poland

1 Introduction

Information in the nervous system is transmitted by spike trains. Spikes from individual neurons in the cerebral cortex can be induced by external stimulation or they can be internally generated by the recurrent connectivity within brain circuits, what is known as spontaneous activity. In this study, we will characterize information content in cortical spike trains. Information content is an abstract concept that refers to the richness of the interspike intervals, but it should not be understood as information necessarily related to an stimulus.

In this article, we study the issue of how neurons cooperate during spontaneous activity. The following questions arise in this context:

- (i) How does the information carried by neighboring neurons relate?

- (ii) Do neurons cooperate always in some specific way or, rather, do they switch from one manner to another?
- (iii) How do the above processes depend on the type of stimuli or more generally on the kind of environment in which the biological organism is currently placed in?
- (iv) How do neurons collaborate in the absence of external stimulation?

To answer some of these questions, we analyze spike trains from the primary visual cortex of the awake, freely moving rat in the absence of visual stimulation. In other words, we analyze spontaneous activity reflecting the way the network processes information based on its internal, recurrent connections only. These recordings were obtained with tetrodes, allowing us to record simultaneously the activity from several closely located neurons as well as the separation of their spike waves.

We then analyze *redundancy* and *relative mutual information* (RMI) within the group of neighboring, spike firing neurons from the primary visual cortex. The concept of redundancy, as introduced in Reich et al. (2001), compares the sum of the information rates transmitted separately by the neurons of a group with the information rate transmitted by the whole group. The second term is diminished by the maximal information rate among the neurons in the group. To measure the information conveyed by a given neuron about another neuron from the same group, we introduce the concept of RMI based on the commonly applied concept of mutual information (Weaver and Shannon 1963; Borst and Theunissen 1999). In general, the RMI between two information sources is simply the ratio between their mutual information and the average information transmitted by them. In our case, the information sources are a pair of neurons in stationary conditions. Note that in the standard approach the ratios between mutual information and either entropy of stimuli ensemble or entropy of spike trains are considered. Both concepts, redundancy and RMI require that entropy rates be estimated. We analyze RMI and redundancy for every instant of time. In order to maintain the correspondence of our results exactly the same experimental data was used to calculate both redundancy and RMI.

Neurons communicate by means of trains of action potentials or “spikes,” which are sharp voltage transitions (in the mV range) that have similar amplitude and shape for each of the neurons and, therefore, those parameters hardly carry any information [but see De Polavieja et al. (2005)]. Information is mostly conveyed therefore by the time of occurrence of such events, the spike times. The usual measures of the information content or entropy of spike trains requires first of all the discretization of these analog signals, which can be achieved in a variety of ways (Szczepanski et al. 2003). The *binary temporal coding* method is a common method of

coding described in the literature (Rieke et al. 1997). Here, the unit time interval is divided into small time bins (their size depending on the desired observation precision) and then for each time bin a 0 is assigned if no spike is occurred in the corresponding time interval, and a 1 if at least one spike occurred. Neurons generating spike trains under stationary conditions can be considered as information sources, where the source comprises everything involved in generating the message (type of neuron, codification, collaboration between neighbors, and stimuli of any kind). Note the importance of noise in the channel, which is responsible for the variability of responses to a given stimulus. Mathematically, such a source corresponds to a stationary stochastic process and the binned spike train to a particular realization of the process. This approach can be extended to general (non-stationary) conditions, provided that the stimuli and the neurons being studied may be considered (locally) stationary in sufficiently long intervals of time.

The most important characterization one can attach to an information source is its entropy rate. This measures the average information per symbol generated by the source (Shannon 1948; Cover and Thomas 1991). The entropy rate is well defined for stationary stochastic processes. Moreover, if the source is *ergodic* (a typical property in nature, at least locally in time), then its entropy rate can be estimated with the information obtained from a single (infinitely long) output, with probability 1 (Cover and Thomas 1991). This result provides in turn methods to estimate the entropy rate in practice, i.e., when the output is finite (see, e.g., Sect. 3.1). In other words, we can estimate the entropy rate of an ergodic information source on the base of a single, sufficiently long output.

Traditionally, mutual information is measured between stimuli and the evoked neuronal outputs (Rieke et al. 1997) or, in the case of redundancy, for a group of nearby neurons that are tuned to similar stimuli (Reich et al. 2001). The intuitive core of these ideas is to measure how much information is repeated from one neuron to other neurons for a given kind of stimuli. Thus, these two concepts are strongly related to a type of stimuli that corresponds in turn to the kind of host environment, in which the biological organism is currently located. Since in many cases there are no visual stimuli, it makes sense to ask the question how the visual neurons collaborate in such situations.

Thus, even though there is neither visual stimulation nor changes in the brain state (sleep–awake transitions) during the recordings, we do not assume that this is a stationary situation but rather we allow for a weak non-stationarity resulting from both motor and (non-visual) sensory processing. This being the case, we need to go beyond standard Information Theory in one respect: we consider sources that are not necessarily stationary, provided that they can be considered stationary in a smaller time scale. We call a source

locally stationary if it can be considered stationary in sufficiently long-time intervals. More generally, these periods of “quasi-stationarity” may be interrupted by comparatively short-transition phases between (quasi-)stationary states, as would happen in alternate periods of sleep and awake states. Locally stationary sources could be assigned, in general, by means of sliding time windows of adequate length T (see Sect. 3.4 for details) to phases in which time-dependent entropy is slowly changing. As a result, one considers rather a family of ergodic information sources, one for every instant of time. The entropy rate of a locally stationary neural signal at time t is then estimated in the time window $[t, t + T]$. We will use for this purpose the normalized Lempel–Ziv 76 complexity (Lempel and Ziv 1976), a technique we applied and tested in Amigó et al. (2004). Precisely, the explicit analysis of locally stationary neural signals (along with the introduction of RMI) is perhaps the most important aspect of this article.

Section 2 describes the experimental procedures. All mathematical tools (entropy rate and its estimations, redundancy, and RMI) have been gathered for the reader’s convenience in Sect. 3. Our results (Sect. 4) show that neurons collaborate in a flexible way: sometimes synergistically and, most frequently, via a leading neuron. In both cases, the mutual information between each pair of neurons is kept at the same relatively high level in relation to the average amount of information being transmitted by both neurons. This hints at a reliable mechanism of information transmission.

2 Experimental procedures

Recordings were obtained from Lister Hooded rats weighing 300–400 g at the time of surgery. In short, rats were chronically implanted with tetrodes in the primary visual cortex (–6 mm posterior and –4 mm lateral from bregma, see Paxinos and Watson 1998). Each tetrode was made from four twisted strands of HM-L-coated 90% platinum–10% iridium wire of 17 or 25 μm diameter (California Fine Wire, Grover Beach, CA). Four tetrodes were held by a cannula that was attached to a microdrive. Rats under deep anesthesia were chronically implanted with microdrives, using a mixture of isoflurane (0.5–1.2%), nitrous oxide (3.0 l/min), and oxygen (1.5 l/min). After 1 week of postoperative recovery recordings began. Electrophysiological recordings of neuronal activity were obtained while the animal was awake and freely moving in his home cage. These recordings were obtained in the dark and in the absence of visual stimuli, the recorded activity therefore being spontaneously generated in the visual cortex. The animal behavior was monitored with an infrared video camera mounted directly above the cage. Microelectrodes were advanced dorsoventrally by

50–75 $\mu\text{m}/\text{day}$ in steps of 25–50 μm . The electrode wires were AC-coupled to unity-gain buffer amplifiers. Lightweight hearing aid wires (2–3 m) connected these to a preamplifier (1,000 \times) and then to the filters and amplifiers of the recording system (Axona, St. Albans, UK). Signals were amplified further up to 10,000–35,000 and highpass filtered (360 Hz cutoff) and acquired using software from Axona Ltd. Each channel was continuously monitored at a sampling rate of 48 kHz, and action potentials were stored as 1 ms threshold-triggered “window” of samples for each spike (200 μs pre-threshold; 800 μs post-threshold) whenever the signal from any of the pre-specified recording channels exceeded a given threshold set by the experimenter for subsequent off-line spike sorting analysis. Cluster cutting (isolating single units from the multiunit recording data) was performed using an Off-Line Spike Sorter (Plexon Inc.). Waveforms were considered as being generated by a single neuron only if they occurred simultaneously in the four electrodes (Fig. 1a) and if they defined a discrete cluster in 3D principal component space that was distinct from clusters either from other units or multiunit activity (Fig. 1c, in 2D representation). The waveforms were sorted into units by an automatic method: the valley-seeking algorithm (Fukunaga 1972) that uses the inter-point distances in space to assign the waveforms to an optimal number of clusters. The degree to which the selected unit clusters are separated in the 2D and 3D cluster views is determined by a Multivariate Analysis of Variance (MANOVA) test. Single units had to exhibit a clearly recognizable refractory period (1 ms) in their ISI histograms (Fig. 1b). Once spikes originated in individual neurons and recorded in proximity were identified, the times of occurrence were used for subsequent analysis. At the end of the experimental recordings, the rat was given an overdose of sodium pentobarbital and perfused transcardially with saline, followed by 4% paraformaldehyde. The brain was sliced coronally into 100- μm thick sections, which were mounted and Nissl stained to aid visualization of the electrode track and tip. Recordings from 102 neurons have been included in this study.

