Behavioral/Systems/Cognitive

Mismatched Decoding in the Brain

Masafumi Oizumi,[1] Toshiyuki Ishii,[2] Kazuya Ishibashi,[1] Toshihiko Hosoya,[2] and Masato Okada[1,2]
[1]University of Tokyo, Kashiwa, 277-8561 Chiba, Japan, and [2]RIKEN Brain Science Institute, 351-0198 Saitama, Japan

"How is information decoded in the brain?" is one of the most difficult and important questions in neuroscience. We have developed a general framework for investigating to what extent the decoding process in the brain can be simplified. First, we hierarchically constructed simplified probabilistic models of neural responses that ignore more than *K*th-order correlations using the maximum entropy principle. We then computed how much information is lost when information is decoded using these simplified probabilistic models (i.e., "mismatched decoders"). To evaluate the information obtained by mismatched decoders, we introduced an information theoretic quantity, *I, which was derived by extending the mutual information in terms of communication rate across a channel. We showed that *I** provides consistent results with the minimum mean-square error as well as the mutual information, and demonstrated that a previously proposed measure quantifying the importance of correlations in decoding substantially deviates from *I** when many cells are analyzed. We then applied this proposed framework to spike data for vertebrate retina using short natural scene movies of 100 ms duration as a set of stimuli and computing the information contained in neural activities. Although significant correlations were observed in population activities of ganglion cells, information loss was negligibly small even if all orders of correlation were ignored in decoding. We also found that, if we inappropriately assumed stationarity for long durations in the information analysis of dynamically changing stimuli, such as natural scene movies, correlations appear to carry a large proportion of total information regardless of their actual importance.**

Introduction

An ultimate goal of neuroscience is to elucidate how information is encoded and decoded by neural activities (Averbeck et al., 2006). One method of investigating the amount of information encoded about certain stimuli in a certain area of the brain is by calculating the mutual information between the stimuli and their neural responses. Because the mutual information quantifies the maximal amount of information that can be extracted from neural responses, it is implicitly assumed that encoded information is decoded by an optimal decoder. In other words, the brain is assumed to have full knowledge of the encoding process, in which stimuli are transformed into noisy neural activities. Considering the probable complexity of optimal decoding, however, the assumption of an optimal decoder in the brain is doubtful; rather, it is more plausible to consider that information is decoded in a suboptimal manner by a simplified decoder that has only partial knowledge of the encoding process. We call this type of a decoder a "mismatched decoder."

An example of a mismatched decoder is an independent decoder, which ignores correlations in neural activities. Independent decoders are potentially important because they are simpler, and the brain might use them rather than take on the task of trying to figure out what the correlation structure in the responses is. An experimental finding that a sufficiently large proportion of total information is obtained by an independent decoder would suggest that the brain may function in a manner similar to an independent decoder. In this context, Nirenberg et al. (2001) computed the information obtained by an independent decoder in pairs of retinal ganglion cells activities and found that no pair of cells showed a loss of information $>11\%$. However, their analysis considered pairs of cells only, and the importance or otherwise of correlations in population activities has not been fully elucidated.

Here, we developed a general framework for investigating the importance of correlations in population activities. Because analysis of population activities generally requires consideration of not only second-order but also higher-order correlations, we hierarchically constructed simplified decoders that ignore more than *K*th-order correlations using the maximum entropy method (Schneidman et al., 2006). We inferred how many orders of correlation should be taken into account to extract sufficient information by evaluating the information obtained by the simplified decoders. To accurately quantify information obtained by the mismatched decoders, we introduce an information theoretic quantity derived in the study by Merhav et al. (1994), *I**. *I** was first introduced in neuroscience in the study by Latham and Nirenberg (2005) to show that the previously proposed information for mismatched decoders in the study by Nirenberg and Latham (2003) is the lower bound of *I**.

Here, we showed that this lower bound can be loose when many cells are analyzed. We also justified the use of *I** from the viewpoint of the minimum mean-square error. Finally, we quantitatively evaluated the importance of correlations in decoding neural activities by applying our theoretical framework to the vertebrate retina.

Received Aug. 19, 2009; revised Jan. 26, 2010; accepted Jan. 29, 2010.

This work was partially supported by Grants-in-Aid for Scientific Research 18079003, 20240020, and 20650019 from the Ministry of Education, Culture, Sports, Science and Technology of Japan (M. Okada). M. Oizumi was supported by Grant-in-Aid 08J08950 for Japan Society for the Promotion of Science Fellows.

Correspondence should be addressed to Masato Okada, University of Tokyo, 701 Kiban-tou, Kashiwa-no-ha, Kashiwa, 277-8561 Chiba, Japan. E-mail: okada@k.u-tokyo.ac.jp.

DOI:10.1523/JNEUROSCI.4360-09.2010

Part of this paper was published in the study by Oizumi et al. (2009).

Materials and Methods

Retinal recording. Details of retinal recording have been described previously (Meister et al., 1994). The dark-adapted retina of a larval tiger salamander was isolated in oxygenated Ringer's medium at 25°C. A piece of retina (2–4 mm) was mounted on a flat array of 61 microelectrodes (MED-P2H07A; Panasonic) and perfused with oxygenated Ringer's solution (2 ml/min; 25°C). Six thousand frames of a movie of natural scenes (van Hateren, 1997) were projected at 30 Hz using a cathode ray tube monitor (60 Hz refresh rate; Dell E551). The mean intensity of light was 4 mW/m^2. Voltages from the electrodes were amplified, digitized, and then stored. Well isolated action potentials were sorted off-line with custom-built software. All procedures concerning animals met the RIKEN guidelines.

Information for mismatched decoders. It is well known that neural responses, even to a single repeated stimulus, are noisy and stochastic. Let us represent this stochastic process with the conditional probability distribution $p(\mathbf{r}|s)$, namely that neural responses $\mathbf{r}$ are evoked by stimulus s. We can say that the stimulus s is encoded by neural response $\mathbf{r}$, which obeys the distribution $p(\mathbf{r}|s)$. We call this $p(\mathbf{r}|s)$ the "encoding model." For the brain to function properly, the brain has to somehow accurately infer what stimulus is presented from the observation of noisy neural responses. We call this inference process the decoding process. To date, we have not known how stimulus information is decoded from noisy neural responses in the brain. Thus, when we investigate neural coding problems, we usually simply consider the limit of decoding accuracy assuming that stimulus information is decoded in an optimal way. Optimal decoding can be done by choosing the stimulus that maximizes the Bayes posterior probability,

$$p(s|\mathbf{r}) = \frac{p(\mathbf{r}|s)p(s)}{p(\mathbf{r})}, \qquad (1)$$

where $p(\mathbf{r}) = \sum_s p(\mathbf{r}|s)p(s)$ and $p(s)$ is the prior probability of stimuli. The mutual information invented by Shannon (1948) is one such quantity that provides the upper bound of decoding accuracy. The mutual information between stimulus s and neural responses $\mathbf{r}$ is given by the following equation:

$$I = -\sum_{\mathbf{r}} p(\mathbf{r}) \log_2 p(\mathbf{r}) + \sum_{\mathbf{r}} \sum_{s} p(s)p(\mathbf{r}|s) \log_2 p(\mathbf{r}|s). \qquad (2)$$

If we experimentally obtain the conditional probability distribution $p(r|s)$, we can easily quantify how accurately the stimulus is decoded from the noisy neural responses with the mutual information.

The mutual information is a useful indicator, which quantitatively shows how much the neural responses are related to the target stimuli. However, it is not evident whether this quantity is biologically relevant because it is implicitly assumed that information about stimuli is optimally decoded in the brain. Taking account of the complexity of optimal decoding and the difficulty of the brain in knowing the actual encoding process $p(\mathbf{r}|s)$, it is more plausible to consider that information about stimuli is decoded in a suboptimal manner in the brain. Let us assume that the brain has only the partial knowledge of the encoding process $p(\mathbf{r}|s)$ and denote the probability distribution that partially matches $p(\mathbf{r}|s)$ by $q(\mathbf{r}|s)$. For instance, if we assume that the brain does not know the complicated correlation structure in neural responses but rather only knows the individual property of neural responses of each neuron, $q(\mathbf{r}|s)$ is expressed by the product of the marginal distribution of $p(\mathbf{r}|s)$, $q(\mathbf{r}|s) = \prod_i p(r_i|s)$.

Here, the important question is how accurately the stimulus is inferred from neural responses only with the partial knowledge of $p(r|s)$. In this case, we assume that the inference is done by choosing the stimulus which maximizes the following posterior probability distribution as follows:

$$q(s|\mathbf{r}) = \frac{q(\mathbf{r}|s)p(s)}{q(\mathbf{r})}, \qquad (3)$$

where $q(\mathbf{r}) = \sum_s q(\mathbf{r}|s)p(s)$. This posterior probability distribution is not equal to the actual distribution (Eq. 1) because $q(\mathbf{r}|s)$ is used instead of the actual encoding model $p(\mathbf{r}|s)$. We call $q(\mathbf{r}|s)$ the "decoding model." When the decoding model $q(\mathbf{r}|s)$ is mismatched with the actual encoding model $p(\mathbf{r}|s)$, the accuracy of the decoding is naturally degraded.

To quantify how much stimulus information would be lost because of the mismatch in the decoding model, we need an information theoretic quantity, which corresponds to the mutual information when the mismatched decoding model is used. Nirenberg and Latham (2003) previously proposed that the information obtained by mismatched decoders can be evaluated using the following:

$$I^{\mathrm{NL}} = -\sum_{\mathbf{r}} p(\mathbf{r}) \log_2 \sum_{s} p(s)q(\mathbf{r}|s) + \sum_{\mathbf{r}} \sum_{s} p(s)p(\mathbf{r}|s) \log_2 q(r|s). \qquad (4)$$

We call their proposed information "Nirenberg–Latham information." By comparing Equations 2 and 4, we can see that I^{NL} is equal to I when the decoding model $q(\mathbf{r}|s)$ is equal to the encoding model $p(\mathbf{r}|s)$. To derive I^{NL}, they adopted the yes/no-question formulation of mutual information given by Cover and Thomas (1991). By extending the mutual information in the yes/no-question framework, they derived I^{NL}. Using a different approach, Pola et al. (2003) derived I^{NL} by decomposing the mutual information. Amari and Nakahara (2006) justified the use of I^{NL} for quantifying the information obtained by mismatched decoding from the point of view of information geometry. Because I^{NL} is easy to understand and appears sound, it has been used in neuroscience (Nirenberg et al., 2001; Golledge et al., 2003; Montani et al., 2007). However, as is shown in Appendix, I^{NL} may be an inappropriate measure, particularly when computed in large neural populations.

