Multineuronal vectorization is more efficient than time-segmental vectorization for information extraction from neuronal activities in the inferior temporal cortex

Hidekazu Kaneko a,*, Hiroshi Tamura b, c, Shunta Tate b, Takahiro Kawashima d, Shinya S. Suzuki a, Ichiro Fujita a, b, c

a Institute for Human Science and Biomedical Engineering, National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central 6, Higashi, Tsukuba, Ibaraki 305-8566, Japan
b Laboratory for Cognitive Neuroscience, Graduate School of Frontier Biosciences, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
c CREST, Japan Science and Technology Agency, 4-1-8, Honcho, Kawaguchi, Saitama 332-0012, Japan
d Cooperative Research Facility Center, Toyohashi University of Technology, 1-1, Hiberigoko, Tempaku-cho, Toyohashi, Aichi 441-8580, Japan

ARTICLE INFO

Article history:
Received 31 October 2009
Accepted 8 March 2010

Keywords:
Information theory
Temporal coding
Population coding
Decoding
Single-unit recording
Vision
Inferior temporal cortex
Brain-machine interface

ABSTRACT

In order for patients with disabilities to control assistive devices with their own neural activity, multineuronal spike trains must be efficiently decoded because only limited computational resources can be used to generate prosthetic control signals in portable real-time applications. In this study, we compare the abilities of two vectorizing procedures (multineuronal and time-segmental) to extract information from spike trains during the same total neuron-seconds. In the multineuronal vectorizing procedure, we defined a response vector whose components represented the spike counts of one to five neurons. In the time-segmental vectorizing procedure, a response vector consisted of components representing a neuron’s spike counts for one to five time-segment(s) of a response period of 1 s. Spike trains were recorded from neurons in the inferior temporal cortex of monkeys presented with visual stimuli. We examined whether the amount of information of the visual stimuli carried by these neurons differed between the two vectorizing procedures. The amount of information calculated with the multineuronal vectorizing procedure, but not the time-segmental vectorizing procedure, significantly increased with the dimensions of the response vector. We conclude that the multineuronal vectorizing procedure is superior to the time-segmental vectorizing procedure in efficiently extracting information from neuronal signals.

1. Introduction

Neural interface/prosthesis technology may enable amputees to voluntarily control their artificial limbs with neuronal activities and allow individuals who have sustained spinal cord injuries to remotely control assistive devices. It was recently demonstrated that electronic devices can be controlled using neuronal activities recorded from animals (Chapin, Moxon, Markowitz, & Nicolelis, 1999; Paninski, Shoham, Fellows, Hatsopoulos, & Donoghue, 2004; Schwartz, Taylor, & Tillery, 2001; Serruya, Hatsopoulos, Fellows, Paninski, & Donoghue, 2003; Wassberg et al., 2000) and tetraplegic patients (Donoghue, Nurmikko, Black, & Hochberg, 2007; Hochberg et al., 2006). These promising results obtained under laboratory conditions raise the possibility that neural interface/prosthesis technology could provide valuable therapeutic measures.

Control signals for prosthetic devices are usually derived from the neuronal activities of motor-related cortices, but signals from other cortices may also be used. For example, Felton, Wilson, Williams, and Garell (2007) recently demonstrated that a prosthetic device can be controlled using neural activity induced in the human temporal lobe by thinking about tones. Indeed, the results suggest that neuronal activities in any cortex can be used to create control signals for prosthetic devices. An important caveat, however, is the reproducibility of the neuronal activities that represent a specific motor, perceptual, or cognitive event.

This reproducibility parameter of neuronal activities can be evaluated using information theory. The amount of information shared by neuronal activities and the corresponding event is often used as a measure of the reproducibility of neuronal activities for a particular event because shared information depends on the probability that the neuronal activities and the event coincide (Cover & Thomas, 1991). When various signal processing methods...
are compared, the methods that extract more information from the neuronal activities are generally considered superior.

The application of information theory to neuronal signals has shown that both the vectorization of spike counts across multiple neurons (multineuronal vectorization) and the vectorization of spike counts over a time period of a single neuron (time-segmental vectorization) can increase the amount of information extracted from spike trains (Aguilopoulous et al., 2005; Gochin et al., 1994; Kaneko, Tamura, Kawashima, Suzuki, & Fujita, 2007; Optican & Richmond, 1987; Reich, Meckler, & Victor, 2001a; Richmond & Optican, 1987; Rolls, Franco, Aguilopoulous, & Recce, 2003; Tovee, Rolls, Treves, & Bellis, 1993). Studies using simultaneously recorded responses from multiple neurons have indicated that a large amount of information can be obtained using the multineuronal vectorizing procedure (Aguilopoulous et al., 2005; Gochin et al., 1994; Kaneko, Tamura, Kawashima et al., 2007; Reich et al., 2001a; Rolls et al., 2003). Time-segmental vectorization was also shown to be effective for extracting more information (Optican & Richmond, 1987; Richmond & Optican, 1987; Tovee et al., 1993). The adoption of both vectorizing procedures, however, prohibitively increases the dimensions of the response vector and may exceed the limited capacity of the data processing resources. This issue is critical for information theoretical approaches to real-time signal processing and the control of portable prosthetic devices.