3 Quantitative tools and methods

3.1 The entropy rate and the Lempel–Ziv estimator

Entropy rate estimators constitute a mathematically rigorous tool to approximate information transmission rates. This is an alternative to the analysis based on firing rate methods. It was shown that Lempel–Ziv complexity as defined in Lempel and Ziv (1976) can be used successfully as such an estimator (Amigó et al. 2004).

We recall the basics of Shannon entropy and Lempel–Ziv complexity. Consider blocks or “words” $x_1^l := x_1 x_2 \dots x_l$ of

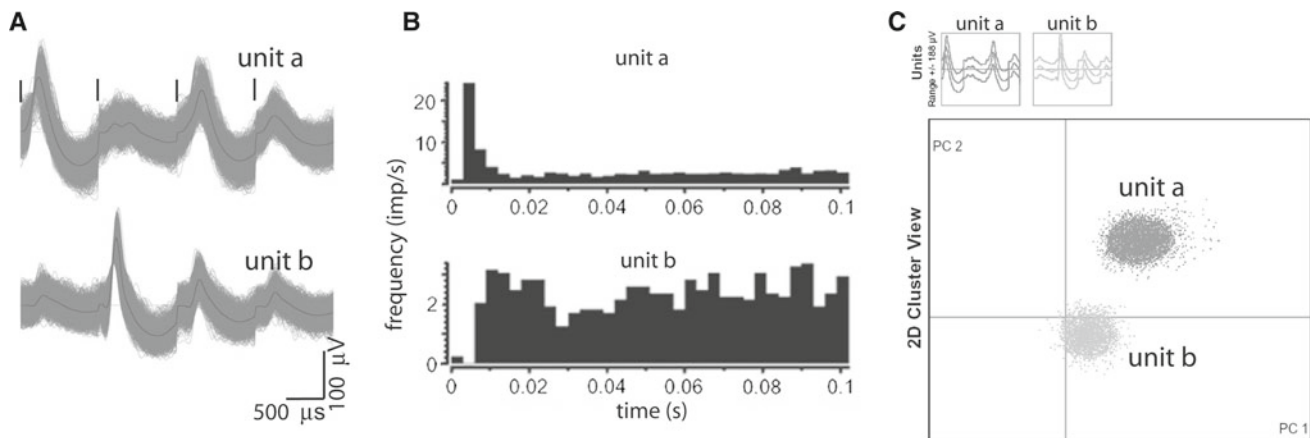


Fig. 1 Spike recording from the auditory cortex in a chronically implanted rat. **a** Spikes from two neurons recorded simultaneously with one tetrode. The vertical black lines determine the time windows correspondent to the four electrodes in the tetrode. Tetrodes are used to improve the capability of separating spikes originating in different neurons. Notice that the spike of “unit a” is of larger amplitude in the first

electrode, while the one of “unit b” is of larger amplitude in the second electrode. **b** Interspike interval histogram of each isolated spike. The refractory period following each spike further confirms that only spikes from one neuron are being included in each cluster. **c** 2D view of the principal component analysis of two units. Notice that the clusters corresponding to each unit do not overlap

length l output by an information source (or, equivalently, stationary stochastic process) X , with symbols or “letters” x_i taken from an “alphabet” $\mathcal{A} = \{a_1, \dots, a_\alpha\}$. With $p(x_1^l)$ denoting the probability of the word x_1^l of the length l , the block entropy of order l in units of bits is

$$H(l) = - \sum_{x_1^l} p(x_1^l) \log_2 p(x_1^l), \quad (1)$$

where the sum is over all possible such words x_1^l . The entropy rate of X , $h(X)$, is then defined as

$$h(X) = \lim_{l \rightarrow \infty} \frac{1}{l} H(l). \quad (2)$$

This theoretical definition is quite problematic when used as a statistical estimator of h in a finite-sized data set. Current entropy rate estimation methods (e.g., Strong et al. 1998) proceed by substituting the observed relative frequencies \hat{p}_i for p_i in Eq. 2, and subsequently extrapolating the finite l behavior of $H(l)/l$ when l goes to infinite (i.e., when $1/l$ goes to 0). Other approaches to entropy rate estimators are also discussed in the literature (London et al. 2002; Kennel et al. 2005; Kontoyiannis et al. 1998).

Intuitively speaking, the complexity of a sequence counts the number of different patterns appearing during its generation. Precise definition of the Lempel–Ziv 76 (LZ 76) complexity $C_{LZ}(x_1^l)$ of a length l word $x_1^l := x_1 x_2 \dots x_l$, with $x_i \in \mathcal{A} = \{a_1, \dots, a_\alpha\}$, is given in Lempel and Ziv (1976). We recall in short this definition. Consider the following sequence: $x_1^4 := x_1 x_2 x_3 x_4$. Take the first letter x_1 and set $C_{LZ} = 1$. Take the second letter x_2 . If $x_2 \neq x_1$, increase C_{LZ} by one; otherwise, consider the substring $x_2^3 := x_2 x_3$. If $x_3 \neq x_2^2 = x_1 x_2$, then a new pattern occurs and C_{LZ}

is increased by one; otherwise, consider $x_2^4 := x_2 x_3 x_4$ and check whether this substring occurs in $x_1^3 = x_1 x_2 x_3$. The procedure continues along these lines for longer sequences, the counter C_{LZ} being increased as soon as a new pattern occurs in the segment comprised x_1, x_2, \dots , up to the last-but-one letter appended. In this way, x_1^l gets decomposed into (minimal) patterns and $C_{LZ}(x_1^l)$ is the number of them.

Furthermore, the generation rate of new patterns along a word x_1^l is measured by the (LZ 76) normalized complexity,

$$c(x_1^l) = \frac{C_{LZ}(x_1^l)}{l / \log_2 l}. \quad (3)$$

Sequences generated by random sources have normalized complexity of about 1 with very high probability, whereas quasiperiodic sequences have very small values. It can be proved (Ziv and Lempel 1978) that, if the source X is ergodic i.e., the time average over trajectory is equal to the ensemble average, then

$$\limsup_{l \rightarrow \infty} c(x_1^l) = h(X) \quad (4)$$

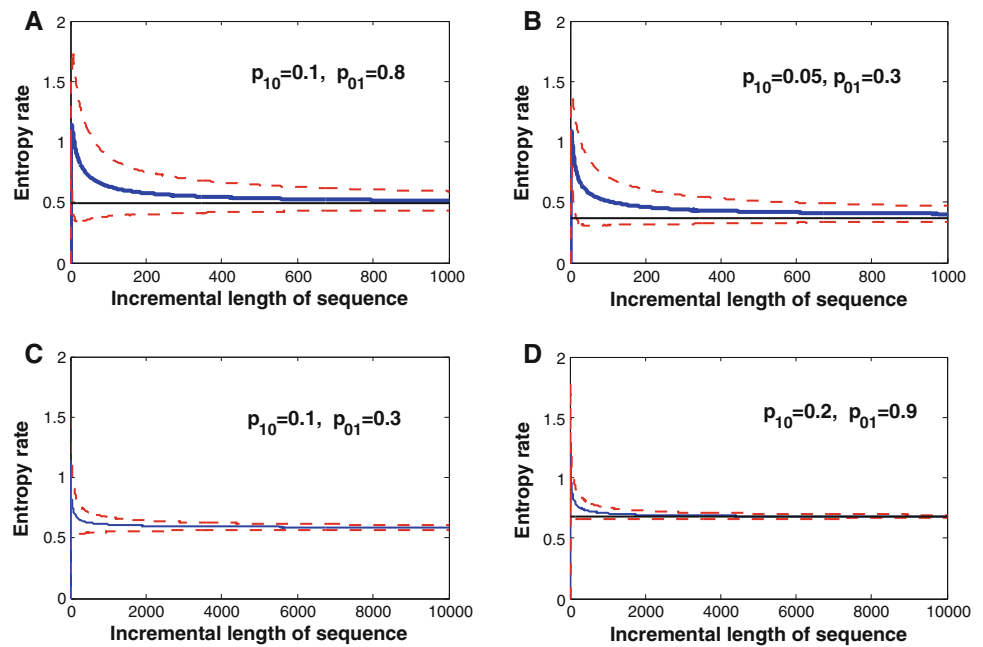
with probability 1. For precise definition of ergodicity, see Durrett (1996).