In the present study, we reintroduce an information theoretic quantity, I^*, which was originally derived by Merhav et al. (1994) by extending the mutual information in the context of the best achievable communication rate when a mismatched decoding model is used (see the next section for the information theoretic meaning of I^*). We call this quantity "information for mismatched decoders." In the present study, we use I^* to quantify the decoding accuracy when the stimulus information is decoded by using mismatched probabilistic models of neural responses. I^* can be computed by the following equations [for the details of the mathematical derivation of I^*, see Merhav et al. (1994) and Latham and Nirenberg (2005)]:

$$I^*(\mathbf{r}; s) = \max_{\beta} \tilde{I}(\beta), \qquad (5)$$

$$\tilde{I}(\beta) = -\sum_{\mathbf{r}} p(\mathbf{r}) \log_2 \sum_{s} p(s)q(\mathbf{r}|s)^{\beta} + \sum_{\mathbf{r}} \sum_{s} p(s)p(\mathbf{r}|s) \log_2 q(\mathbf{r}|s)^{\beta}. \qquad (6)$$

To compute I^*, we need to maximize $\tilde{I}(\beta)$ with respect to β. Thus, the equations for I^* have no closed-form solution. However, we can easily find the maximum of $\tilde{I}(\beta)$ numerically by the standard gradient ascent method because this is convex optimization (Latham and Nirenberg, 2005).

By comparing Equations 4 and 6, we can see that I^{NL} is equal to $\tilde{I}(\beta)$ when $\beta = 1$ (Latham and Nirenberg, 2005). Because I^* is the maximum value of $\tilde{I}(\beta)$ with respect to β, I^* is always larger than or equal to I^{NL}. Thus, I^{NL} is a lower bound of I^*. I^* was first introduced into neuroscience in Latham and Nirenberg (2005) to show that their proposed information, I^{NL}, provides a lower bound on I^*. To our knowledge, however, no application of I^* in neuroscience has appeared.

As is shown in Appendix, this lower bound provided by I^{NL} can be loose, and can be negative when many cells are analyzed. It is also shown that I^* gives consistent results with the minimum mean-square error, whereas I^{NL} does not. Taking account of these facts, we consider that I^* should be used instead of I^{NL}.

Information theoretic meaning of information I *and* I*. In the previous section, we introduced mutual information as a measure that quantifies how accurately a stimulus is inferred from the observation of noisy neural responses. In information theory, the mutual information has a rig-

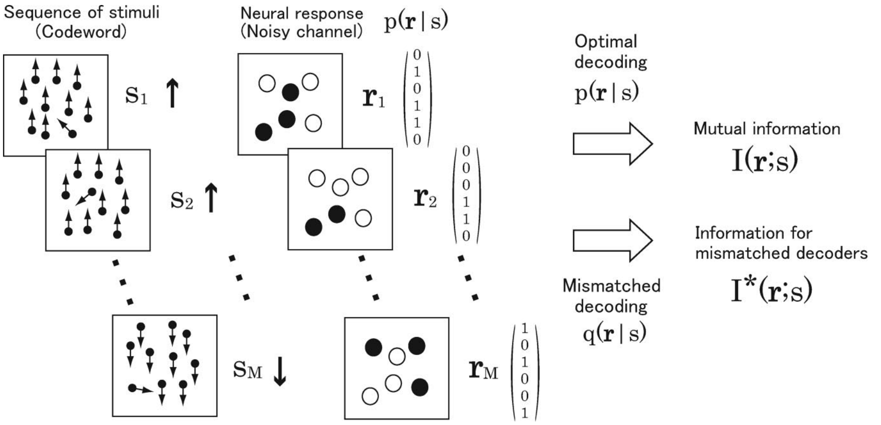

Figure 1. Information transmission using stimulus s and neural responses $\mathbf{r}$ as symbols when the neural population is considered as a noisy channel. Random-dot stimuli moving upward or downward are considered. Code words encoded with the sequence of stimuli, for example, $s_1 s_2 \ldots s_M = \uparrow\ \uparrow \cdots \downarrow$, are sent and neural responses of the six neurons to each stimuli, $\mathbf{r}_1 \mathbf{r}_2 \ldots \mathbf{r}_M$, are received. Neural responses $\mathbf{r}$ are binary, either firing ("1") (filled circles) or silent ("0") (open circles). The neurons stochastically fire in response to each stimulus s according to the conditional probability distribution $p(\mathbf{r}|s)$. The receiver infers which code word is sent from the received neural responses (decoding). When decoding is performed using the actual probability distribution $p(\mathbf{r}|s)$, the maximum number of code words which can be sent error-free is quantified by the mutual information $I(\mathbf{r};s)$ (Eq. 2). In contrast, when decoding is performed using a mismatched probability distribution $q(\mathbf{r}|s)$, the maximum number of code words which can be sent error-free is quantified by the information for mismatched decoders $I^*(\mathbf{r}; s)$ (Eqs. 3, 4).

orous quantitative meaning [i.e., it gives the upper bound of the amount of information that can be reliably transmitted over a noisy channel (see below)]. In this section, we first review the meaning of the mutual information I within the framework of information theory using the language of neuroscience. We then explain the meaning of I^* as an extension of the mutual information.

Let us consider information transmission using a set of stimuli s and neural responses $\mathbf{r}$ (Fig. 1). We will consider a random-dot stimulus moving upward, ↑, or downward, ↓, as an example of stimulus s. The sequence of stimuli $s_1 s_2 \ldots s_M$ (Fig. 1, $\uparrow\ \uparrow \ldots \downarrow$) is sent over a noisy channel, which in this case is a neural population, and the sequence of noisy neural responses $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$ to each stimulus is then received (Fig. 1). We assume that the channel is memoryless; that is, the neural responses $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$ are mutually independent. This sequence of stimuli is called a code word. We consider the limit that the length of code word M tends to infinity, $M \to \infty$.

Here, we introduce an important concept, the codebook. A codebook is the assembly of transmitted code words. The sender and the receiver share the codebook. The job of the receiver is to determine which code word was sent from observed neural responses $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$ by consulting the codebook. In this setting, let us consider the following question: How many code words can be sent error-free when the transmitted code words are "optimally" decoded? In other words, how many code words can the codebook contain?

Optimal decoding is done by choosing a code word that maximizes Bayes posterior probability given by the observed sequence of neural responses $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$ from the codebook. The decoding procedure is described by the following equations:

$$p(c|\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M) = \frac{p(\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M|c)p(c)}{\sum_{c'} p(\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M|c')p(c')}, \tag{7}$$

$$p(\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M|c = s_1(c)s_2(c) \ldots s_M(c)) = \textstyle\prod_i p(\mathbf{r}_i|s_i(c)), \tag{8}$$

$$\hat{c}_{\mathrm{optimal}} = \operatorname{argmax} p(c|\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M), \tag{9}$$

where $s_i(c)$ means the ith stimulus of the sequence of stimuli corresponding to code word c. A uniform prior distribution on c is usually assumed, in which case Equation 7 becomes the maximum-likelihood estimation.

If stimuli ↑ and ↓ evoke nonconfusable neural responses, 2^M code words can be sent error-free. However, when there is an overlap between neural responses to stimuli ↑ and ↓, the question "How many code words can be sent error-free?" is not easily answered. In this case, we cannot let our codebook contain all of possible 2^M code words but rather need to sparsely select the transmission of some of them so as to avoid confusable neural responses to each code word. Shannon's mutual information gives the answer to this nontrivial question (Shannon, 1948). If we denote the upper bound of the number of code words that can be sent error-free by 2^K, K is given by the following:

$$K/M = I(\mathbf{r}; s), \tag{10}$$

where I is the mutual information given by Equation 2. This relationship can be mathematically proved by taking advantage of the law of large numbers (Shannon, 1948; Cover and Thomas, 1991). The ratio K/M is called the communication rate or information rate. Thus, within the framework of information theory, the mutual information defined by Equation 2 has the meaning of the upper bound of communication rate (i.e., the number of code words that can be sent error-free).

When we have full knowledge of the channel property, $p(\mathbf{r}|s)$, the mutual information gives the upper bound of the number of code words that can be sent error-free. The next question is how many code words can be sent error-free when we only partially know the channel property. In other words, we assume that the mismatched probability distribution $q(\mathbf{r}|s)$, which partially matches with the actual channel property $p(\mathbf{r}|s)$, is used for decoding. Similarly to Equations 7–9, decoding is done by the following equations:

$$q(c|\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M) = \frac{q(\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M|c)p(c)}{\sum_{c'} q(\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M|c')p(c')}, \tag{11}$$

$$q(\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M|c = s_1(c)s_2(c) \ldots s_M(c)) = \textstyle\prod_i q(r_i|s_i(c)), \tag{12}$$

$$\hat{c}_{\mathrm{optimal}} = \operatorname{argmax} q(c|\mathbf{r}_1\mathbf{r}_2 \ldots \mathbf{r}_M). \tag{13}$$

Note that $q(\mathbf{r}|s)$ is used instead of $p(\mathbf{r}|s)$. Merhav et al. (1994) provided the answer to this question: if we denote the upper bound of the number of code words that can be sent error-free by 2^{K^*} when the mismatched decoding model $q(\mathrm{r}|s)$ is used, K^* ($<K$) is given by the following:

$$K^*/M = I^*(\mathbf{r}; s), \tag{14}$$

where I^* is information for mismatched decoders given by Equations 5 and 6. This relationship can be also mathematically proved by making use of the large deviation theory (Merhav et al., 1994; Latham and Nirenberg, 2005). Thus, I^* gives the upper bound of the number of code words that can be sent error-free for mismatched decoders. In this sense, I^* is a natural extension of the mutual information I.