In the present study, we compared these two vectorizing procedures based on their ability to efficiently extract information from extracellular spike activities from a population of neurons. For this comparison, we obtained multineuronal data from the inferior temporal (IT) cortex of monkeys with a multisite microelectrode while they were presented with a set of visual stimuli. The IT is an association area that is crucial for visual object recognition (D’Esposito et al., 1997; Miyashita, 1988; Roland & Gulyas, 1994, 1995). In our multineuronal vectorizing procedure, we defined a response vector whose components represented the spike counts of multiple neurons. In the time-segmental vectorizing procedure, a response vector consisted of components representing a neuron’s spike counts for one to five time-segment(s) of the whole response period (1 s). To fairly compare these vectorizing procedures, the total neuron-seconds of analyzed data was fixed at 1 neuron-second. We then determined which of these two strategies enabled us to obtain more information with the same number of vector dimensions.

2. Methods

2.1. Recording multiple neuronal activities

Neuronal responses to 64 visual stimuli were recorded from the IT of four anesthetized monkeys (Macaca fuscata; body weight: 5.2–7.5 kg; see Tamura, Kaneko, Kawasaki, and Fujita (2004) for details). All experimental procedures were based on guidelines from the National Institutes of Health of the United States (1996). Furthermore, the Osaka University animal experiment committee approved the procedures.

The general experimental procedures were similar to those described previously (Tamura et al., 2004). The monkeys were prepared for repeated recording by undergoing initial aseptic surgery under sodium pentobarbital anesthesia. For recording experiments, the monkeys were anesthetized with isoflurane (i.e., heart rate, body temperature, end-tidal CO2, and arterial oxygen saturation level) were monitored throughout the experiments. The monkeys’ eyes were covered with contact lenses. For multichannel recording of neuronal activity, a seven-core electrode (seven recording sites, impedance of each site: 1–2 MΩ at 1 kHz; Heptode, UWE Thomas Recording, Germany) was inserted into the IT through a craniotomy. To ensure stable recordings, we immobilized the brain surface with paraffin. After placing an electrode at each recording position we waited at least 20 min before the data acquisition was started. To prevent eye movement, the monkeys were paralyzed with pancuronium bromide.

Because the sampling radius of the seven-core electrode is approximately 150 μm, multichannel recordings were made at 300 μm intervals to avoid sampling the same neuron twice. The neuronal activity from each recording site was amplified 10,000 times, band-pass filtered (500 Hz to 3 kHz), and digitized at 20 kHz for offline spike sorting and analysis.

Multiple single-unit recording of nearby neurons was achieved by applying spike sorting to the multichannel recording data (Kaneko, Suzuki, Okada, & Akamatsu, 1999; Kaneko, Tamura, & Suzuki, 2007; McNaughton, O’Keefe, & Barnes, 1983). We employed a custom-made spike sorter consisting of three procedures: spike detection, burst detection, and spike classification (Kaneko et al., 1999; Kaneko, Tamura, & Suzuki, 2007). Multineuronal spikes were detected by matching the observed waveforms with a set of spike templates with different durations (spike detection). For each neuron spike, six waveforms were recorded simultaneously at six recording sites (i.e., the tip and five lateral sites of the seven-core electrode; one recording site was not used for technical reasons). The amplitudes of these waveforms constituted a spike-amplitude vector. A burst of spikes was identified based on short inter-spike intervals (1.5–15 ms) with an allowance for small amplitude ratio variations (burst detection). The amplitude vector of the first intraburst spike was used for spike clustering. Finally, clusters of spike-amplitude vectors were classified by bottom-up hierarchical clustering (spike classification).

2.2. Visual stimulation

We used 64 visual stimuli (see Fig. 1B in Tamura et al. (2004)), including 53 two-dimensional geometric shapes (e.g., circles, squares, triangles, bars, stars, gradation patterns, and gratings) and 11 photographs of natural objects (e.g., banana, apple, human face, monkey face, and hand). During each of the 10 presentation sessions, each of the 64 visual stimuli was presented for 1 s in random order at the center of the receptive field against a homogeneous gray background (15.7 cd/m2). The interstimulus interval was 1 s. The onset and offset of the stimulus presentation were timed by the V-SYNC signals of the display. The entire recording period lasted 1280 s plus the sum of the V-SYNC signal delays (5.6 s). Because the response latency in the IT is approximately 80 ms, on average across neurons (Richmond, Wurtz, & Sato, 1983), we defined the response period as a 1 s period starting 80 ms after the onset of each stimulus presentation.