Equation 4 provides a simple way to estimate the entropy rate of an ergodic source via the normalized complexity of a sequence x_1^L of finite length $L \gg 1$ produced by it, namely,

$$c(x_1^L) \simeq h(X). \quad (5)$$

However, while the performance of the standard entropy rate estimators (based on maximum likelihood) depends critically on l as compared to L because of undersampling, the applicability of the normalized complexity rate relies rather on

Fig. 2 Illustration of the convergence rate of the Lempel–Ziv 76 entropy estimator as a function of the incremental length (window size) of a sequence. A typical Markov processes with relatively small entropy characteristic for spike trains have been considered. In each case 1,000 trajectories have been generated. **a** and **b** show the short-length behavior, while **c** and **d** show the long-length estimation. The true value (straight horizontal line), the average value (solid thick line) of estimated entropy and two SD bounds are depicted (dashed lines). In the top panels, we see that a few hundred bit long window provides a satisfactory estimation of entropy



the convergence speed of $c(x_l^l)$ as $l \rightarrow L$, i.e., as the pattern count goes on. Therefore, fast convergence of the normalized complexity entails not only reliable entropy rate estimations, but also the possibility to assess the entropy rate of the neuron with short data series. In particular, one expects this to be the case when the source entropy is small (i.e., when the source is far from random), because then the pattern count is low. This is important because periods of stationarity can be comparatively short. For more details, see Amigó et al. (2004).

In order to assess the choice of the window length and to justify the quality of the Lempel–Ziv estimator applied (especially for parameters typical of spike trains), we computed its convergence rate for a few Markov processes, see Fig. 2. For these processes the entropy rate is known also analytically. This figure shows that the convergence rate is quite satisfactory, especially when we take into account biological applications. Moreover, it is worth to stress that an estimator for the variance of Lempel–Ziv complexity was proposed recently (Amigó and Kennel 2006). It was shown that for three-state Markov processes and for 400 symbols long sequences the standard deviation is no larger than 0.08. Thus, for the two-state process (this is the case of the redundancy) the accuracy should be even better. In order to estimate the joint entropy for the case of RMI the four-state process was used, which leads to slightly worse accuracy. Moreover, numerical simulations show that LZ 76 overestimates the entropy rate, which in the case of estimation of indicators that are constructed as a ratio (like RMI and redundancy) of entropy rates of some quantities leads to a lower bias. This is the advantage of the RMI index over MI one.

3.2 Redundancy

Redundancy is defined in the literature in several ways; see, e.g., Barlow (2001), Gawne and Richmond (1993), Panzeri et al. (1999), Brenner et al. (2000), Machens et al. (2001), Simoncelli and Olshausen (2001), Schneidman et al. (2003), Rolls et al. (2004), Latham and Nirenberg (2005), and Puchalla et al. (2005). In general, the main idea is based on comparison of the entropy rate of the given memoryless, uniformly distributed (that is with maximal possible entropy rate to be reached) input information source with the output entropy rate. In this article, we apply the definition adapted in neuroscience and proposed in Reich et al. (2001). This definition is based on the comparison between the amount of information conveyed by the whole neuronal ensemble and the information carried by the individual cells. We start by recalling the basic concepts that we need. This idea is illustrated in Fig. 3.

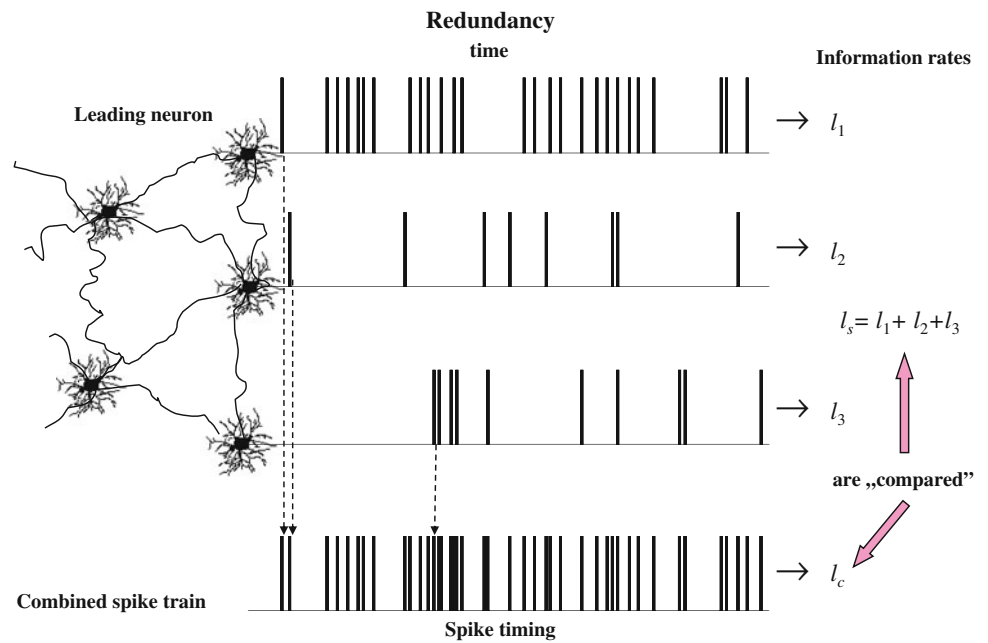
Let us recall that if X is a discrete random variable with probability function $p(x_i)$, then the information (in the Shannon sense) conveyed by the event $X = x_i$ is equal (in units of bits) to

$$I(x_i) = -\log_2 p(x_i). \tag{6}$$

Observe that an event conveys the more information, the less likely it is. The entropy of X , $H(X)$, is the average information conveyed by the observation of an event.

In the more general case of a sequence of random variables (stochastic processes) $X_1^\infty := X_1 \dots X_n \dots$, the information $I(x_1, \dots, x_n)$ conveyed by the event $X_1 = x_1, \dots, X_n = x_n$, is given by

Fig. 3 The concept of redundancy is illustrated. The sum of information for each neuron separately is compared to information transmitted by the combined spike train



$$I(x_1, \dots, x_n) = -\log_2 p(x_1, \dots, x_n), \tag{7}$$

with $p(x_1, \dots, x_n)$ being the probability distribution of the vector random variable $X_1^n := X_1 \dots X_n$. The *information rate* of X_1^n is then defined as

$$l(x_1, \dots, x_n) = -\frac{1}{n} \log_2 p(x_1, \dots, x_n). \tag{8}$$

According to the Shannon–McMillan theorem [also called the Asymptotic Equipartition Property, see Cover and Thomas (1991)],

$$\lim_{n \rightarrow \infty} l(x_1, \dots, x_n) = h(X_1^\infty) \tag{9}$$

with probability 1, if the finite-valued stationary process (or information source) X_1^∞ is ergodic. This result will allow us below to estimate information rates of ergodic sources by means of entropy rate estimators (such as Lempel–Ziv complexity) applied to single, sufficiently long outputs.

The concept of redundancy we consider henceforth was introduced in Reich et al. (2001) in order to measure the cooperation of a group of neurons: do they work synergistically or rather in an uncoordinated way? Its definition is as follows:

Consider a group of neighboring neurons $\mathcal{G} = \{N_1, N_2, \dots, N_k\}$ and denote by:

- l_c the information rate of the combined spike train, i.e., the signal obtained by aggregating the spikes emitted individually by the neurons of \mathcal{G} .
- l_s the sum of information rates for each cell separately.
- l_i the information rate of neuron N_i .

Then, the redundancy R of \mathcal{G} is defined by

$$R = \frac{l_s - l_c}{l_s - \max_{1 \leq i \leq k} \{l_i\}}. \tag{10}$$

Furthermore, to estimate l_c we apply the so-called “summed-population code” (Reich et al. 2001), so that the information rate l_c is estimated with the same accuracy as the information rates l_i for single neurons because only two letters are used in summed-population code too. Thus, the essence of this concept is just to compare the amount of information transmitted by each single neuron coming from a group with information transmitted when the group is treated as a single source of information. In the article, by Reich, Mechler and Victor stimuli are treated as a parameter.