Stationarity assumption about neural responses. We used a movie of natural scenes, which was 200 s long and repeated 45 times, as a stimulus. We divided the movie into many short segments as is shown in Figure 2 and considered them as stimuli over which information contained in neural activities was computed. We assumed that neural responses were stationary while each short natural scene movie was presented. Thus, the length of each stimulus should be short enough for us to assume the stationarity of neural responses. To determine the appropriate length of the stimuli, we computed the correlation coefficients between the temporally separated frames of the natural scene movie. The correlation coefficient between two frames separated by time τ, $C(\tau)$, is computed by the following:

$$C(\tau) = \frac{1}{T}\sum_{t=1}^{T} \frac{(\boldsymbol{x}(t) - \langle \boldsymbol{x} \rangle) \cdot (\boldsymbol{x}(t+\tau) - \langle \boldsymbol{x} \rangle)}{|(\boldsymbol{x}(t) - \langle \boldsymbol{x} \rangle)|^2}, \tag{15}$$

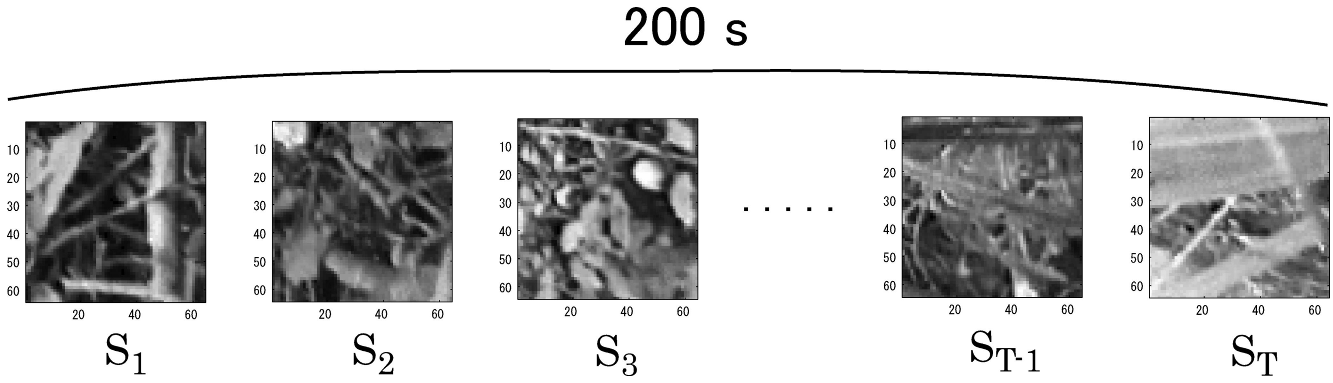

Figure 2. Schematic of a set of stimuli over which mutual information was computed. Each short segment, extracted from a movie of natural scenes of 200 s duration, $s_1, s_2, s_3, \ldots, s_{T-1}, s_T$, was considered as one stimulus.

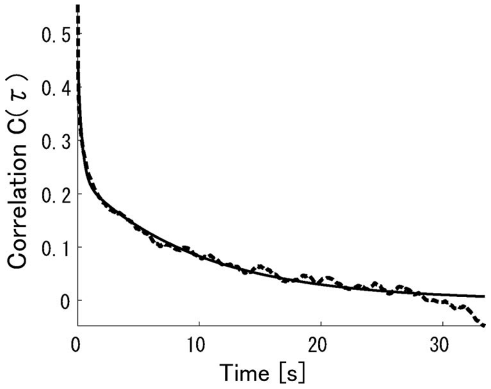

Figure 3. Correlation coefficients between the temporally separated frames of a natural scene movie (dashed line). The solid line is a least-squares fit. The fitted function is of the form $y(\tau) = a_1 \exp(-\tau/\tau_1) + a_2 \exp(-\tau/\tau_2)$.

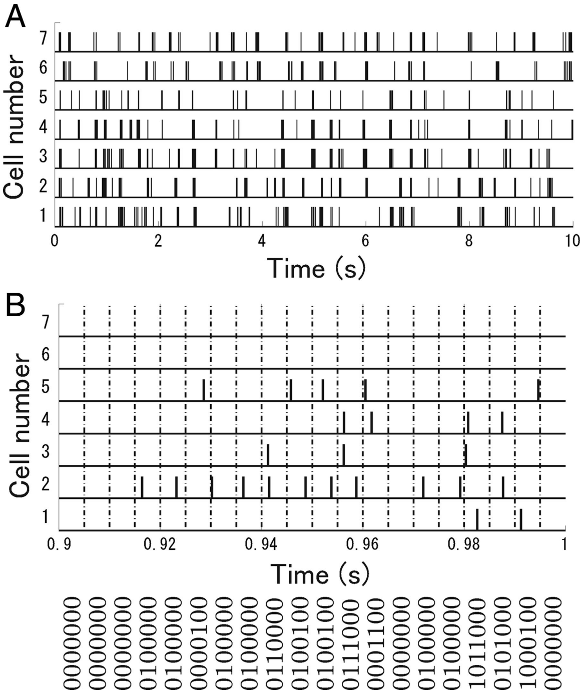

Figure 4. ***A***, Raster plot of seven retinal ganglion cells responding to a natural scene movie. ***B***, Transformation of spike trains into binary words.

where $\boldsymbol{x}(t)$ is the grayscale pixel value of the frame at time t and $\langle \boldsymbol{x} \rangle$ is the averaged pixel value of the frames over the total time of the natural scene movie. $C(\tau)$ is shown as a dotted line in Figure 3. $C(\tau)$ rapidly decays initially and then slowly approaches 0. We fit $C(\tau)$ with the sum of two exponents $y(\tau) = a_1 \exp(-\tau/\tau_1) + a_2 \exp(-\tau/\tau_2)$ by the least-squares method. The fitted line is shown as a solid line in Figure 3. The fitted time constants τ_1 and τ_2 are $\tau_1 = 332$ ms and $\tau_2 = 9.77$ s. This result indicates that the length of stimuli should be shorter than the faster time constant, $\tau_1 = 332$ ms.

Constructing mismatched decoding models by the maximum entropy method. Figure 4*A* shows the response of seven retinal ganglion cells to natural scene movies from 0 to 10 s in length. To apply information theoretic techniques, we first discretized the time into small time bins $\Delta\tau$ and indicated whether or not a spike was emitted in each time bin with a binary variable: $r_i = 1$ means that the cell i spiked and $r_i = 0$ means that it did not. We set the length of the time, $\Delta\tau$, to 5 ms so that it was short enough to ensure that two spikes did not fall into the same bin. In this way, the spike pattern of ganglion cells was transformed into an N-letter binary word, $\mathbf{r} = \{r_1, r_2, \ldots, r_N\}$, where N is the number of neurons (Fig. 4*B*). We then determined the frequency with which a particular spike pattern, $\mathbf{r}$, was observed during each stimulus and estimated the conditional probability distribution $p_{\text{data}}(\mathbf{r}|s)$ from experimental data. If we set the length of stimuli to 100 ms, there were, effectively, a total of 900 (=20 bins × 45 repeats) samples for estimating the conditional probability distribution $p_{\text{data}}(\mathbf{r}|s)$ of each stimulus because each 5 ms bin within the 100 ms segment was assumed to come from the same stimulus. Using these estimated conditional probabilities, we evaluated the information contained in N-letter binary words $\mathbf{r}$.

Generally, the joint probability of N binary variables can be written as follows (Amari, 2001; Nakahara and Amari, 2002):

$$p_N(\mathbf{r}) = \frac{1}{Z} \exp\left[\sum_i \theta_i r_i + \sum_{i<j} \theta_{ij} r_i r_j + \cdots + \theta_{12\ldots N} r_1 r_2 \ldots r_N \right]. \quad (16)$$

This type of representation of probability distribution is called a log-linear model. Because the number of parameters in a log-linear model is equal to the number of all possible configurations of an N-letter binary word $\mathbf{r}$, we can determine the values of parameters so that the log-linear model $p_N(\mathbf{r})$ exactly matches the empirical probability distribution $p_{\text{data}}(\mathbf{r})$: that is, $p_N(\mathbf{r}) = p_{\text{data}}(\mathbf{r})$.

To compute the information for mismatched decoders, we constructed simplified probabilistic models of neural responses that partially match the empirical distribution, $p_{\text{data}}(\mathbf{r})$. The simplest model was an "independent model," $p_1(\mathbf{r})$, in which only the average of each r_i agreed with the experimental data: that is, $\langle r_i \rangle_{p1}(\mathbf{r}) = \langle r_i \rangle_{p\text{data}}(\mathbf{r})$. Many possible probability distributions satisfied these constraints. In accordance with the maximum entropy principle (Jaynes, 1957;

Schneidman et al., 2003, 2006), we chose the one that maximized entropy H, $H = -\sum_{\mathbf{r}} p_1(\mathbf{r}) \log p_1(\mathbf{r})$.

The resulting maximum entropy distribution is as follows:

$$p_1(\mathbf{r}) = \frac{1}{Z_1} \exp\left[\sum_i \theta_i^{(1)} r_i\right], \tag{17}$$

in which model parameters $\theta^{(1)}$ are determined so that the constraints are satisfied. This model corresponds to a log-linear model in which all orders of correlation parameters $\{\theta_{ij}, \theta_{ijk}, \ldots, \theta_{12\ldots N}\}$ are omitted. If we perform maximum-likelihood estimation of model parameters $\theta^{(1)}$ in the log-linear model, the result is that the average r_i under the log-linear model equals the average r_i found in the data: that is, $\langle r_i \rangle_{p1}(\mathbf{r}) = \langle r_i \rangle_{p\text{data}}(\mathbf{r})$ $\langle r_i r_j \rangle_{p2}(\mathbf{r}) = \langle r_i \rangle_{p\text{data}}(\mathbf{r})$. This result is identical with the constraints of the maximum entropy model. Generally, the maximum entropy method is equivalent to the maximum-likelihood fitting of a log-linear model (Berger et al., 1996).

