Decoding visual object categories from temporal correlations of ECoG signals

Kei Majima a,b, Takeshi Matsuo c,d, Keisuke Kawasaki c, Kensuke Kawai d, Nobuhito Saito d, Isao Hasegawa c,e, Yukiyasu Kamitani a,b,*

a ATR Computational Neuroscience Laboratories, 2-2-2 Keihanna Science City, Kyoto 619-0288, Japan
b Department of Neurosurgery, The University of Tokyo Graduate School of Medicine, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan
c Department of Physiology, Niigata University School of Medicine, 1-757 Asahi-machi St., Chuo-ku, Niigata 951-8510, Japan
d Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan
e Center for Transdisciplinary Research, Niigata University, 1-757 Asahi-machi St., Chuo-ku, Niigata 951-8510, Japan

Abstract
How visual object categories are represented in the brain is one of the key questions in neuroscience. Studies on low-level visual features have shown that relative timings or phases of neural activity between multiple brain locations encode information. However, whether such temporal patterns of neural activity are used in the representation of visual objects is unknown. Here, we examined whether and how visual object categories could be predicted (or decoded) from temporal patterns of electrocorticographic (ECoG) signals from the temporal cortex in five patients with epilepsy. We used temporal correlations between electrodes as input features, and compared the decoding performance with features defined by spectral power and phase from individual electrodes. While using power or phase alone, the decoding accuracy was significantly better than chance, correlations alone or those combined with power outperformed other features. Decoding performance with correlations was degraded by shuffling the order of trials of the same category in each electrode, indicating that the relative time series between electrodes in each trial is critical. Analysis using a sliding time window revealed that decoding performance with correlations began to rise earlier than that with power. This earlier increase in performance was replicated by a model using phase differences to encode categories. These results suggest that activity patterns arising from interactions between multiple neuronal units carry additional information on visual object categories.

Keywords:
ECoG
IT cortex
Object category
Decoding
Temporal coding

Introduction
Response selectivity to visual object categories is a hallmark of the temporal visual cortex. Single neurons in the inferior temporal cortex (IT) respond to specific object categories such as faces, hands, and buildings (Desimone et al., 1984; Kreiman et al., 2006; Perrett et al., 1982; Tanaka, 1996; Tsao et al., 2006). While previous studies found that object category information can be coded in the activity of a single neuron or region, other studies have shown that object categories are represented by activity patterns of a population of neurons in the monkey inferotemporal (IT) cortex (Kiani et al., 2007) and fMRI voxels in the human IT cortex (Kriegeskorte et al., 2008).

While analyses of multiple neurons or voxels have revealed detailed object representations, these studies might have overlooked the possibility that correlations between neuronal units in single-trial signals contribute to object representation. Theoretical studies have suggested that neural representations can be achieved using temporal patterns over multiple neuronal units, including the order of response latencies (Gautrais and Thorpe, 1998; Thorpe et al., 2001; Van Rullen et al., 1998), spike sequences with millisecond precision (Abeles, 1991; Abeles et al., 1993; Lesiennne and Strehler, 1987; Oram et al., 1998), and gamma-band synchronization (Eckhorn et al., 1988; Engel et al., 1991; Gray et al., 1989). Experimental results have also shown that such temporal patterns encode information on low-level visual features, such as light intensity in the retina (Golisch and Meister, 2008), orientation in the primary visual cortex (Celebrini et al., 1993; Gawne et al., 1996; Shriki et al., 2012), and co-occurrence of edges (Eckhorn et al., 1988; Engel et al., 1991; Gray et al., 1989). Thus, it is of great interest whether temporal patterns are used to represent visual object categories.

In this study, we used electrocorticogram (ECoG) to record object-evoked neural responses in the temporal cortex from five patients with epilepsy. ECoG allowed for simultaneous, high-temporal resolution measurement of neural responses at multiple sites over a wide range of the temporal cortex, which is generally difficult to perform using fMRI or single unit recording. We calculated the temporal correlations between...
ECO G electrodes using single-trial time series, and constructed a statistical classifier (decoder) to predict the category of a presented object from those signal features on a trial-by-trial basis (Kamitani and Tong, 2005). Temporal correlation measures the degree of in-phase synchronization between two time series (Varela et al., 2001). For comparison, we used the spectral power or phase values of ECoG signals in individual electrodes as features, and compared the decoding performances to examine whether temporal correlations carry additional information on object categories.

In addition, we investigated whether ECoG responses that produce informative correlations are time-locked to the stimulus or not. Informative correlations could be generated by category specific ECoG responses with constant latencies across trials, or alternatively by relative time series of ECoG signals with variable latencies across trials. To examine this, we conducted a shuffling analysis (Averbeck et al., 2006; Yamashita et al., 2008). In this analysis, we created data in which relative timing/phase differences between electrodes within each trial were destroyed (trial-shuffled data) from the original ECoG data by randomly permuting trials with the same stimulus in each electrode. We compared the decoding performance using the shuffled data to the decoding performance of the original data.

To characterize the time course of decoding performance, we calculated decoding performance as a function of time (see Liu et al., 2009) and compared this between spectral power values and temporal correlations. Finally, to specify which encoding model explained the results of the time course analysis at the neuronal electric source level, we performed a simulation analysis in which two representative models were tested; a model where sources encode information with their response latencies (Celebrini et al., 1993; Gautrais and Thorpe, 1998; Gollisch and Meister, 2008; Shriki et al., 2012; Van Rullen et al., 1998), and a model where sources encode using their phase differences (Eckhorn et al., 1988; Engel et al., 1991; Gray et al., 1989). The results from these analyses suggest that activity patterns arising from interactions between multiple neuronal units carry a significant amount of information on visual object categories.