Observe that the denominator is always positive, while the numerator can have both signs. In particular, (i) $R > 1$ if stimulus-related information in jointly analyzed responses is contradictory or confusing, and (ii) $R < 0$ if the neurons fire synergistically. Indeed, case (i) means that $\max_{1 \leq i \leq k} \{l_i\} > l_c$, i.e., there is a single neuron for which the information rate is greater than the information rate of the spike train of the whole ensemble; this is possible if, e.g., one neuron fires evenly, while the firing patterns of the rest is biased in the same way. Case (ii) means that $l_s < l_c$, i.e., the sum of the information rates coming from all neurons separately is smaller than the information rate of the spike train of the whole ensemble. Needless to say, all these quantities can be evaluated with any entropy estimator, say, with the normalized complexity.

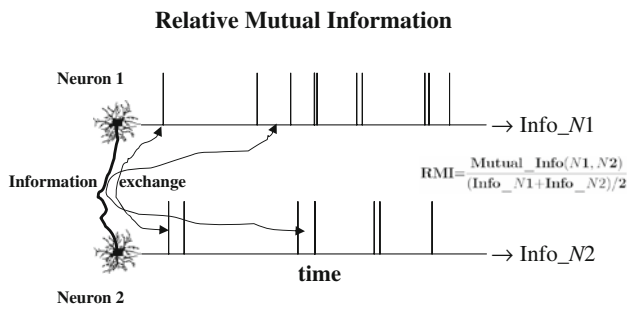


Fig. 4 The idea of relative mutual information concept is introduced. The neurons communicate via the synapses, which constitutes synaptic activity, which is the basis for mutual information. This mutual information is compared to the average information transmitted by both neurons

3.3 Relative mutual information

The amount of information that a random variable Y carries about another random variable X is a central question in Information Theory. The answer is given by the *mutual information* $I(X, Y)$ between X and Y (Cover and Thomas 1991; Paninski 2003; London et al. 2008),

$$\begin{aligned}
 I(X, Y) &:= H(X) - H(X|Y) \\
 &= H(X) + H(Y) - H(X, Y) \\
 &= I(Y, X),
 \end{aligned}
 \tag{11}$$

where $H(X)$ is the entropy of X , $H(X|Y)$ is the conditional entropy measuring our uncertainty about X knowing the realization of Y , and $H(X, Y)$ is the joint entropy of X and Y .

The relative mutual information $RMI(X, Y)$ between the random variables X and Y is just the ratio between their mutual information $I(X, Y)$ and the average of information transmitted by X and Y :

$$RMI(X, Y) = \frac{H(X) + H(Y) - H(X, Y)}{[H(X) + H(Y)]/2}.
 \tag{12}$$

In other words, $RMI(X, Y)$ measures the reduction of uncertainty concerning realization of X having knowledge about the realization of Y and compares with the average uncertainty of realization of both X and Y . This concept is essentially different and can be treated as complementary to cross-correlations analysis. The basic idea of the RMI concept is illustrated in Fig. 4. In contrast to the standard approach that compares $I(X, Y)$ with $H(X)$ or $H(Y)$, a more natural idea is to compare $I(X, Y)$ with information transmitted just by both neurons, since we are interested in cooperation between neurons during information transmission.

Note that $RMI(X, Y)$ is non-dimensional. In the appendix, it is proven that $0 \leq RMI(X, Y) \leq 1$. If the random variables X, Y are independent, then $I(X, Y) = 0$ and, hence, $RMI(X, Y) = 0$ too. (Intuitively, this means that the knowledge of the realizations of Y does not reduce on average

the uncertainty about the realizations of X .) On the contrary, if the dependence between random variables X, Y is deterministic (i.e., knowing the realization of X implies certainty about the realization of Y), then $RMI(X, Y) = 1$.

This definition can be generalized without formal changes to sequences of random variables (stochastic processes) $X_1^\infty := X_1 \dots X_n \dots$ and $Y_1^\infty := Y_1 \dots Y_n \dots$, just by first considering finite length blocks X_1^n, Y_1^n with RMI

$$\begin{aligned}
 RMI(X_1^n, Y_1^n) &= \frac{H(X_1^n) + H(Y_1^n) - H(X_1^n, Y_1^n)}{[H(X_1^n) + H(Y_1^n)]/2}
 \end{aligned}
 \tag{13}$$

(since, for each n fixed, X_1^n and Y_1^n can be treated as random variables) and then taking the limit

$$RMI(X_1^\infty, Y_1^\infty) = \lim_{n \rightarrow \infty} RMI(X_1^n, Y_1^n).
 \tag{14}$$

In real time series analysis, $RMI(X_1^\infty, Y_1^\infty)$ has to be estimated by $RMI(X_1^n, Y_1^n)$, for n sufficiently long.

We would like to stress that we estimate $H(X, Y)$ in a natural way: after encoding (see Sect. 3.4) spike trains X and Y , we assign one of the four letters a, b, c, d to each bin:

- a —when in this bin there is a spike coming from X and no spike coming from Y ,
- b —when in this bin there is no spike coming from X and there is a spike coming from Y ,
- c —when in this bin there is a spike coming from X and there is a spike coming from Y ,
- d —when there is no spike in the bin.

Thus, for the pair of spike trains (X, Y) we obtain corresponding sequence of symbols (based on a four-letter alphabet) representing its joint distribution and, next, we calculate the normalized complexity with four letters formula (Amigó et al. 2004), so the dimensionality of letters does not increase significantly.

3.4 Coding and sliding windows

As already mentioned in the Sect.1, spike trains can be encoded in many ways. We have used in our analysis the binary temporal coding (Rieke et al. 1997). If the time interval T is divided into q bins, the resulting *time resolution* is $\Delta\tau = T/q$. Thus, within each time window $[t, t + T]$, a spike train is encoded as a sequence of q bits; the resulting message translates the information being generated by the neuron (up to a resolution time $\Delta\tau$) to a format amenable to the methods of Information Theory. The window length T has to be long enough to guarantee a good entropy estimation (this depends on the estimation method used). Moreover, if