Similarly, we can consider a "second-order correlation model" $p_2(\mathbf{r})$, which is consistent with not only the averages of r_i but also the averages of all products $r_i r_j$ found in the data. Maximizing the entropy with constraints $\langle r_i \rangle_{p2}(\mathbf{r}) = \langle r_i \rangle_{p\text{data}}(\mathbf{r}) \langle r_i r_j \rangle_{p2}(\mathbf{r}) = \langle r_i \rangle_{p\text{data}}(\mathbf{r})$ and $\langle r_i r_j \rangle_{p2}(\mathbf{r}) = \langle r_i r_j \rangle_{p\text{data}}(\mathbf{r})$ $\langle r_i r_j \rangle_{p2}(\mathbf{r}) = \langle r_i \rangle_{p\text{data}}(\mathbf{r})$, we obtain the following:

$$p_2(\mathbf{r}) = \frac{1}{Z_2} \exp\left[\sum_i \theta_i^{(2)} r_i + \sum_{i,j} \theta_{ij}^{(2)} r_i r_j\right], \tag{18}$$

in which model parameters $\theta^{(2)}$ are determined so that the constraints are satisfied.

The procedure described above can also be used to construct a "Kth-order correlation model" $p_K(\mathbf{r})$. If we substitute the simplified models of neural responses $p_K(\mathbf{r}|s)$ into mismatched decoding model $q(\mathbf{r}|s)$ in Equation 6, we can compute the amount of information that can be obtained when more than Kth-order correlations are ignored in the decoding as follows:

$$I_K^{\star} = \max_{\beta} \tilde{I}_K(\beta), \tag{19}$$

$$\tilde{I}_K(\beta) = -\sum_{\mathbf{r}} p_N(\mathbf{r}) \log_2 \sum_s p(s) p_K(\mathbf{r}|s)^{\beta} + \sum_s p(s) \sum_{\mathbf{r}} p_N(\mathbf{r}|s) \log_2 p_K(\mathbf{r}|s)^{\beta}. \tag{20}$$

By evaluating the ratio of information, $I_K^{\star}/I$, we can infer how many orders of correlation should be taken into account to extract sufficient information.

Limited sampling problem in estimating mutual information. It is well known that estimating mutual information in Equation 2 with a limited amount of neuronal data causes a sampling bias problem (Panzeri and Treves, 1996). With a small amount of data, the mutual information is biased upward. Recently, tight data-robust lower bounds to mutual information, I_{sh}, were developed (Montemurro et al., 2007). I_{sh} was derived using "shuffling," namely, the shuffling of neural responses across trials, to cancel out the upward bias of the mutual information. I_{sh} can be computed by the following equation:

$$I_{\text{sh}} = I_{\text{LB}-1} + \Delta I_{1-\text{sh}}, \tag{21}$$

$$I_{\text{LB}-1} = -\sum_{\mathbf{r}} p(\mathbf{r}) \log_2 \sum_s p(s) p_1(\mathbf{r}|s) + \sum_{\mathbf{r}} \sum_s p(s) p_1(\mathbf{r}|s) \log_2 p_1(\mathbf{r}|s), \tag{22}$$

$$\Delta I_{1-\text{sh}} = I + \sum_{\mathbf{r}} p(\mathbf{r}) \log_2 \sum_s p(s) p_1(\mathbf{r}|s) - \sum_{\mathbf{r}} \sum_s p(s) p_{1-\text{sh}}(\mathbf{r}|s) \log_2 p_{1-\text{sh}}(\mathbf{r}|s), \tag{23}$$

where I is the mutual information in Equation 2, and $p_1(\mathbf{r}|s)$ is the independent model, that is, $p_1(\mathbf{r}|s) = \prod_i p(r_i|s)$, and $p_{1-\text{sh}}(\mathbf{r}|s)$ is the distribution of shuffled neural responses. Using $p_{1-\text{sh}}(\mathbf{r}|s)$ instead of $p_1(\mathbf{r}|s)$ in Equation 23, the upward bias of I is canceled out by a downward bias of the third term of $\Delta I_{1-\text{sh}}$. As a result, $\Delta I_{1-\text{sh}}$ is mildly biased downward. Since $I_{\text{LB}-1}$ is virtually unbiased, I_{sh} is mildly biased downward. Using I in Equation 2 and I_{sh} in Equation 21, the real mutual information, I_{real}, is bounded upward and downward as follows:

$$I_{\text{sh}} < I_{\text{real}} < I. \tag{24}$$

We computed both I and I_{sh}. We found that the difference between I and I_{sh} was markedly small even when all recorded cells were analyzed (Fig. 5). This meant that we had a sufficient amount of data to accurately estimate the mutual information. Thus, in Results, we show the value of mutual information that is directly computed from Equation 2 only.

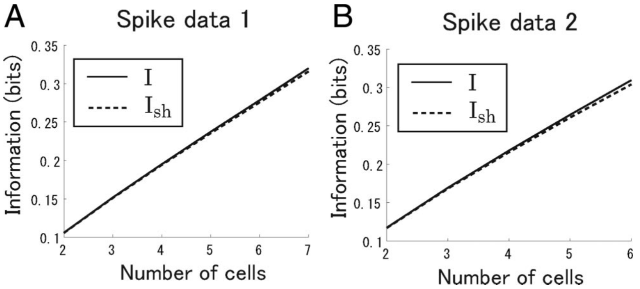

Figure 5. Difference between I (solid line) and I_{sh} (dashed line). Spike data and length of stimuli are the same as in Figure 9, *A1* and *A2*. I is the value of the mutual information that is directly computed from Equation 2. I_{sh} is computed using Equation 21. I provides the upper bound of the real value of the mutual information I_{real}, and I_{sh} provides the lower bound of I_{real}. In other words, $I_{\text{sh}} < I_{\text{real}} < I$. The difference between I and I_{sh} is markedly small even when all recorded cells ($N = 7$ in spike data 1; $N = 6$ in spike data 2) are analyzed. ***A***, Spike data 1. ***B***, Spike data 2.

Results

Information conveyed by correlated activities is negligibly small despite the presence of substantial correlations

We quantitatively evaluated the importance of correlated activities by comparing the mutual information I (Eq. 2) with the information for mismatched decoders $I_K^{\star}$ (Eqs. 19, 20). We considered the independent model $p_1(\mathbf{r})$ in Equation 17 and the second-order correlation model $p_2(\mathbf{r})$ in Equation 18 as mismatched decoders. We analyzed two spike data recorded from isolated retinas of different salamanders. Seven neurons were simultaneously recorded in spike data 1 and six in spike data 2. The same 200 s natural scene movie was used as a stimulus for spike data 1 and 2.

We computed the spike-triggered averages of all recorded neurons responding to the natural scene movie stimulus in spike data 1 and 2. The recorded cells were all OFF cells. The fits of two-dimensional Gaussian functions to the spike-triggered averages are shown in Figure 6. As can been seen, the receptive fields mostly overlapped in both spike data 1 and 2. Figure 7 shows cross-correlograms of all pairs of cells in spike data 1 and 2. Many pairs show strong peaks with a width of ~100 ms around the origin. To show the degree of correlation in the population activities of the retinal ganglion cells, we investigated how accurately the independent model and the second-order correlation model predicted the actual neural responses, following previous studies (Schneidman et al., 2006; Shlens et al., 2006). Figure 8 shows the observed frequency of N-letter binary words $\mathbf{r}$ against the predicted frequency of the independent model and the second-order correlation model. As can be seen from Figure 8, the independent model roundly failed to capture the observed statistics of firing patterns. The second-order correlation model substantially improved the prediction of the observed pattern rate. We therefore consider that correlations need to be taken into account to explain the observed neural responses. However, this does not necessarily mean that they need to be taken into account in decoding neural activities (see Discussion).

We computed the ratio of information obtained by independent model, I_1^*/I, and that obtained by a second-order correlation model, I_2^*/I. Considering the decay speed $\tau_1 = 332$ ms of the correlations between the frames of the natural scene movie (see Materials and Methods), we set the length of stimuli to 100 ms, providing 2000 stimuli from the 200 s movie. With a uniform stimulus length of 100 ms, no spikes occurred when some stimuli were presented. We removed these stimuli and used the remaining stimuli for analysis. Figure 9*A* shows I_1^*/I and I_2^*/I when the number of cells analyzed was changed. Although I_1^*/I decreased slightly as the number of cells analyzed increased, I_1^*/I was $>$90% in both spike data 1 and 2 even when all cells ($N = 7$ in spike data 1 and $N = 6$ in spike data 2) were analyzed. This result means that the loss of information associated with ignoring correlations was minor.

We computed the mutual information between all stimuli and neural responses. In terms of average, the percentage of information conveyed by correlations was low. However, it is possible that correlations play an important role in discriminating some stimuli. To test this possibility, we computed I and I_1^* for pairs of 100 ms natural scene movie stimuli selected from all stimuli. Figure 10 shows the histogram of I_1^*/I when all recorded cells were analyzed. I_1^*/I was $>$90% for $\sim$95% of pairs of stimuli. Pairs whose correlations carried a large proportion of total information were extremely rare. This result also supports the idea that almost all stimulus information could be extracted even if correlations were ignored in decoding.

An important point is that the amount of information conveyed by correlations was markedly small (Figs. 9*A*, 10) even though there were significant correlations in population activities of ganglion cells (Figs. 7, 8). This result shows that, to assess the importance of correlations in information processing in the brain, we should not only evaluate the degree by which the actual neural responses differ from the independent model but should also compute the information obtained by the independent model, I_1^*/I.

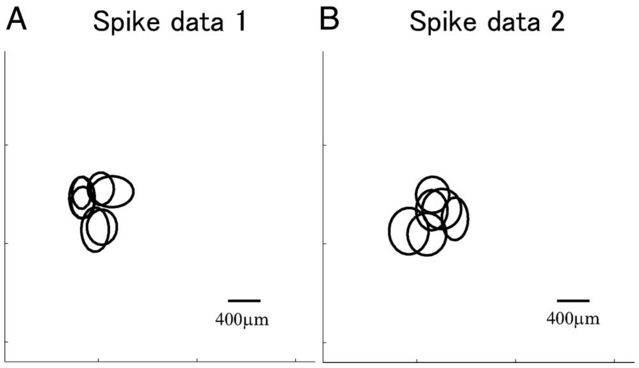

Figure 6. ***A***, ***B***, Receptive fields of seven OFF cells in spike data 1 (***A***) and six OFF cells in spike data 2 (***B***). Ellipses represent 1 SD of the Gaussian fit to the spatial profile of the spike-triggered averages measured from the natural scene movie stimulus.