2.3. Response vector

For both the multineuronal and time-segmental vectorizing procedures, we expressed the responses of multiple neurons and their temporal patterns mathematically as vectors (Gochin et al., 1994; Optican & Richmond, 1987; Richmond & Optican, 1987). When we vectorize the responses of multiple neurons, the number of spikes (i.e., spike count) generated from each neuron during the response period is assigned to a component in the response vector. Let n be the number of neurons used for the multineuronal vectorization and the spike count of the i-th neuron be equal to nj. We represent the response vector of the multiple neurons as

\[ \mathbf{r} = (r_1, r_2, \ldots, r_n) \]  

(1)

To make the total neuron-seconds of used data equal to 1 neuron-second, the period for evaluating nj is fixed at \((1/n)\) s.
To uniformly sample data in the response period, a response vector is constructed at every \((1/n)\) s time-segment. This treatment minimizes the effects of particular time-segments on the subsequent multinatural information analysis, e.g., "ON" responses of neurons at the beginning of the response period may contain slightly more information (Maldonado et al., 2008; Osborne, Bialek, & Lisberger, 2004; Reich, Mechler, & Victor, 2001b; Tovee et al., 1993). In a posteriori analysis of all analyzed data sets, the amount of information was more or less uniformly distributed in the entire response period (Fig. 1). Thus, the subsequent multinatural analyses were not affected by particular time-segments of the response period or by the "ON" responses.

When the temporal response of a neuron is vectorized, the spike count in each of the multiple time-segments is assigned to a component of the response vector. Let the response period \((1 s)\) be separated into \(n\) segments \((n = 1, 2, \ldots, 5)\) and the spike count in the \(i\)-th segment be equal to \(r_i\), i.e., the neuron-spike counts of the data used is 1 neuron-second. The response vector is also represented as Eq. (1).

The response vectors obtained by the above two vectorizing procedures can be treated in the same way for calculating the amount of information, because the vectorial form is identical in the two procedures.

### 2.4. Amount of transmitted information

Based on Shannon’s information theory (Cover & Thomas, 1991), we calculated the amount of information on the visual stimulus that was transmitted by neuronal responses. Let an \(n\)-component response vector \(\mathbf{r}\) be the one obtained after one visual stimulus \(s\) out of 64 visual stimuli \(S\) was presented. The response vectors obtained after the presentation of the same visual stimulus scatter in an \(n\)-dimensional vector space \(\mathbf{R}\). The amount of information on \(S\) that was transmitted by neuronal responses in \(\mathbf{R}\), \(I(S; \mathbf{R})\), is derived from the joint probability of a visual stimulus \(s\) and a response vector \(\mathbf{r}, p(s, \mathbf{r})\), as follows:

\[
I(S; \mathbf{R}) = \sum_{s=1}^{64} \int_{\mathbf{R}} p(s, \mathbf{r}) \log_2 \frac{p(s, \mathbf{r})}{p(s)p(\mathbf{r})} d\mathbf{r},
\]

where \(p(s)\) and \(p(\mathbf{r})\) indicate the marginal probability density functions of \(p(s, \mathbf{r})\).

Although a Poisson process is often used to model the spike firing of a neuron, the distribution of spike counts of a neuron does not always fit a Poisson distribution (Amarasingham, Chen, Geman, Harrison, & Steinberg, 2006; Gershon, Wiener, Latham, & Richmond, 1998; Shimomoto, Miyazaki, Tamura, & Fujita, 2005). As a suboptimal solution, we used a non-parametric method to estimate the stimulus–response joint probability distribution function \(p(s, \mathbf{r})\) (Fukunaga, 1990; Optican & Richmond, 1987; Silverman, 1986). In this method, the spike count distribution of a neuron in response to a stimulus is given by blurring the spike counts with a Gaussian function. Thus, \(p(s, \mathbf{r})\) is expressed as a function of the \(n\)-component response vector \(\mathbf{r}_k\) to the \(k\)-th presentation of the \(s\)-th stimulus as follows:

\[
p(s, \mathbf{r}) = \frac{1}{64K} \sum_{k=1}^{K} \frac{1}{\sqrt{(2\pi)^n|\Sigma|^n}} e^{-\frac{1}{2} (\mathbf{r} - \mathbf{r}_k)^T \Sigma^{-1} (\mathbf{r} - \mathbf{r}_k)},
\]

where

\[
\Sigma = \left( \frac{4}{n+2} K - \frac{n}{K} \right)^2 K - 1 \Sigma_k,
\]

\(\Sigma_k\) indicates the population variance–covariance matrix estimated from the responses to the \(s\)-th stimulus presented \(K\) times, \(K\) indicates the transpose of a vector \(\mathbf{a}\), and \(\mathbf{B}^{-1}\) indicates the inverse of a matrix \(\mathbf{B}\). Since each visual stimulus was presented once per session for \(10\) sessions, \(K = 10\). To compensate for estimating the joint probability distribution function from the limited number of observations or trials, we used Eq. (4) (Fukunaga, 1990; Silverman, 1986). Eq. (4) is optimized for nonparametric density estimation of normal distributions with normal kernel function by minimizing the integral mean-square error between the estimated and true normal probability density functions (Silverman, 1986). The precision of the estimation improves as the number of trials for the same stimulus increases.