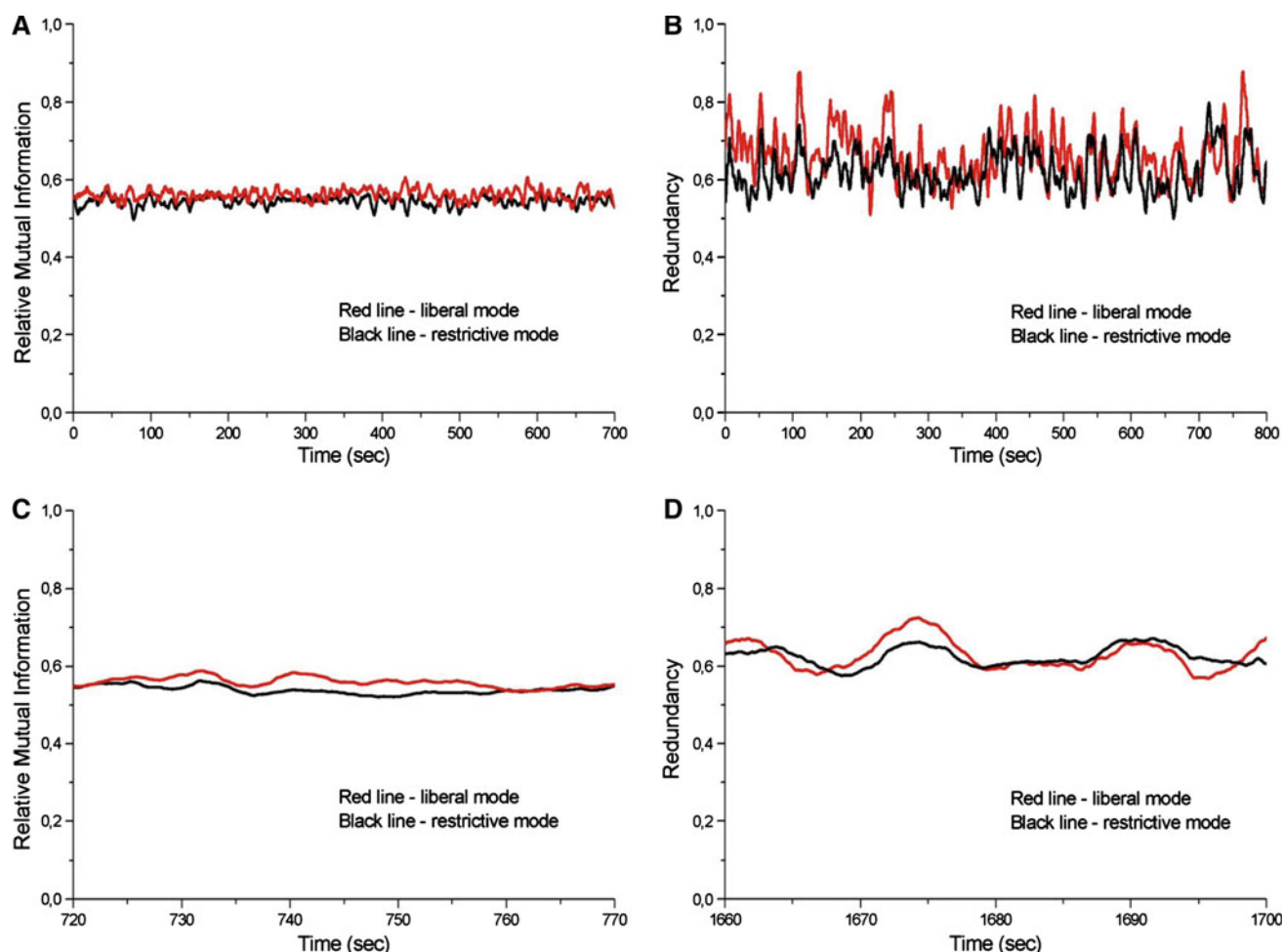


Fig. 5 **a** RMI for the same pair of neurons with restrictive and non-restrictive spike cutting. The long-time scale result is presented. **b** Redundancy for the same group of neurons with restrictive and non-

restrictive spike cutting. **c, d** The later stage of experiment is presented. For better illustration, a shorter time period has been plotted

the neuron is not considered stationary but rather only locally stationary, as we do in the case of a rat moving freely in the dark, T must be also short enough to guarantee (approximate) stable conditions within the window. Numerical calculations show that $T = 5$ s is a good trade-off between these two conditions.

The quantity q is the other encoding parameter in this scheme. In our analyses, we set the *encoding frequency* $f = 1/\Delta\tau$ equal to 80 Hz. This choice is based on the so-called saturation levels studied in [Szczepanski et al. \(2003\)](#): from $f = 80$ Hz on, the increase of encoding frequency does not affect the values of the normalized complexity obtained. In other words, above $f = 80$ Hz we cannot extract more information from the signal. Thus, the length of the words inside each window is $l = fT = 400$ symbols. These binary sequences can be now treated as stochastic trajectories for calculations of the redundancy and as realizations of random variables for the calculation of the RMI.

4 Results

Multiunit recordings were obtained by means of chronically implanted tetrodes in the visual cortex of freely moving rats. During the experiment, a voltage threshold was set through visual inspection of the recording. All events that crossed the threshold were recorded and sorted off-line (see Sect. 2). In order to explore to what extent being more or less restrictive in the spike cluster cutting would affect the estimations of both redundancy and RMI, we compared two different forms of cluster cutting: “restrictive” and “non-restrictive.” The clusters were represented in the 3D PC space using off-line sorter (Plexon Inc.) and the centroids of the clusters (Fig. 1c) calculated, as well as the Mahalanobis distance ([Mahalanobis 1936](#)) between each point in the cluster and the centroid. We considered “restrictive” cutting the one when all points farther than $\geq 2SD$ from the centroid were removed as outliers, and “non-restrictive” when no outliers were removed.

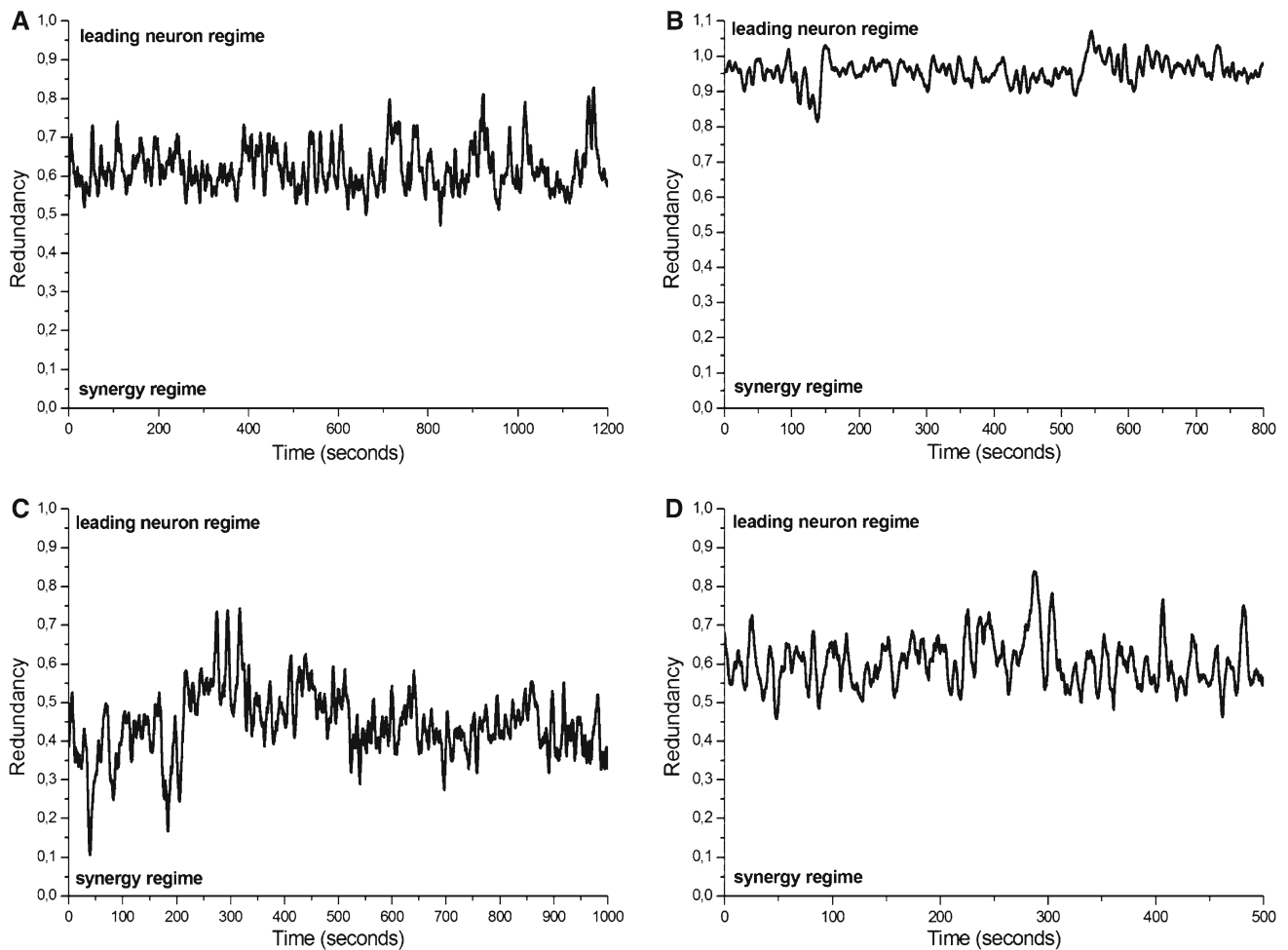


Fig. 6 **a** Evolution for the redundancy in a spike train during a 1,200 s recording. Large fluctuations in redundancy can be observed, but close to the leading neuron regime (above 0.5). The conjecture is that neurons can exchange the leading role during transmission. **b** Neurons can also work in one regime for an extended period of time. This figure corresponds to a leading neuron regime of transmission. Probably neurons did not exchange the leading role in this experiment. **c** Transitions between different regimes of information transmission (0.1–0.75) can

be observed in this figure. Note how these transitions are brisker than the ones presented above, e.g., **b**. **d** In this experiment, the redundancy of a group of neurons changes between 0.5 and 0.9. The degree of cooperation within a group of neurons depends on the regime of information transmission (uncoordinated or driven by a leading neuron). Here, the transitions to different stage regimes take place in a mild way. This figure presents the early stage of experiment

In Fig. 5, we display four examples of the results obtained with a restrictive and non-restrictive sorting for both redundancy and RMI. As can be observed, there is no significant difference between the two parameters independently of how restrictive the spike sorting was. This analysis was performed for the recordings of 11 neurons and no significant differences were detected. Indeed, the mean and the standard deviation for restrictive and non-restrictive sorting differed by <5% in all cases, both for redundancy and RMI. The fact that these curves almost coincide for liberal and restrictive types of sorting suggests that these sortings practically do not affect neither the RMI nor the redundancy. To be more specific these results show that most of the information is carried by the spikes with “a typical profile,” i.e., those which

are close to the center of “mass” in the sense of the Mahalanobis distance. The removed outliers (those being farther than two standard deviations from the centroid) do not significantly affect the amount of transmitted information. In this sense, the precise spike-timing codes are also not very much affected. We conclude therefore that the type of the sorting mode does not affect in an essential way neither the redundancy nor the RMI within a group of neurons. Thus, in the rest of the study non-restrictive spike sorting is used.