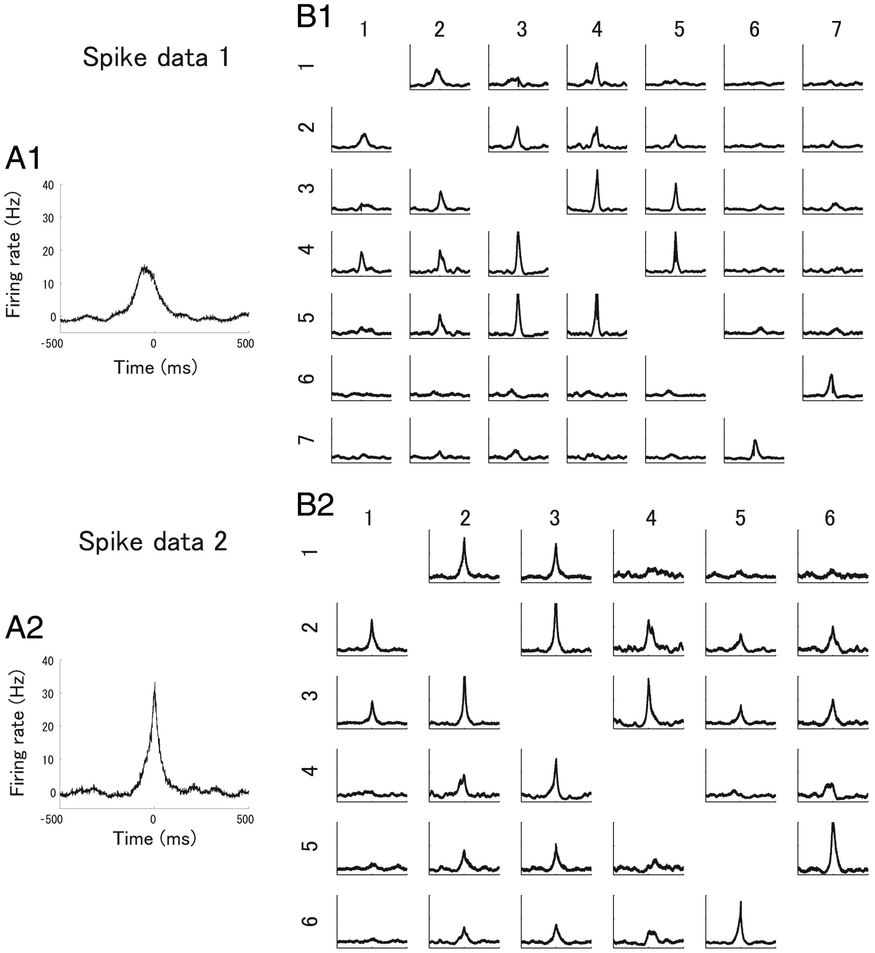

Figure 7. Synchronous firing in a population of retinal ganglion cells. ***A1***, ***A2***, Example cross-correlograms in spike data 1 (***A1***) and spike data 2 (***A2***) showing the firing rate of one cell when the time difference between spikes of one cell and the other cell is given. Mean firing rates are subtracted so that the vertical axis shows the excess firing rate from the baseline firing rate. ***B1***, ***B2***, Cross-correlograms of all pairs of recorded cells in spike data 1 (***B1***) and spike data 2 (***B2***). The range of the vertical and horizontal axes is the same as that in the example cross-correlograms in ***A1*** and ***A2***.

Pseudo-importance of correlations arising from stationarity assumption about neural responses

We also computed I_1^*/I and I_2^*/I when the length of stimuli was set to 10 s to see what happens if the stimulus length is made considerably longer than the time constant of the stimulus autocorrelation, $\tau_1 = 332$ ms. Figure 9*B* shows I_1^*/I and I_2^*/I when the length of stimuli was set to 10 s. When only two cells were considered, I_1^*/I exceeded 90%, which means that, consistent with the result obtained by Nirenberg et al. (2001), ignoring correlation leads to only a small loss of information. However, when all cells were used in the analysis, I_1^*/I was only $\sim$60% with both spike data 1 and 2. Thus, we reached different conclusions when the length of stimuli was set to 10 s from those when it was 100 ms. This is

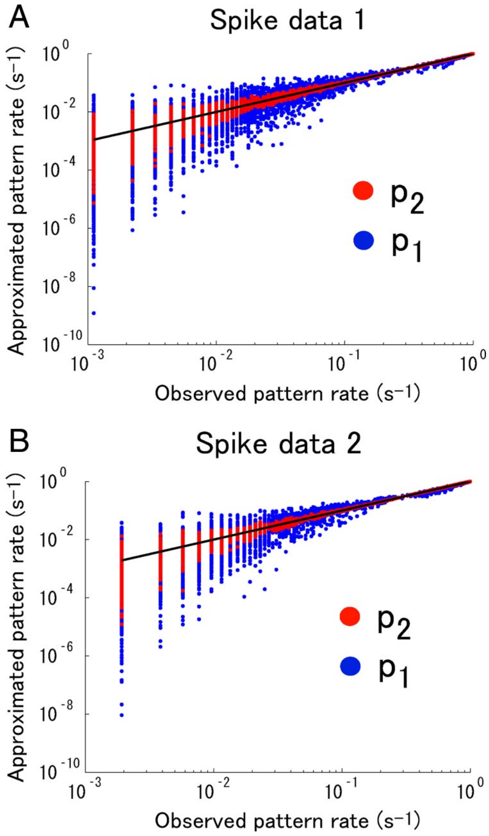

Figure 8. Relationship between the observed frequency of firing patterns and the predicted frequency of firing patterns from an independent model p_1 (blue dots) and second-order correlation model p_2 (red dots) constructed using the maximum entropy method. Natural scene movies of 100 ms duration were used as stimuli. $p_K(\mathbf{r}|s)$ $(K = 1, 2)$ for all stimuli are plotted against $p_{\text{data}}(\mathbf{r}|s)$. The black line shows equality of the observed frequency and the predicted frequency of firing patterns. ***A***, Spike data 1. ***B***, Spike data 2.

because 10 s is too long to be considered as one stimulus during which neural responses are stationary, that is, during which neural responses obey the same conditional probability distributions $p(\mathrm{r}|s)$. If we assume stationarity when neural responses are not in fact stationary, correlations may carry a large proportion of information that is irrelevant to the actual importance of correlated activities.

Figure 11 shows $I_1^{\star}/I$ and $I_2^{\star}/I$ when the duration of stimuli was changed. When the length of stimuli is appropriately set, >90% of information can be extracted even if correlations are ignored in decoding neural activities. However, when the length of stimuli is too long, correlations appear to carry a large proportion of total information because of the stationarity assumption about neural responses.

To clarify why the correlation becomes less important as the stimulus is shortened, we used the toy model shown in Figure 12. We considered the case in which two cells fire independently in accordance with a Poisson process and performed an analysis similar to that for the actual spike data. We used simulated spike data for the two cells generated in accordance with the firing rates shown in Figure 12*A*. Firing rates with a 2 s stimulus sinusoidally changed with time. We divided the 2 s stimulus into two 1 s stimuli, s_1 and s_2, as shown in Figure 12*B*. We then computed mutual information I and the information obtained by independent model $I_1^{\star}$ over s_1 and s_2. Because the two cells fired independently, there were essentially no correlations between them. However, pseudocorrelation arose because of the assumption of stationarity for the dynamically changing stimulus. The pseudocorrelation was high for s_1 and low for s_2. In contrast to the difference in the degree of "correlation" between the two stimuli, s_1 and s_2, the mean firing rates of the two cells during each stimulus were equal. If the stimulus is 1 s long, therefore, we cannot discriminate two stimuli using the independent model, namely $I_1^{\star} = 0$. This implies that, when the stationarity of neural responses is assumed for long durations, correlations could carry a large proportion of total information irrespective of its actual importance.

We also considered the case in which the stimulus was 0.5 s long, as shown in Figure 12*C*. In this case, pseudocorrelations again appeared, but there was a significant difference in mean firing rates between the stimuli. Thus, the independent model could be used to extract almost all the information. The dependence of $I_1^{\star}/I$ on stimulus length is shown in Figure 11*C*. Behaviors similar to those in this figure were also observed in analysis of the actual spike data for retinal ganglion cells (Fig. 11*A*,*B*). Even if we observe that correlation carries a significantly larger proportion of information for longer stimuli compared with the speed of change in the firing rates, this may simply have resulted from meaningless correlation. Thus, to assess the role of correlation in information processing, the stimuli used should be sufficiently short that the neural responses to these stimuli can be considered to obey the same probability distribution. Considering the response speed of retinal ganglion cells, 100 ms, to which we set the stimulus length in the present study, is still not short enough for the stationarity assumption. However, we kept the stimulus length equal to or longer 100 ms to ensure sufficient data to allow the mutual information to be reliably estimated. If the stimulus length is shortened, the ratio of information carried by correlations could be smaller, as suggested by the analysis in this section (Fig. 11*C*).

Comparison between I^{NL} and $I^{\star}$

In Appendix, we show a simple example in which the difference between I^{NL} and $I^{\star}$ is large particularly when many cells are analyzed. To see the difference between I^{NL} and $I^{\star}$ in the actual spike data, we computed I_1^{NL}, which corresponds to the information obtained by the independent decoder, $I_1^{\star}$. The dot-dashed lines in Figure 9 plot I_1^{NL}. Although the difference between I^{NL} and $I^{\star}$ increases slightly as the number of cells analyzed increases, the lower bound of $I_1^{\star}$ provided by I_1^{NL} was relatively tight, even when all recorded cells were analyzed. These results suggest that the values of I^{NL} previously reported in the analysis of pair of cells were also probably close to $I^{\star}$ (Nirenberg et al., 2001; Golledge et al., 2003).