In a posteriori analysis of all analyzed neurons, the mean spike count of each neuron per stimulus was not significantly different across the sessions \((p = 0.959, Kruskal–Wallis test)\). Thus, this satisfied the assumption that the neuronal responses were in a stationary state, which is necessary to estimate \(p(s, \mathbf{r})\) accurately. Furthermore, the distribution of spike counts for every combination of neurons and stimuli was not significantly different from the distribution of \(p(s, \mathbf{r})\) (multiple Kolmogorov–Smirnov tests for goodness-of-fit with Bonferroni corrections, \(p > 0.05\)).

### 2.5. Bias reduction

The limited number of trials/observations causes biases in the calculation of the amount of information transmitted by neuronal activities. Ideally, the number of trials for each stimulus should be sufficiently large to alleviate this problem. Practically, it is difficult, if not impossible, to obtain data for a sufficiently large number of trial repetitions; this is because the large number of stimuli needed in our study reduces the number of trials per stimulus, and accidental electrode movements or changes in the condition of the animal can disrupt stable long-term neuronal recordings. To address this issue, multiple bias reduction measures have been used to calculate the net amount of information transmitted by the neuronal activities (Panzeri, Senatore, Montemurro, & Petersen, 2007). In this study, we used a battery of bias reduction measures (Methods 1, 2, and 3, see below).

Method 1 is an unbiased estimation of the stimulus–response joint probability distribution function with a Gaussian mixture model (Fukunaga, 1990; Silverman, 1986, see previous section). The blurring function (i.e., Eq. (4)) in the Gaussian mixture model decreased the bias by reducing the resolution of the response vector space; this is similar to reducing the number of response patterns in the calculation of entropies.
Method 2 is used to remove the bias caused by pseudo relationships between the stimuli and responses (Tovee et al., 1993; Treves & Panzeri, 1995). For example, in a small number of trials, some relationships between stimuli and responses may be accidental and thus misleading us into thinking more information was transmitted by the neuronal activities. This bias is reduced by evaluating the net amount of information as follows:

\[ I_{\text{net}}(S; R) = I_{\text{rep}}(S; R) - I_{\text{shuf}}(S; R), \]  

(5)

where \( I_{\text{net}}(S; R) \) represents the net amount of information, \( I_{\text{rep}}(S; R) \) is \( I(S; R) \) obtained from spike trains during the response period, and \( I_{\text{shuf}}(S; R) \) is \( I(S; R) \) obtained from spike trains shifted between different stimuli.

\( I_{\text{rep}}(S; R) \) and \( I_{\text{shuf}}(S; R) \) were calculated by constructing response vectors using multineuronal or time-segmental procedures. Since the relationships between the stimuli and responses are broken by shifting spike trains between different stimuli, \( I_{\text{shuf}}(S; R) \) different from zero should be eliminated as a bias caused by the limited number of observations.

Method 3 is used to reduce the bias caused by estimation errors of stimulus-independent correlations between the components of the response vectors (Montemurro, Senatore, & Panzeri, 2007; Panzeri et al., 2007). Since estimation errors of stimulus-independent correlations between the components of the response vectors might cause a bias in calculating \( I_{\text{rep}}(S; R) \) and \( I_{\text{shuf}}(S; R) \) in Eq. (5), this bias is reduced by shuffling the responses of a neuron to the same stimulus across trials (for multineuronal) or by shuffling the responses of a neuron to the same stimulus in the same segment across trials (for time-segmental). This shuffling procedure was applied to both \( I_{\text{rep}}(S; R) \) and \( I_{\text{shuf}}(S; R) \):

\[ I_{\text{rep}}(S; R) = I_{\text{rep}}(S; R_{\text{raw}}) - I_{\text{rep}}(S; R_{\text{shuf}}), \]  

(6)

\[ I_{\text{shuf}}(S; R) = I_{\text{shuf}}(S; R_{\text{raw}}) - I_{\text{shuf}}(S; R_{\text{shuf}}), \]  

(7)

where \( R_{\text{raw}} \) is the response vector space \( R \) when the probability of a response vector, \( p(r) \), is calculated from raw responses, \( R_{\text{shuf}} \) is \( R \) when \( p(r) \) is calculated from the shuffled responses of a neuron across trials or the shuffled responses of a neuron in the same segment across trials, and \( R_{\text{independent}} \) is \( R \) when \( p(r) = p(r_1, r_2, \ldots, r_n) \) is set to be the product of marginal probabilities \( p(r_1), p(r_2), \ldots, p(r_n) \) (i.e., the components of \( r \) are fully independent of each other).