Next, we summarize the main results of our calculations.

(i) We analyzed the redundancy of 19 groups of neurons. The number of neurons was 3–13, depending on the group. Figures 6 and 7 show four different behaviors of the redundancy. In general, the redundancy of the information

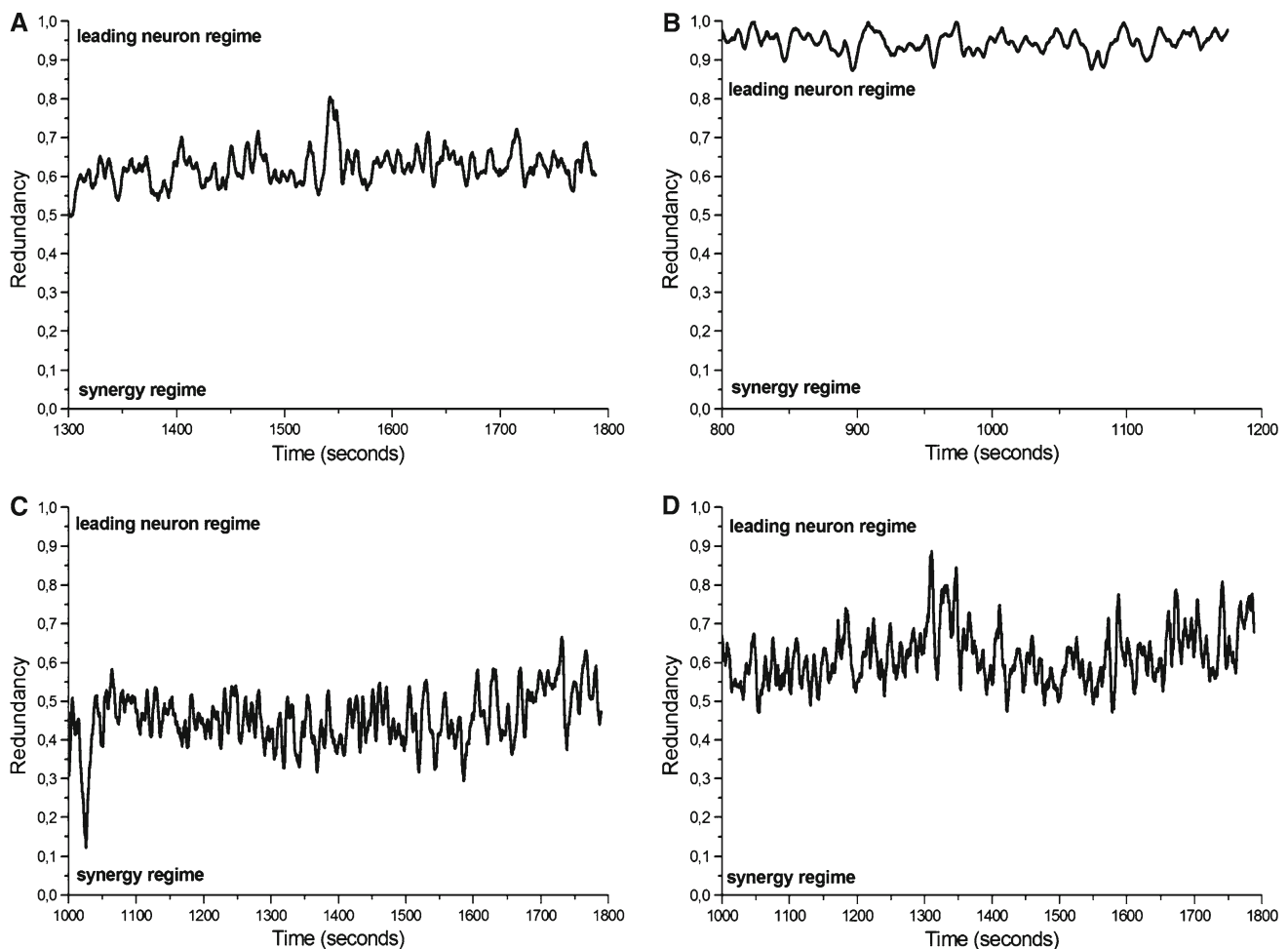


Fig. 7 Same as Fig. 6, but the later stage of experiment is presented showing the unchanged temporal behavior of the redundancy index

transmitted by a group of neurons is characterized by a large variability (the standard deviation is 0.08); the mean value can be essentially different depending on the group of neurons (mean value varies from 0.45 to 0.90 depending on the group).

We have found different types of group collaboration between neurons which we describe next.

The first type of collaboration that we observed (Figs. 6a, 7a) was characterized by a sharp oscillation in relatively short periods of time during the whole recording (1,200 s). The regime of this collaboration was somewhere between synergistic and “leading neuron” [called “contradictory” in Reich et al. (2001)] methods.

For another group of neurons, the redundancy was characterized by small fluctuations (Figs. 6b, 7b) showing an stable redundancy curve. This kind of collaboration takes place when the leading neuron plays the crucial role in a transmission process. The value of redundancy in these cases is close to one.

In the third case, we have observed that during transmission the regime of collaboration changes from one type to another (Figs. 6c, 7c). In Fig. 6c for the initial 200 s one can see the periods of synergistic collaboration (redundancy drops as low as to 0.1), while in the time interval 300 ± 40 s the redundancy jumps up to the values as high as 0.8 which corresponds to a “leading neuron” regime. Then, during the final 500 s neurons collaborated in a mid-regime (the redundancy value oscillated about 0.5). Similar behavior one can be seen in Fig. 6c.

The fourth type of neuron collaboration we have observed was a slow transition (Figs. 6d, 7d) from “the leading neuron regime” (redundancy about 0.9) to “mid-regime” (redundancy about 0.5), followed by a slow increase up to the value of 0.8 (Fig. 7d). Similar behavior was observed in an earlier stage of this experiment (Fig. 7d). The fluctuations of this redundancy curve are relatively large especially in comparison with the curve presented in Figs. 6b and 7b.