Discussion

Here, we describe a general framework for investigating to what extent the decoding process in the brain can be simplified. In this framework, we first constructed a simplified decoding model (i.e., mismatched decoding model), using the maximum entropy method. We then computed the amount of information that can be extracted using the mismatched decoders. We introduced the information for mismatched decoders, $I^{\star}$, which was derived in terms of communication rate in information theory (Merhav et al., 1994). By analytical computations, we showed that both the mutual information I and the information for mismatched decoders $I^{\star}$ are inversely proportional to the minimum mean-square error under the condition that neural responses obey

Gaussian statistics. We also pointed out that the difference between the previously proposed information I^{NL} (Nirenberg and Latham, 2003) and I^* may become large when many cells are analyzed. By using the information theoretic quantity I^*, we showed that >90% of the information encoded in population activities of retinal ganglion cells can be decoded even if all orders of correlation are ignored in decoding. Our results imply that the brain uses a simplified decoding strategy in which correlation is ignored.

Below, we discuss differences between the present and previous studies using the maximum entropy approach (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008); limitations and extensions of the methodology used in this work; and future directions, which concern animal behavior experiments (Stopfer et al., 1997; Ishikane et al., 2005).

Presence of significant correlated activities does not necessarily mean the importance of correlations in decoding

Previous studies using the maximum entropy approach (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008) emphasized the discrepancy between the independent model and actual probability distribution. That is, their results show that there are significant correlations in large neural populations. The impact of such significant correlated neural activities on information encoding has been recently addressed (Montani et al., 2009). In the present study, we addressed how important the correlations are in information decoding. Our results indicate that, even if the independent model fails to capture the statistics of population activities, it does not necessarily mean that correlations play an important role in extracting information about stimuli. Assume that we experimentally obtained the probability distribution of neural responses to two different stimuli, $p_{\mathrm{data}}(\mathbf{r}|s_1)$ and $p_{\mathrm{data}}(\mathbf{r}|s_2)$, respectively. Even when the independent models of two stimuli, $p_1(\mathbf{r}|s_1)$ and $p_1(\mathbf{r}|s_2)$, mostly deviate from the data distribution $p_{\mathrm{data}}(\mathbf{r}|s_1)$ and $p_{\mathrm{data}}(\mathbf{r}|s_2)$, if the two independent models $p_1(\mathbf{r}|s_1)$ and $p_1(\mathbf{r}|s_2)$ are significantly different from each other, correlations are not important in decoding neural activities. In fact, the information conveyed by correlated activity in our analysis represented only 10% of the total, albeit that we observed a large deviation in the independent model from the data distribution in our spike data, as in previous studies (Fig. 8). As shown in Figure 8, the independent model fails disastrously in predicting the actual probability distribution. However, the second-order correlation model considerably improves the fitting accuracy of the actual probability distribution, as was shown in the previous studies (Schneidman et al., 2006; Shlens et al., 2006). If we consider only the discrepancy between the independent model and the actual probability distribution (Fig. 8), we may mistakenly conclude that correlations play an important role in information processing in the brain. To assess the importance of correlations, we rather need to evaluate the difference between the mutual information and the information obtained by simplified probabilistic models I^*, as was done in the present study.

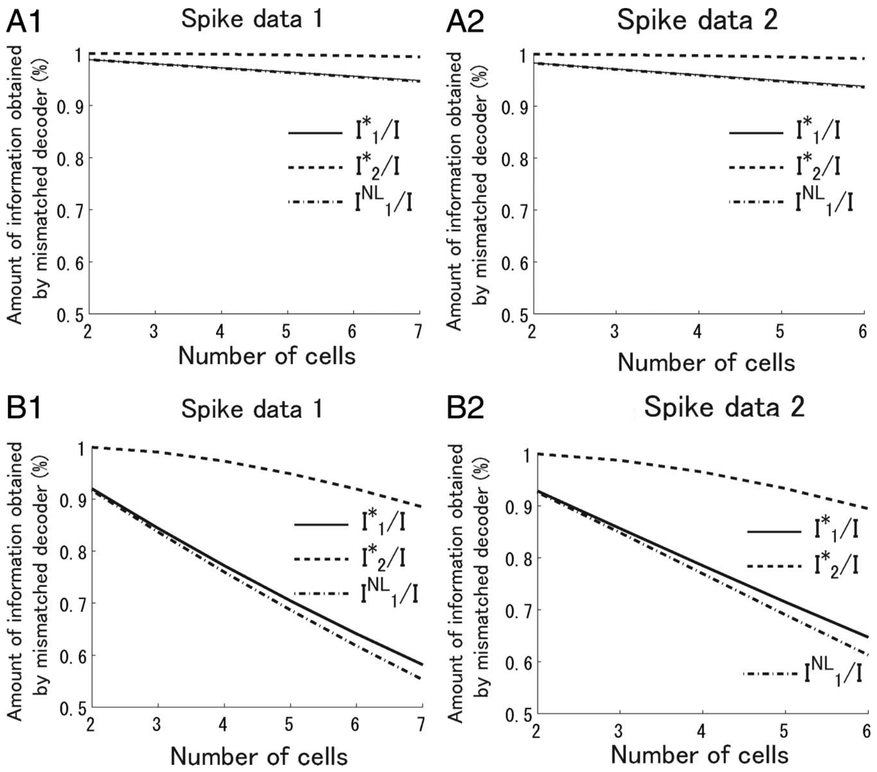

Figure 9. Dependence of the amount of information obtained by simplified decoders on the number of ganglion cells analyzed. The average values of I^*_K for $K = 1, 2$ over all possible combinations of recorded cells is shown when the number of cells analyzed is given. Spike data 1 is used in ***A1*** and ***B1***, and spike data 2 in ***A2*** and ***B2***. ***A1***, ***A2***, A natural scene movie of 100 ms duration was considered as the stimulus. ***B1***, ***B2***, A natural scene movie of 10 s duration was considered as the stimulus.

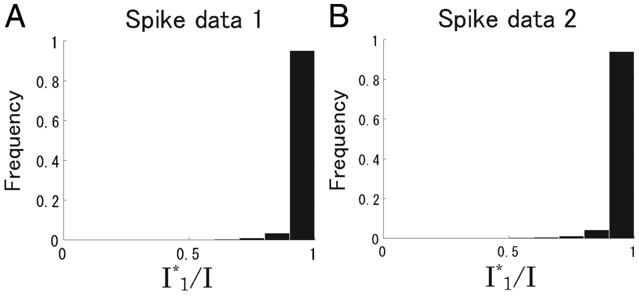

Figure 10. Histogram of I^*_i/I. All recorded cells ($N = 7$ in spike data 1; $N = 6$ in spike data 2) were analyzed. ***A***, Spike data 1. ***B***, Spike data 2.

Temporal correlations across time bins

In this study, we focused on synchronous firing within one time bin, on the basis of suggestions that synchronous firing has functional importance (Gray et al., 1989; Meister et al., 1995; Meister, 1996; Stopfer et al., 1997; Dan et al., 1998; Perez-Orive et al., 2002; Ishikane et al., 2005), and spike timing-based computations taking advantage of synchronous firing can be implemented in a biologically relevant network architecture (Hopfield, 1999; Brody and Hopfield, 2003). Given previous findings that neurons carry substantial sensory information in their response latencies (Panzeri et al., 2001; Reich et al., 2001; Gollisch and Meister, 2008), consideration of temporal correlations across the time bins may be important. Statistical models that take account of time-lagged correlations can be constructed based on the maximum entropy method with a Markovian assumption of temporal evolution (Marre et al., 2009) or based on a generalized linear model (Pillow et al., 2005, 2008). By comparing the amount of

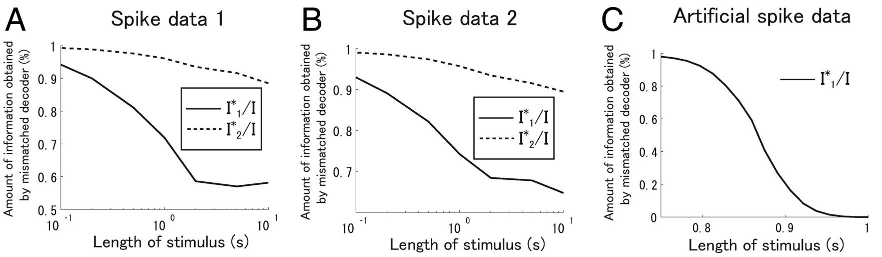

Figure 11. Dependence of the amount of information obtained by simplified decoders on the length of stimuli (Oizumi et al., 2009). All recorded cells ($N = 7$ spike data 1; $N = 6$ in spike data 2) were analyzed. ***A***, Spike data 1. ***B***, Spike data 2. ***C***, Artificial spike data generated according to the firing rates shown in Figure 12*A*.

information obtained by a probabilistic model that takes account of time-lagged correlations with that obtained by a probabilistic model that only takes account of simultaneous firing within a short time bin, we can quantitatively evaluate the amount of information conveyed by the complex temporal correlations between spikes.

Using a different approach than ours, Pillow et al. (2008) reported that model-based decoding that exploits time-lagged correlations between neurons extracted 20% more information about the visual scene than decoding under the assumption of independence. Decoding performance was quantified using the log signal-to-noise ratio. Our results showed that the second-order correlation model, which takes account of correlations within one time bin only, extracts only ~10% more about the visual scene than the independent model. The difference in improvement of decoding performance from the independent model between this work and the work of Pillow et al. may be attributable to the amount of information conveyed by time-lagged correlations. Besides, this could be also explained by the fact that they analyzed more cells (27 cells) than we did. Additional investigations of the importance of time-lagged correlations in information processing in the brain is required.