In this compensation, the difference between \( I_{\text{rep}}(S; R_{\text{shuf}}) \) and \( I_{\text{rep}}(S; R_{\text{independent}}) \), both of which reach the same convergence value in an infinite number of trials/observations, is regarded as a bias included in \( I_{\text{rep}}(S; R_{\text{shuf}}) \). The same is true for \( I_{\text{shuf}}(S; R_{\text{shuf}}) \).

This method is equivalent to the method proposed by Panzeri et al. (2007) (see Appendix).

2.6. Estimated upper and lower bounds

Even after bias reduction by the methods described above, some biases still remain in the amount of information, especially at higher dimensions of \( R \). These residual biases can only be avoided by collecting a sufficient amount of data. For a suboptimal method to evaluate the residual bias, we defined the upper and lower bounds of the true value of the amount of transmitted information, \( I_{\text{true}}(S; R) \) (Borst & Theunissen, 1999; Panzeri et al., 2007). Since \( I_{\text{true}}(S; R) \) includes a positive residual bias, we can consider \( I_{\text{true}}(S; R) \) as the upper bound of \( I_{\text{true}}(S; R) \). To estimate the lower bound of \( I_{\text{true}}(S; R) \), a decoding process of response vectors was added to the information analysis (Aagelopoulou et al., 2005; Borst & Theunissen, 1999). In this process, every response is decoded into one of the presented visual stimuli by the maximum likelihood method under the leave-one-out treatment (Aagelopoulou et al., 2005; Rolls, Treves, & Tovee, 1997). When we represent the decoding function as \( D \), the lower bound of \( I_{\text{true}}(S; R) \) can be expressed as \( I_{\text{true}}(S; D(R)) \). The bias reduction measures mentioned above as Methods 1, 2, and 3 were also applied to the calculation of \( I_{\text{true}}(S; D(R)) \):

\[ I_{\text{true}}(S; D(R)) = I_{\text{rep}}(S; D(R)) - I_{\text{shuf}}(S; D(R)), \]  

(8)

\[ I_{\text{rep}}(S; D(R)) = I_{\text{rep}}(S; D(R_{\text{true}})) - (I_{\text{rep}}(S; D(R_{\text{shuf}})) - I_{\text{shuf}}(S; D(R_{\text{independent}}))), \]  

(9)

\[ I_{\text{true}}(S; D(R)) = I_{\text{rep}}(S; D(R_{\text{true}})) - (I_{\text{rep}}(S; D(R_{\text{shuf}})) - I_{\text{shuf}}(S; D(R_{\text{independent}}))), \]  

(10)

2.7. Redundancy analysis

As per a procedure published by Schneidman, Bialek, and Berry (2003), we analyzed the redundancy of activities between neurons for the multineuronal vectorizing procedure and between segments for the time-segmental vectorizing procedure. Since the response vectors are explained in the same way for the two vectorizing procedures (Eq. (1)), the redundancies can also be defined in the same way. Assuming that a response vector space \( R \) includes two-component response vectors (i.e., \( n = 2 \), and each component is obtained from half of the response period) and that another response vector space \( R_l \) includes one-component response vectors each of which has the \( i \)-th component of a response vector in \( R \) as the component (i.e., \( i = 1 \) or \( i = 2 \), the redundancy is estimated as follows:

Reuspondancy(R) = Itrue(S; R1) + Itrue(S; R2) - Iretrue(S; R)  

(11)

for the upper bound, and

\[ \text{Redundancy}(D(R)) = I_{\text{true}}(S; D(R_1)) + I_{\text{true}}(S; D(R_2)) - I_{\text{true}}(S; D(R)), \]  

(12)

for the lower bound. If the value is positive, then the two components of a response vector in \( R \) are redundant with each other. If the value is zero, then they are independent. If the value is negative, then they are synergistic, and more information is obtained in this case than when the value is positive or zero. Previous studies indicated that the redundancy, independency, and synergy were related to the signal and noise correlation between neurons (Averbeck, Latham, & Pouget, 2006; Montani, Kohn, Smith, & Schultz, 2007; Narayanan, Kimchi, & Laubach, 2005; Panzeri, Schultz, Treves, & Rolls, 1999; Petersen, Panzeri, & Diamond, 2001; Schneidman et al., 2003).