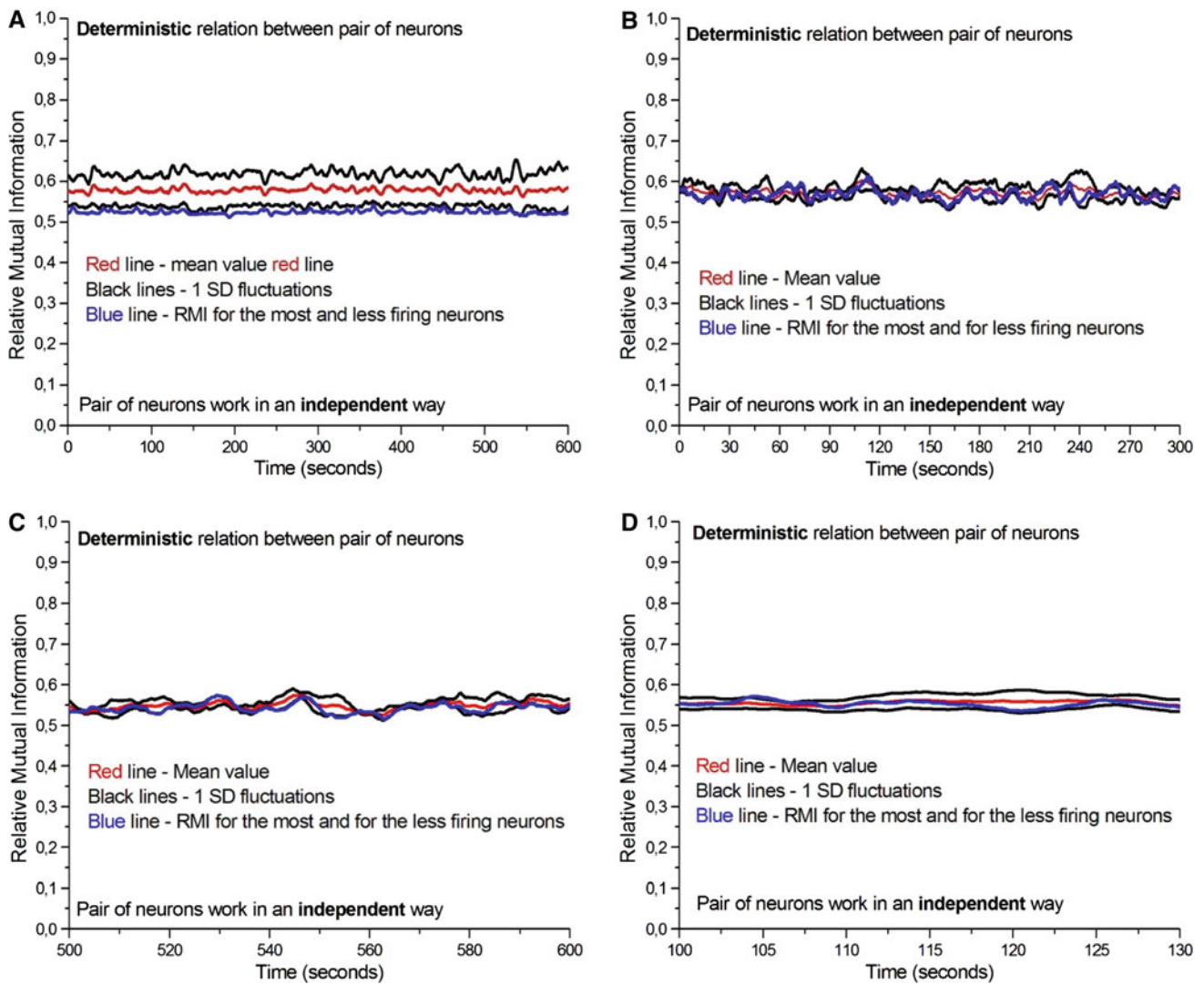


Fig. 8 Relative Mutual Information plots for four groups of neighboring neurons. Within each group RMI for each pair of neurons was calculated. At each moment of time the average RMI and one standard

deviation (1 SD) departure from the average value are depicted. In **a** (13 neurons) and **b** (six neurons) RMI curves for longer time scale, while in **c** (three neurons), **d** (five neurons) for short-time scale are presented

(ii) RMI was studied for four groups of neurons. Within each group, RMI between each pair of neurons was calculated. Our results show that the information carried by one neuron about the spike activity of another neuron (for nearby cortical neurons) is almost the same for long-time spans (the mean value of the RMI being around 0.5–0.6) and the variability of this relation is very small (the standard deviation is 0.015).

The typical time behavior of RMI is presented in Figs. 8 and 9. In Figs. 8a, b and 9a, b, we show results for longer time scale. On Figs. 8a and 9a, a group consisting of 13 neighboring neurons was considered. For each pair of neurons (78 pairs) the RMI was calculated. At each moment of time the average RMI and \pm one standard deviation (1 SD) departure from the average value were depicted. We included

also the RMI calculated for the single pair consisting of the most and the least firing neurons within the group. In Figs. 8b and 9b, longer time scale RMI runs for a smaller group consisting of six neurons are presented. Finally, Figs. 8c and 9c (three neuron group) and Figs. 8d and 9d (five neuron group) present the plots of RMI time runs for relatively short-time scales.

In all cases, the RMI curves exhibit a relatively high mean value and small relative fluctuations measured by ratio standard deviation and average value.

By definition of RMI this relatively large mean value proves that each neuron conveys a large amount of information about the activity of its neighbor. Moreover, our results show that the RMI fluctuations are significantly smaller than the redundancy fluctuations. The above results show that each

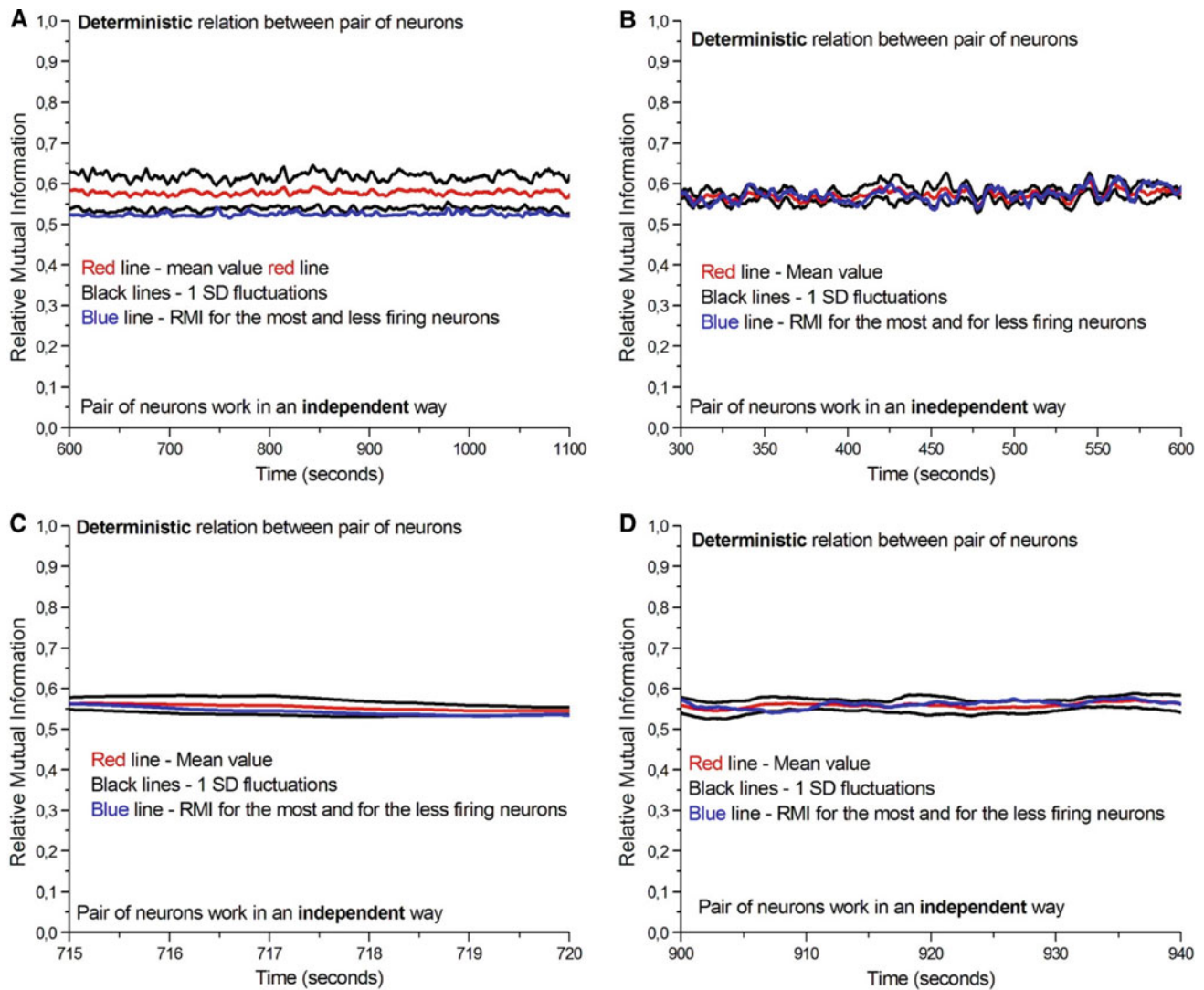


Fig. 9 Same as Fig. 8, but the later stage of experiment is presented showing the unchanged temporal behavior of the RMI index

pair of neighboring neurons exchange information in a very regular way, while the cluster of neurons collaborate in a variable way although usually within leading neuron regime (redundancy above 0.5).

(iii) Information transmission via a leading neuron (a role of which is often passed from one neuron to another of the same group) combined with the property (ii), is efficient and robust.

5 Discussion

We have studied the redundancy and RMI in groups of neighboring neurons from the primary visual cortex in the awake, freely moving rat. Given that no specific visual stimuli were presented, we may assume that the spike trains are locally stationary, allowing us to estimate those parameters by means of 5-s long, sliding windows. The estimation of the necessary

information and entropy rates was obtained by means of the normalized Lempel–Ziv complexity, a technique that the authors have studied in previous works; see, e.g., Szczepanski et al. (2003), Amigó et al. (2004). The encoding parameters and window length were chosen precisely on the grounds of our experience with this estimator.