Quantitative investigation of the relationship between synchronized activity and animal behavior

We showed that synchronized activity does not convey much information about stimuli from a natural scene. In some experiments, however, a strong correlation between synchronized activity and animal behavior has been demonstrated (Stopfer et al., 1997; Ishikane et al., 2005). Stopfer et al. (1997) showed that picrotoxin-induced desynchronization impaired the discrimination of molecularly similar odorants in honeybees but did not prevent coarse discriminations of dissimilar odorants. Ishikane et al. (2005) showed that bicuculline-induced desynchronization suppressed escape behavior in frogs. The important point in these studies is that pharmacological blockade of GABA receptors strongly affected synchronization only, and had little effect on the firing rate of neurons. If the firing rate of neurons relevant to the behavior did not change at all, we could say without doubt that synchronized activity is essential to the decoding of neural activities. However, some ambiguity remains because it is impossible that pharmacological blockade does not alter the firing rate of any neuron at all. To resolve this ambiguity, the information for mismatched decoders, I^*, may be helpful.

Let us assume that we experimentally obtain normal neural responses to a specific stimulus s, $\mathbf{r}_1$, and altered neural responses to the same stimulus s after pharmacological blockade of neurotransmitter receptors, $\mathbf{r}_2$. If animal behavior between $\mathbf{r}_1$ and $\mathbf{r}_2$ differed, this would mean that the brain interpreted that two "different" stimuli were presented when $\mathbf{r}_1$ and $\mathbf{r}_2$ were evoked, even though the same stimulus, s, had in fact been presented. The important question is what difference in neural activities before and after the pharmacological blockade determined the judgment of the brain. This question can be quantitatively answered by computing the mutual information, I, between the two "different" stimuli interpreted by the brain and the corresponding neural responses and by comparing I with the information for mismatched decoders, I^*. For example, if I_1^*/I is high, it can be said that the decision of the brain is mainly based on the difference in firing rate between two neural responses $\mathbf{r}_1$ and $\mathbf{r}_2$. However, if I_1^*/I is low, the difference in correlated activities plays a crucial role in discriminating the stimulus. Applying the information theoretic measures, I and I^*, to behavioral experiments with physiological measurements will provide profound insights into how information is decoded in the brain.

Appendix: Theoretical evaluation of information I, I^*, and I^{NL}

In this appendix, we compared three measures of information contained in neural activities, namely mutual information I, information obtained by mismatched decoding I^*, and Nirenberg–Latham information I^{NL}, by analytical computation. Two results were obtained: (1) I and I^* provide consistent results with the minimum mean-square error, and (2) the difference between I^* and I^{NL} may increase when many cells are analyzed and I^{NL} can take negative values.

First, let us consider the problem in which mutual information is computed when stimulus s, which is a single continuous variable, and slightly different stimulus $s + \Delta s$ are presented. We assume the prior probability of stimuli $p(s)$ and $p(s + \Delta s)$ are equal: $p(s) = p(s + \Delta s) = 1/2$. Neural responses evoked by the stimuli are denoted by $\mathbf{r}$, which is considered here to be the neuron firing rate. When the difference between two stimuli is small, the conditional probability $p(\mathbf{r}|s + \Delta s)$ can be expanded with respect to Δs as follows:

$$p(\mathbf{r}|s + \Delta s) = p(\mathbf{r}|s) + p'(\mathbf{r}|s)\Delta s + \frac{1}{2}p''(\mathbf{r}|s)(\Delta s)^2 + \cdots, \tag{25}$$

where $'$ represents differentiation with respect to s. Using Equation 25, to leading order of Δs, we can write mutual information I as follows:

$$I = \frac{\Delta s^2}{8}\int d\mathbf{r}\frac{(p'(\mathbf{r}|s))^2}{p(\mathbf{r}|s)}, \tag{26}$$

where $\int d\mathbf{r}\frac{(p'(\mathbf{r}|s))^2}{p(\mathbf{r}|s)}$, is the Fisher information. The Fisher information has also been widely used in neuroscience as the maximal amount of information that can be extracted from neural responses (Paradiso, 1988; Seung and Sompolinsky, 1993; Abbott and Dayan, 1999; Gutnisky and Dragoi, 2008) because the inverse of the Fisher information gives the lower bound of the mean-

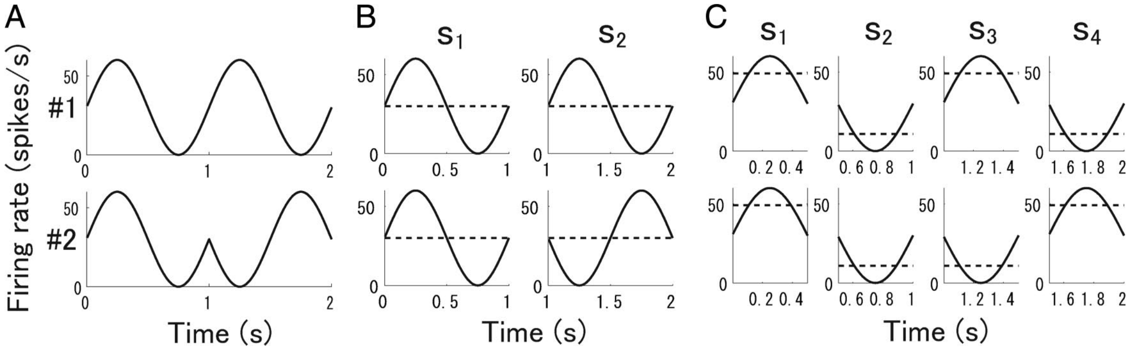

Figure 12. Firing rates of two model cells. Rate of cell 1 is shown in top panel; rate of cell 2 is shown in bottom panel (Oizumi et al., 2009). ***A***, Firing rates from 0 to 2 s. ***B***, Firing rates (solid line) and mean firing rates (dashed line) when stimulus duration was 1 s. ***C***, Firing rates (solid line) and mean firing rates (dashed line) when stimulus duration was 500 ms.

square error when the stimulus s is optimally estimated (i.e., the minimum mean-square error). As we can see in Equation 26, the mutual information is proportional to the Fisher information when Δs is small. Similarly, $\tilde{I}(\beta)$ (Eq. 6) can be written as follows:

$$\tilde{I}(\beta) = \frac{\Delta s^2}{8}\left(-\beta^2\int d\mathbf{r}p(\mathbf{r}|s)\left(\frac{q'(\mathbf{r}|s)}{q(\mathbf{r}|s)}\right)^2 + 2\beta\int d\mathbf{r}\frac{p'(\mathbf{r}|s)q'(\mathbf{r}|s)}{q(\mathbf{r}|s)}\right). \quad (27)$$

By maximizing $\tilde{I}(\beta)$ with respect to β, we obtain the correct information I^* for mismatched decoders as follows:

$$I^* = \frac{\Delta s^2}{8}\left(\int d\mathbf{r}\frac{p'(\mathbf{r}|s)q'(\mathbf{r}|s)}{q(\mathbf{r}|s)}\right)^2\left(\int d\mathbf{r}\frac{p(\mathbf{r}|s)(q'(\mathbf{r}|s))^2}{q(\mathbf{r}|s)^2}\right)^{-1}. \quad (28)$$

By substituting $\beta = 1$ into $\tilde{I}(\beta)$, we can obtain the Nirenberg–Latham information I^{NL} as follows:

$$I^{\mathrm{NL}} = \frac{\Delta s^2}{8}\left(-\int d\mathbf{r}p(\mathbf{r}|s)\left(\frac{q'(\mathbf{r}|s)}{q(\mathbf{r}|s)}\right)^2 + 2\int d\mathbf{r}\frac{p'(\mathbf{r}|s)q'(\mathbf{r}|s)}{q(\mathbf{r}|s)}\right). \quad (29)$$

We can also easily check that $\tilde{I}(\beta)$ becomes equal to the mutual information I when $q(\mathbf{r}|s) = p(\mathbf{r}|s)$ and $\beta = 1$. Taking into consideration the proportionality of the mutual information to the Fisher information, we can interpret $\left(\int d\mathbf{r}\frac{p'(\mathbf{r}|s)q'(\mathbf{r}|s)}{q(\mathbf{r}|s)}\right)^2$ $\left(\int d\mathbf{r}\frac{p(\mathbf{r}|s)(q'(\mathbf{r}|s))^2}{q(\mathbf{r}|s)^2}\right)^{-1}$ in Equation 28 as being a Fisher information-like quantity for mismatched decoders.

We assume that the encoding model $p(\mathbf{r}|s)$ obeys the Gaussian distribution as follows:

$$p(\mathbf{r}|s) = \frac{1}{Z}\exp\left(-\frac{1}{2}(\mathbf{r}-\mathbf{f}(s))^T\mathbf{C}^{-1}(\mathbf{r}-\mathbf{f}(s))\right), \quad (30)$$

where ^{T} stands for the transpose operation, $\mathbf{f}(s)$ is the mean firing rates given stimulus s, and $\mathbf{C}$ is the covariance matrix. We consider an independent decoding model $q(\mathbf{r}|s)$ that ignores correlations as follows:

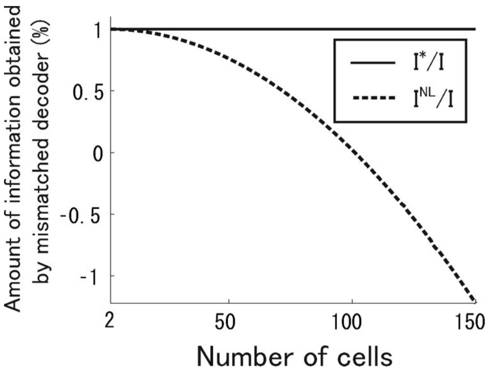

Figure 13. Difference between I^*/I (solid line) and I^{NL}/I (dotted line) in a Gaussian model in which correlations and derivatives of mean firing rates are uniform (Oizumi et al., 2009). Correlation parameter $c = 0.01$.