3. Results

A total of seven data sets, each containing spike trains of five to 10 neurons (47 neurons in total), were used in this study. In the multineuronal vectorizing procedure, \( n \)-neuron groups (up to five neurons per group) randomly selected from each data set were used for constructing the response vectors. More specifically, if the number of such combinations of \( n \) neurons was less than 100 patterns, all combinations of \( n \)-neuron groups were analyzed. The results were averaged across all analyzed neuron groups and time-segments. In the time-segmental vectorizing procedure, each of 47 neurons was analyzed by dividing the response period of 1 s into one to five segments, and the calculated values were averaged across neurons. We then compared the dimensionality effects (\( n = 1, 2, \ldots, 5 \)) of the two vectorizing procedures for the same total neuron-seconds.
3.1. Use of bias reduction measures

We first used simulation data to evaluate the efficiency of the bias reduction measures on the amount of multineuronal information obtained from the limited number of trials. Fig. 2 shows the amount of information for a data set in which every neuron carries 1 bit of information with complete redundancy between neurons. The data set was generated by assigning random values from a normal distribution (mean: 5 spikes, variance: 11.1) to the neuronal responses for half of the visual stimuli and assigning random values from another normal distribution (mean: 25 spikes, variance: 11.1) to the neuronal responses for the other half. The other conditions, i.e., 10 trials per stimulus and 64 visual stimuli, were the same as in the recording experiments. Since the activities of any two neurons were completely redundant, the amount of information should not increase with the number of neurons used. However, the estimated upper bounds (circles in Fig. 2) of the amount of information increased with the dimensions of the response vector as a result of biases caused by the limited number of trials, whereas the biases were suppressed in the estimated lower bounds (triangles in Fig. 2). It must be noted that, with a combination of all bias reduction measures (Fig. 2(d)), the deviations of the estimated upper and lower bounds, $I_{\text{net}}(S; R)$ and $I_{\text{net}}(S; D(R))$, from the true amount of transmitted information were minimized, and furthermore, the gap between the two bounds was also minimized. Thus, we considered $I_{\text{net}}(S; R)$ and $I_{\text{net}}(S; D(R))$ as the upper and lower bounds of $I_{\text{true}}(S; R)$, respectively.

Because the response vectors for the multineuronal and time-segmental procedures have mathematically the same structure as explained in Section 2.3, the demonstration presented above is applicable to both procedures.

3.2. Use of multineuronal spike counts

Stimulus preference correlations are higher between nearby neurons than between distant neurons (Gochin, Miller, Gross, & Gerstein, 1991; Tamura, Kaneko, & Fujita, 2005; Wang, Fujita, & Murayama, 2000), suggesting a clustered structure of neurons with similar selectivity in the IT (Fujita, 2002; Fujita, Tanaka, Ito, & Cheng, 1992; Kreiman et al., 2006; Tsunoda, Yamane, Nishizaki, & Tanifuji, 2001). Although the similarity in stimulus
the lower bound \( I_{net}(S; D(R)) \) was only 0.265 bits (SD: 0.218). Neuronal responses to the visual stimuli were redundant in the IT neurons (Fig. 4(a) and (b)) because the mean value of Redundancy(R) and Redundancy(D(R)) was slightly positive (mean: 0.020 bits, SD: 0.054 for Redundancy(R); mean: 0.016 bits, SD: 0.121 for Redundancy(D(R))). The ratio of the redundancy to the mean amount of information transmitted by single neurons was approximately 7%.

Although neuron responses were somewhat redundant, more information could be extracted with multineuronal activities. Fig. 5(a) shows the relationship between the amount of extracted information and the number of examined neurons. A regression analysis enabled us to estimate the response of a neuron from that of the other, i.e., their responses are redundant. This means that the similarity in stimulus preferences and the redundancy of neuronal activities are not directly linked. We evaluated the mean amount of information transmitted by single neurons and the redundancy between neurons. The upper bound \( I_{net}(S; R) \) averaged over 47 IT neurons was only 0.285 bits (SD: 0.217), while

Fig. 3. The stimulus preferences of simultaneously recorded neurons (#16, #49, and #144). These three adjacent neurons had different stimulus preferences. Visual stimuli producing salient responses are indicated with arrows. The abscissa indicates 64 visual stimuli. The ordinate indicates the mean 1 s spike counts over 10 trials for each stimulus.

Preference varies with the ratio of the sampled excitatory and inhibitory neurons (Tamura et al., 2005), the similarity in stimulus preferences between nearby neurons is lower than originally thought (Sato, Uchida, & Tanifuji, 2009; Tamura et al., 2005, 2004). Even nearby neurons have different responses to the same visual stimulus. Fig. 3 shows the stimulus preferences of three nearby neurons that were recorded simultaneously. All three neurons exhibited some responses to a white square, a white pinched square, and a hand. However, they had the strongest responses to different visual stimuli. Thus, more information can be extracted by combining the responses of multiple neurons than by using the responses of a single neuron. For the data in Fig. 3, the upper bound \( I_{net}(S; R) \) increased with the number of neurons at a rate of 0.229 bit/neuron and reached 1.020 bits with the three neurons. The lower bound \( I_{net}(S; D(R)) \) also increased at a rate of 0.127 bit/neuron and reached 0.792 bits.