A large variability of the redundancy for groups of neurons (as in Figs. 6a, c, d and 7a, c, d) suggests that during the process of information transmission, a group of neurons can collaborate in a flexible way. Specifically, neurons can collaborate in both ways: with a leading neuron with higher transmission rate (more frequently) or closer to a synergetic way, i.e., with the information rate of the whole group higher than the sum of the information rates of each individual neuron. The values obtained by Reich et al. (2001) correspond to the first case. This flexibility in the collaboration corroborates that the transmission of information by the brain is characterized by high reliability.

To the best of our knowledge, the concept of RMI has been introduced by the present article. Being the mutual information of two random variables (modeling two neurons) normalized by the average of their entropies, RMI is a dimensionless parameter ranging from 0 to 1 that is very well suited for our purposes. The values of the RMI measured in our experiments were quite constant over time, in spite of the different activities being carried out by the rat in that time span. These values are above 0.5, the value 1.0 corresponding to a deterministic relation. This means that during information transmission, neurons convey information about the patterns, transmitted by each neighboring neuron at a similar relatively high level (with relatively small variance) in relation to the total information being transmitted. More explicitly, the results of our calculations (RMI close to 0.55), show that the “uncertainty” of spiking of one neuron reduces by half (in comparison to the average “uncertainty” of spiking of both neurons) provided we have information about spiking of the other neuron. We conjecture that this is to make the transmission resistant against noise that naturally occurs in biological systems.

Finally, we highlight the fact, supported by our results, that neurons can cooperate in different ways (as indicated by the redundancy) while keeping the RMI between neighboring neurons at the same level.

Acknowledgments This research was the result of a collaboration supported by the Spanish-Polish Scientific Cooperation Program CSIC-PAS, grant 17/R04/R05. Experimental research was supported by the Spanish Ministry of Science and Innovation and the EU Future and Emerging Technologies Program (PRESENCIA project) to MVSV.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Lemma *Let X and Y be two random variables. The following properties of*

$$RMI(X, Y) \equiv \frac{H(X) + H(Y) - H(X, Y)}{[H(X) + H(Y)]/2} \tag{15}$$

hold true:

1. $0 \leq RMI(X, Y) \leq 1$;
2. $RMI(X, Y) = 0$ if and only if X and Y are independent;
3. $RMI(X, Y) = 1$ if and only if there exists a deterministic relation between X and Y.

Proof 1. First of all (Ash 1965; Theorem 1.4.4),

$$\begin{aligned} H(X, Y) &= H(X) + H(Y|X), \\ H(X, Y) &= H(Y) + H(X|Y), \end{aligned} \tag{16}$$

where $H(X|Y) \geq 0$ and $H(Y|X) \geq 0$ are conditional entropies. It follows

$$H(X) \leq H(X, Y), \quad H(Y) \leq H(X, Y), \tag{17}$$

Adding both inequalities, we get

$$\frac{H(X) + H(Y)}{2} - H(X, Y) \leq 0 \tag{18}$$

and, adding now $\frac{H(X)+Y}{2}$ to both sides, we end up with

$$H(X) + H(Y) - H(X, Y) \leq \frac{H(X) + H(Y)}{2}. \tag{19}$$

If $H(X) \neq 0$ or $H(Y) \neq 0$, then it follows

$$\frac{H(X) + H(Y) - H(X, Y)}{[H(X) + H(Y)]/2} \leq 1. \tag{20}$$

Furthermore, $RMI(X, Y) \geq 0$ follows trivially from $I(X, Y) = H(X) + H(Y) - H(X, Y) \geq 0$ and $H(X) \geq 0, H(Y) \geq 0$.

The case $H(X) = H(Y) = 0$ we put aside before (20) corresponds to random variable with δ -distributions (i.e., $X = x$ and $Y = y$ with probability 1), so that X and Y are constant (and deterministic) for all practical purposes. Since this is a limiting case in the probability distribution, we pass to the limit and define $RMI(X, Y) = 1$.

2. X and Y are independent if and only if $I(X, Y) = 0$ (Cover and Thomas 1991), and $I(X, Y) = 0$ if and only if $RMI(X, Y) = 0$.

3. In (17) equalities hold if and only if $H(X|Y) = H(Y|X) = 0$ (see (16)), which means that the relation between X and Y is deterministic (Ash 1965; p. 51). \square

References

Amigó JM, Kennel MB (2006) Variance estimators for the Lempel–Ziv entropy rate estimator. *Chaos* 16:043102

Amigó JM, Szczepanski J, Wajnryb E, Sanchez-Vives MV (2004) Estimating the entropy rate of spike trains via Lempel–Ziv complexity. *Neural Comput* 16:717–736

Ash RB (1965) *Information theory*. Wiley, New York

Barlow H (2001) Redundancy reduction revisited. *Netw Comput Neural Syst* 12(3):241–253

Borst A, Theunissen FE (1999) Information theory and neural coding. *Nat Neurosci* 2:947–957

Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck R (2000) Synergy in neural code. *Neural Comput* 12:1531–1552

Cover TM, Thomas JA (1991) *Elements of information theory*. Wiley, New York

De Polavieja GG, Harsch A, Kleppe I, Robinson HP, Juusola M (2005) Stimulus history reliably shapes action potential waveforms of cortical neurons. *J Neurosci* 25:5657–5665

Durrett R (1996) *Probability: theory and examples*. Duxbury Press, Belmont

Fukunaga K (1972) *Introduction to statistical pattern recognition*. Academic Press, New York

Gawne TJ, Richmond BJ (1993) How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13:2758–2771

- Kennel MB, Shlens J, Abarbanel HDI, Chichilnisky EJ (2005) Estimating entropy rates with Bayesian confidence intervals. *Neural Comput* 17:1531–1576
- Kontoyiannis I, Algoet PH, Suhov YM, Wyner AJ (1998) Nonparametric entropy estimation for stationary processes and random fields, with applications to English text. *IEEE Trans Inf Theory* 44:1319–1327
- Latham PE, Nirenberg S (2005) Synergy, redundancy and independence in population codes, revisited. *J Neurosci* 25:5195–5206
- Lempel A, Ziv J (1976) On the complexity of individual sequence. *IEEE Trans Inf Theory* IT-22:75–88
- London M, Schreibman A, Hausser M, Larkum ME, Segev I (2002) The information efficacy of a synapse. *Nat Neurosci* 5:332–340
- London M, Larkum ME, Hausser M (2008) Predicting the synaptic information efficacy in cortical layer 5 pyramidal neurons using a minimal integrate-and-fire model. *Biol Cybern* 99:393–401
- Machens CK, Stemmler MB, Prinz P, Krahe R, Ronacher B, Herz AV (2001) Representation of acoustic communication signals by insect auditory receptor neurons. *J Neurosci* 21:3215–3227
- Mahalanobis PC (1936) On the generalised distance in statistics. *Proc Natl Inst Sci India* 2:49–55
- Paninski L (2003) Estimation of entropy and mutual information. *Neural Comput* 15(6):1191–1253
- Panzeri S, Schultz SR, Treves A, Rolls E (1999) Correlations and the encoding of information in the nervous system. *Proc R Soc Lond B* 266:1001–1012
- Paxinos G, Watson C (1998) *The rat brain in stereotaxic coordinates*. Academic Press, New York
- Puchalla JL, Schneidman E, Harris RA, Berry MJ (2005) Redundancy in the population code of the retina. *Neuron* 46:493–504
- Reich DS, Mechler F, Victor JD (2001) Independent and redundant information in nearby cortical neurons. *Science* 294:2566–2568
- Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1997) *Spikes: exploring the neural code*. MIT Press, Boston
- Rolls ET, Aggelopoulos NC, Franco L, Treves A (2004) Information encoding in the inferior temporal visual cortex: contributions of the firing rates and the correlations between the firing of neurons. *Biol Cybern* 90:19–32
- Schneidman E, Bialek W, Berry MJ (2003) Synergy, redundancy and independence in population codes. *J Neurosci* 23:11539–11553
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423, 623–656
- Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. *Annu Rev Neurosci* 24:1193–1216
- Strong SP, Koberle R, de Ruyter van Steveninck R, Bialek W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200
- Szczepanski J, Amigó JM, Wajnryb E, Sanchez-Vives MV (2003) Application of Lempel–Ziv complexity to the analysis of neural discharges. *Netw Comput Neural Syst* 14:335–350
- Weaver W, Shannon CE (1963) *The mathematical theory of communication*. University of Illinois Press, Urbana
- Ziv J, Lempel A (1978) Compression of individual sequences via variable-rate coding. *IEEE Trans Inf Theory* IT-24:530–536