$$q(\mathbf{r}|s) = \frac{1}{Z_D}\exp\left(-\frac{1}{2}(\mathbf{r}-\mathbf{f}(s))^T\mathbf{C}_D^{-1}(\mathbf{r}-\mathbf{f}(s))\right), \quad (31)$$

where $\mathbf{C}_D$ is the diagonal covariance matrix obtained by setting the off-diagonal elements of $\mathbf{C}$ to 0. If the Gaussian integral is performed for Equations. 26, 28, and 29, I, I^*, and I^{NL} can be written as follows:

$$I = \frac{\Delta s^2}{8}\mathbf{f}'^T(s)\mathbf{C}^{-1}\mathbf{f}'(s), \quad (32)$$

$$I^* = \frac{\Delta s^2}{8}\frac{(\mathbf{f}'^T(s)\mathbf{C}_D^{-1}\mathbf{f}'(s))^2}{\mathbf{f}'^T(s)\mathbf{C}_D^{-1}\mathbf{C}\mathbf{C}_D^{-1}\mathbf{f}'(s)}, \quad (33)$$

$$I^{\mathrm{NL}} = \frac{\Delta s^2}{8}(-\mathbf{f}'^T(s)\mathbf{C}_D^{-1}\mathbf{C}\mathbf{C}_D^{-1}\mathbf{f}'(s) + 2\mathbf{f}'^T(s)\mathbf{C}_D^{-1}\mathbf{f}'(s)). \quad (34)$$

Next, let us consider the minimum mean-square error when stimulus s is presented. The optimal estimate of stimulus s when we know the actual encoding model $p(\mathbf{r}|s)$ is the value of $\hat{s}$ that maximizes the likelihood $p(\mathbf{r}|s)$. Similarly, the optimal estimate of stimulus s when we can only use the independent model $q(\mathbf{r}|s)$ is the value of $\hat{s}$ that maximizes the likelihood $q(\mathbf{r}|s)$. Previously, Wu et al. (2001) computed the minimum mean-square error when the optimal decoder is applied, MMSE, and the minimum mean-square error when the independent decoder is applied, MMSE* (Wu et al., 2001). These are given by the following:

$$\mathrm{MMSE}(s) = \frac{1}{\mathbf{f}'^{T}(s)\mathbf{C}^{-1}\mathbf{f}'(s)}, \tag{35}$$

$$\mathrm{MMSE}^{*}(s) = \frac{\mathbf{f}'^{T}(s)\mathbf{C}_{D}^{-1}\mathbf{C}\mathbf{C}_{D}^{-1}\mathbf{f}'(s)}{(\mathbf{f}'^{T}(s)\mathbf{C}_{D}^{-1}\mathbf{f}'(s))^{2}}. \tag{36}$$

If we compare Equation 32 with Equation 35, we can see that mutual information I is inversely proportional to the minimum mean-square error when the optimal decoder is applied. Similarly, as can be seen in Equations 33 and 36, I^* is also inversely proportional to the minimum mean-square error when the independent decoder is applied. Thus, I^* corresponds to the mutual information not only from the viewpoint of communication rate across a channel but also from that of the minimum mean-square error. However, I^{NL} is not inversely proportional to the minimum mean-squared error.

As a simple example that demonstrates a large discrepancy between I^* and I^{NL}, we considered a uniform correlation model (Abbott and Dayan, 1999; Wu et al., 2001) in which covariance matrix $\mathbf{C}$ is given by $C_{ij} = \sigma^2\,[\delta_{ij} + c(1 - \delta_{ij})]$ and assumed that the derivatives of the firing rates were uniform: that is, $f_i{=}f'$. In this case, I, I^*, and I^{NL} become the following:

$$I = \frac{\Delta s^2}{8}\,\frac{Nf'^{2}}{\sigma^2(Nc + 1 - c)}, \tag{37}$$

$$I^{*} = \frac{\Delta s^2}{8}\,\frac{Nf'^{2}}{\sigma^2(Nc + 1 - c)}, \tag{38}$$

$$I^{\mathrm{NL}} = \frac{\Delta s^2}{8}\,\frac{(-c(N-1)+1)Nf'^{2}}{\sigma^2}, \tag{39}$$

where N is the number of cells. We can see that I^* is equal to I, which means that information is not lost even if correlation is ignored in the decoding process. Figure 13 shows I^{NL}/I and I^*/I when the degree of correlation c is 0.01. As shown in Figure 13, the difference between the correct information I^* and Nirenberg–Latham information I^{NL} is markedly large when the number of cells N is large. When $N > \frac{c+1}{c}$, I^{NL} is negative. Analysis showed that the use of Nirenberg–Latham information I^{NL} as a lower bound of the correct information I^* can lead to erroneous conclusions, particularly when many cells are analyzed. In the spike data used in this study, we did not observe a large discrepancy between I^* and I^{NL}, possibly because the number of cells analyzed was small (Fig. 9).

References

Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. Neural Comput 11:91–101.

Amari S (2001) Information geometry on hierarchy of probability distributions. IEEE Trans Inform Theory 47:1701–1711.

Amari S, Nakahara H (2006) Correlation and independence in the neural code. Neural Comput 18:1259–1267.

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358–366.

Berger A, Della Pietra S, Della Pietra C (1996) A maximum entropy approach to natural language processing. Comput Linguistics 22:1–36.

Brody CD, Hopfield JJ (2003) Simple networks for spike-timing-based computation, with application to olfactory processing. Neuron 37:843–852.

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

Dan Y, Alonso JM, Usrey WM, Reid RC (1998) Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. Nat Neurosci 1:501–507.

Golledge HD, Panzeri S, Zheng F, Pola G, Scannell JW, Giannikopoulos DV, Mason RJ, Tovée MJ, Young MP (2003) Correlations, feature-binding and population coding in primary visual cortex. Neuroreport 14:1045–1050.

Gollisch T, Meister M (2008) Rapid neural coding in the retina with relative spike latencies. Science 319:1108–1111.

Gray CM, König P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature 338:334–337.

Gutnisky DA, Dragoi V (2008) Adaptive coding of visual information in neural populations. Nature 452:220–224.

Hopfield JJ (1999) Odor space and olfactory processing: collective algorithms and neural implementation. Proc Natl Acad Sci U S A 96:12506–12511.

Ishikane H, Gangi M, Honda S, Tachibana M (2005) Synchronized retinal oscillations encode essential information for escape behavior in frogs. Nat Neurosci, 80:1087–1095.

Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106:62–79.

Latham PE, Nirenberg S (2005) Synergy, redundancy, and independence in population codes, revisited. J Neurosci 25:5195–5206.

Marre O, El Boustani S, Frégnac Y, Destexhe A (2009) Prediction of spatiotemporal patterns of neural activity from pairwise correlations. Phys Rev Lett 102:138101.

Meister M (1996) Multineuronal codes in retinal signaling. Proc Natl Acad Sci U S A 93:609–614.

Meister M, Pine J, Baylor DA (1994) Multi-neuronal signals from the retina: acquisition and analysis. J Neurosci Methods 51:95–106.

Meister M, Lagnado L, Baylor DA (1995) Concerted signaling by retinal ganglion cells. Science 2700:1207–1210.

Merhav N, Kaplan G, Lapidoth A, Shamai Shitz S (1994) On information rates for mismatched decoders. IEEE Trans Inform Theory 40:1953–1967.

Montani F, Kohn A, Smith MA, Schultz SR (2007) The role of correlations in direction and contrast coding in the primary visual cortex. J Neurosci 27:2338–2348.

Montani F, Ince RA, Senatore R, Arabzadeh E, Diamond ME, Panzeri S (2009) The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. Philos Transact A Math Phys Eng Sci 367:3297–3310.

Montemurro MA, Senatore R, Panzeri S (2007) Tight data-robust bounds to mutual information combining shuffling and model selection techniques. Neural Comput 19:2913–2957.

Nakahara H, Amari S (2002) Information-geometric measure for neural spikes. Neural Comput 14:2269–2316.

Nirenberg S, Latham PE (2003) Decoding neural spike trains: how important are correlations? Proc Natl Acad Sci U S A 100:7348–7353.

Nirenberg S, Carcieri SM, Jacobs AL, Latham PE (2001) Retinal ganglion cells act largely as independent encoders. Nature 411:698–701.

Oizumi M, Ishii T, Ishibashi K, Hosoya T, Okada M (2009) A general framework for investigating how far the decoding process in the brain can be simplified. Adv Neural Inform Process Syst 21:1225–1232.

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

Panzeri S, Petersen RS, Schultz SR, Lebedev M, Diamond ME (2001) The role of spike timing in the coding of stimulus location in rat somatosensory cortex. Neuron 29:769–777.

Paradiso MA (1988) A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. Biol Cybern 58:35–49.

Perez-Orive J, Mazor O, Turner GC, Cassenaer S, Wilson RI, Laurent G (2002) Oscillations and sparsening of odor representations in the mushroom body. Science 297:359–365.

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. J Neurosci 250:11003–11013.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 4540:995–999.

Pola G, Thiele A, Hoffmann KP, Panzeri S (2003) An exact method to quantify the information transmitted by different mechanisms of correlational coding. Network 14:35–60.

Reich DS, Mechler F, Victor JD (2001) Temporal coding of contrast in primary visual cortex: when, what, and why. J Neurophysiol 85:1039–1050.

Schneidman E, Still S, Berry MJ 2nd, Bialek W (2003) Network information and connected correlations. Phys Rev Lett 91:238701.

Schneidman E, Berry MJ 2nd, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440:1007–1012.

Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. Proc Natl Acad Sci U S A 90:10749–10753.

Shannon CE (1948) A mathematical theory of communication. Bell System Tech J 27:379–423, 623–656.

Shlens J, Field GD, Gauthier JL, Grivich MI, Petrusca D, Sher A, Litke AM, Chichilnisky EJ (2006) The structure of multi-neuron firing patterns in primate retina. J Neurosci 260:8254–8266.

Stopfer M, Bhagavan S, Smith BH, Laurent G (1997) Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. Nature 390:70–74.

Tang A, Jackson D, Hobbs J, Chen W, Smith JL, Patel H, Prieto A, Petrusca D, Grivich MI, Sher A, Hottowy P, Dabrowski W, Litke AM, Beggs JM (2008) A maximum entropy model applied to spatial and temporal correlations from cortical networks *in vitro*. J Neurosci 28:505–518.

van Hateren JH (1997) Processing of natural time series of intensities by the visual system of the blowfly. Vision Res 37:3407–3416.

Wu S, Nakahara H, Amari S (2001) Population coding with correlation and an unfaithful model. Neural Comput 13:775–797.