A low similarity in stimulus preferences between nearby neurons does not always result in low redundancy of neuronal activities (Averbeck et al., 2006; Gawne & Richmond, 1993; Schneidman et al., 2003). For example, if the correlation of stimulus preferences of two neurons is nearly equal to minus one, the negative correlation enables us to estimate the response of a neuron from that of the other; i.e., their responses are redundant. This means that the similarity in stimulus preferences and the redundancy of neuronal activities are not directly linked. We evaluated the mean amount of information transmitted by single neurons and the redundancy between neurons. The upper bound \( I_{net}(S; R) \) averaged over 47 IT neurons was only 0.285 bits (SD: 0.217), while...
Tovee et al. analyzed for the same total amount of neuron seconds. In this paper, Tamura, Kawashima et al. hypothesized that the multineuronal vectorizing procedure is more effective than the time-segmental vectorizing procedure in which multineuronal data were recorded simultaneously and used to compare the two vectorizing procedures under conditions in which multineuronal data were recorded simultaneously and analyzed for the same total amount of neuron-seconds. In this section, we first discuss the validation of the evaluation and then hypothesize that the multineuronal vectorizing procedure is more efficient than the time-segmental vectorizing procedure to extract information from neuronal activities. Finally, we compare the efficiencies of these procedures from the perspective of real-time signal processing and the control of portable prosthetic devices.

### 4. Discussion

Previous studies demonstrated that spike counts from multiple neurons (multineuronal vectorizing procedure) and the temporal response patterns of each neuron (time-segmental vectorizing procedure) can be used to efficiently extract information from spike trains (Aggelopoulos et al., 2005; Gochin et al., 1994; Kaneko, Tamura, Kawashima et al., 2007; Optican & Richmond, 1987; Reich et al., 2001a; Richmond & Optican, 1987; Rolls et al., 2003; Tovee et al., 1993). To our knowledge, the present study is the first to compare the two vectorizing procedures under conditions in which multineuronal data were recorded simultaneously and analyzed for the same total amount of neuron-seconds. In this section, we first discuss the validation of the evaluation and then hypothesize that the multineuronal vectorizing procedure is more efficient than the time-segmental vectorizing procedure to extract information from neuronal activities. Finally, we compare the efficiencies of these procedures from the perspective of real-time signal processing and the control of portable prosthetic devices.

#### 4.1. Effectiveness of bias reduction measures

The calculated values of the amount of information may include some biases due to a limited number of trials or observations. Panzeri et al. (2007) reported that a combination of multiple bias reduction procedures is necessary to accurately estimate the amount of information carried by neurons. We combined three measures to reduce the biases in our study with the following:

2. The reduction of bias caused by pseudo relationships between stimuli and responses (Tovee et al., 1993; Treves & Panzeri, 1995).
3. The Bonferroni-Dunn post hoc test.
The intersection of the vertical and horizontal solid lines indicates values and the time segmental information values are plotted on the y-axis, respectively. The time-course of information obtained from the limited number of simulation trials (Fig. 2). Combining all of the bias reduction measures minimized the divergences of the upper and lower bounds from the true amount of transmitted information and narrowed the gap between the upper and lower bounds (Fig. 2(d)). When only Method 2 was implemented (Fig. 2(b) and (d)), the estimated lower bounds, $I_{seg}(S; D(R_{sp})) - I_{seg}(S; D(R_{sp}))$ and $I_{se}(S; D(R))$, tended to be lower than the true value. Thus, Method 2 was absolutely necessary to maintain the true value between the upper and lower bounds. Method 3 was also effective in reducing the bias included in the upper bound values (Fig. 2(a) and (b)). Consequently, we implemented all of these bias reduction measures to estimate the upper and lower bounds of the net amount of information.

### 4.2. Comparison of the two vectorizing procedures

We compared the amount of information obtained from the same neuron-seconds between the multineuronal and time-segmental vectorizing procedures. With the multineuronal vectorizing procedure, the upper and lower bounds of the amount of transmitted information increased with the number of neurons (Fig. 5(a)). In contrast, the time-segmental vectorizing procedure did not help increase the upper and lower bounds (Fig. 5(b)). The two-dimensional plot based on the same five-neuron groups (Fig. 7) indicates that the amount of information is always larger for the multineuronal vectorizing procedure than for the time-segmental vectorizing procedure (see also Fig. 5(c)). Our results are different from those of previous studies that reported an increase in the amount of information with increased time-segments (Optican & Richmond, 1987; Richmond & Optican, 1987; Tovee et al., 1993). This discrepancy may be caused by how the bias reduction measures were combined for the information analysis. In an early study of temporal coding in IT neurons, using only one bias reduction measure that was similar to Method 1, the temporal pattern was expected to give more than twice the information of the mean firing rate of a single neuron (Optican & Richmond, 1987). However, in a subsequent study (Tovee et al., 1993), the combination of bias reduction measures similar to Methods 1 and 2 resulted in a reduction of the estimated amount of information gained by temporal coding. Since we used three bias reduction measures (Methods 1, 2, and 3) in the present study, the gain in the amount of information by temporal patterns was statistically undetectable. Thus, the efficiency of temporal coding should be investigated in further studies.

This difference in efficiency between the two procedures was not accounted for by differences in the amount of transmitted information at $n = 1$, because the mean amount of information transmitted by the activities of single neurons in 1 s of the response period was the same for the multineuronal and time-segmental vectorizing procedures. The difference was caused by differences in the redundancy between the components of the response vectors. For a given neuron-second, responses were less redundant between two neurons than between two time-segments of a single neuron (Fig. 4(e)). In other words, the spike counts of two neurons were more independent than the spike counts of two time-segments from a single neuron.

### 4.3. Use of a vector for prosthetic control

The dimensions of the response vector should be reduced when applied to the real-time control of prosthetic devices, because an increase in the vector dimension may deplete the limited...
resources for processing signals and controlling devices. The use of simultaneously recorded responses of multiple neurons and the temporal patterns of neuronal activities can be used to effectively extract more information from neuronal activities (Aggelopoulos et al., 2005; Gochin et al., 1994; Kaneko, Tamura, Kawashima et al., 2007; Optican & Richmond, 1987; Richmond & Optican, 1987; Rolls et al., 2003; Tovee et al., 1993). However, the dimensions of the response vector increased the amount of information less efficiently when the temporal patterns of the neuronal responses were used in our study (Fig. 5(a) and (b)). Therefore, using multineuronal activities seems more advantageous to control an artificial limb in real time.

If the resources for processing signals and controlling devices can accommodate more information, then using both multineuronal and time-segmental vectorizing procedures may be more efficient than using either of them alone. The results of our experiments showed that the time-segmental vectorizing procedure did not increase the amount of transmitted information under the conditions with 1 total neuron-second (Fig. 5(b)). If the observation time increases and therefore the total neuron-seconds increases, then the efficiency of the time-segmental vectorizing procedure may also increase. This expectation stems from the observation that neuronal activities are transitional from state to state and never remain in a specific state (Muirpour & Esteky, 2008; Stern, Kincaid, & Wilson, 1997). This property of the two combined vectorizing procedures must be addressed in future studies.

5. Conclusions

To extract information more efficiently from neuronal activities, we compared two vectorizing procedures: one procedure used simultaneously recorded responses of multiple neurons and the other used temporally segmented responses of single neurons. When both procedures were examined for the same total neuron-seconds, the amount of information increased more efficiently with an increase in the dimensions of the response vector of multiple neurons than with an increase in the dimensions of the vector of the segmented activities of single neurons. Compared to using temporal patterns of single or smaller number of neurons, the use of multineuronal activities may be an advantageous strategy to control prosthetic devices.

Acknowledgements

This work was supported by grants from the National Institute of Advanced Industrial Science and Technology (HK and SS), the Ministry of Education, Culture, Sports, Science and Technology of Japan (HT, ST, and IF), and the Japan Science and Technology Agency, Core Research for Evolutional Science and Technology (IF).

Appendix. An alternative explanation of equations used in Method 3

Here we show that the equations for calculating $I_{\text{resp}}(S; R)$ and $I_{\text{resp}}(S; R)$ are mathematically similar to those used in the bias reduction measure proposed by Panzeri et al. (Montemurro et al., 2007; Panzeri et al., 2007). Because the mutual information between the response and stimulus can be explained by differences between the entropy of the response and the conditional entropy of the response to the stimulus (Cover & Thomas, 1991), each term in Eq. (6) can be rewritten as follows:

$$I_{\text{resp}}(S; R) = H_{\text{resp}}(R_{\text{raw}}) - H_{\text{resp}}(R_{\text{raw}}|S), \quad (A.1)$$

$$I_{\text{resp}}(S; R) = H_{\text{resp}}(R_{\text{daffle}}) - H_{\text{resp}}(R_{\text{daffle}}|S), \quad (A.2)$$

$$I_{\text{resp}}(S; R)_{\text{independent}} = H_{\text{resp}}(R_{\text{independent}}) - H_{\text{resp}}(R_{\text{independent}}|S). \quad (A.3)$$

Here the entropy of response $R_{\text{independent}}$ is approximately equal to that of $R_{\text{independent}}$, because shuffling the responses obtained from different trials for the same visual stimulus makes components of the response vector independent;

$$H_{\text{resp}}(R_{\text{independent}}) \approx H_{\text{resp}}(R_{\text{independent}}). \quad (A.4)$$

More specifically, if the order of the response vector is one ($n = 1$) or if the number of trials is infinite, then we have equality in Eq. (A.4). Thus, Eq. (6) can be represented as follows:

$$I_{\text{resp}}(S; R) \approx H_{\text{resp}}(R_{\text{raw}}) - H_{\text{resp}}(R_{\text{independent}}) + H_{\text{resp}}(R_{\text{daffle}}) - H_{\text{resp}}(R_{\text{daffle}}|S). \quad (A.5)$$

This equation is the same type of equation that was used in the bias reduction measure proposed by Panzeri et al. (2007). Eq. (7) for $I_{\text{resp}}(S; R)$ can be rewritten in the same way.

References