Materials and methods

Subjects

Five patients with medically intractable epilepsy (5 female, 22–42 years, Table 1) participated in our experiments. All patients were admitted to The University of Tokyo Hospital, Japan. Patients underwent electrode implantation for the purpose of localizing seizure foci to guide neurosurgical treatment (Fig. 1). The locations of electrodes were determined solely by therapeutic considerations. We obtained written informed consent from the patients, and all experimental protocols were approved by the institutional review board at the hospital (#1797(3)).

ECoG recordings

Patients were implanted with subdural electrode arrays arranged in grids or strips (Unique Medical Co., Ltd., Tokyo, Japan). Each grid/strip contained 4–20 electrodes. Each electrode contact was 3 mm in diameter with 10 mm separation, or 1.5 mm in diameter with 5 mm separation. The number and location of the recording sites in the temporal, occipital, and frontal lobes were determined exclusively by clinical criteria. Data were obtained from a total of 628 recording sites, with 120–127 electrodes per subject. Recorded signals were amplified using a reference electrode placed on the scalp, filtered between 0.55 and 150 Hz and sampled at 400 Hz (Nicoleton, CareFusion, San Diego, CA, USA). In addition to this bandpass filtering, the signals were filtered to pass each of five frequency bands (1–10, 10–20, 20–30, 30–80 and 80–150 Hz) when calculating ECoG features in bands of interest. All data were acquired during periods without seizure events.

Electrode localization

To localize electrodes, we integrated the anatomical information of the brain provided by preoperative magnetic resonance imaging (MRI), and spatial information of the electrode positions provided by postoperative computer tomography (CT). For each subject, the 3D brain surface was reconstructed and an automatic registration based on mutual information was performed using Avizo (Maxnet Co., Ltd., Tokyo, Japan). Because the location of the recording site depended on clinical criteria, various ventral and lateral cortical areas were evaluated for each patient, and the coordinate of each electrode contact in their stereotactic scheme was measured (using the Talairach coordinate system). Coordinates were used to anatomically localize contacts using the proportional atlas of Talairach and Tournoux (1993), after a linear scale adjustment to correct size differences between the patient’s brain and the Talairach model using Statistical Parametric Mapping, Version 8 software (SPM8, Wellcome Trust, London, UK) (http://www.fil.ion.ucl.ac.uk/spm/). Electrodes on the IT cortex that consists of the middle and inferior temporal gyri were selected and used for the analysis of this study (55 electrodes for S1, 84 electrodes for S2, 117 electrodes for S3, 104 electrodes for S4, 106 electrodes for S5; Fig. 1, Table 1).

Stimulus presentation

We prepared 120 colored photographs of objects from 24 different categories as stimuli (Fig. 2A). There were five different exemplars per category. The stimuli were presented on a 27-inch LCD monitor at a viewing distance of 57 cm with a central fixation point (0.5°) (Fig. 2B). Each stimulus subtended a 6° × 6° visual angle and was presented for 300 ms followed by a 900-ms interval period (Fig. 2C). The stimuli were presented in pseudorandom order and each stimulus was presented either 10 or 11 times per subject. We instructed the subjects to fix their eyes on the fixation point and to perform a one-back task, indicating whether an exemplar was repeated successively or not by pressing a button (55–120 repetition trials per subject). We excluded trials with button presses from any analyses reported in this study.

ECoG features

Signal features used as input to decoding analyses were calculated from single-trial time series of the individual electrodes in the temporal

| Subject | Age | Sex | Task performance (success/total) | # of all implanted electrodes | # of electrodes on IT cortex (left side/left bottom/right side/right bottom) |
|---------|-----|-----|---------------------------------|-----------------------------|-------------------------------------------------------------------|
| S1      | 42  | F   | No record                       | 120                         | 55 (15/40/0/0)                                                    |
| S2      | 36  | F   | 93/112                          | 127                         | 84 (10/22/10/32)                                                  |
| S3      | 22  | F   | 108/113                         | 127                         | 117 (5/52/5/55)                                                  |
| S4      | 35  | F   | 111/120                         | 127                         | 104 (0/52/0/52)                                                  |
| S5      | 38  | F   | 53/55                           | 127                         | 106 (2/52/0/52)                                                  |
cortex within a time window from 0 to 300 ms relative to the stimulus onset or a 100/300-ms sliding time window with a step size of 25 ms.

“Power features” were the power values of five frequency bands (1–10, 10–20, 20–30, 30–80 and 80–150 Hz) calculated from the power spectrum for each time window (the number of electrodes times the five frequency bands). “Correlation features” were the correlation coefficients of time series for all pairs of electrodes within each time window (1485–6786 features for each subject). In the analysis to compare specific frequency bands (the five bins of power spectral frequency), power features were limited to those in the frequency bands of interest. Correlation features were re-calculated from the band-pass filtered time series.

Correlation features are higher-order variables based on the products of (normalized) signal amplitudes between electrodes at each time point. As a result, the number of available features is much larger than that of power features. To control for the effect of higher-order feature extraction and the number of available features, we used the products of power values as an additional feature type (“product-of-power features”). In each frequency band, the products of the power values were calculated for all pairs of electrodes (five [frequency bands] times the number of correlation features).

We also used the phase values of ECoG time series as features (Hammer et al., 2013; Lopour et al., 2013). We calculated the phase values of five frequencies (6, 16, 26, 56 and 116 Hz; near-middle values of the five frequency bands) in each time window using the fast Fourier transform (FFT), and took the sine and cosine values. These values were used as “phase features” (the number of electrodes \( \times 5 \) [frequencies] \( \times 2 \) [trigonometric functions]). Note that individual phase features do not explicitly indicate relative timings between electrodes, although the linear decoder may detect relative timings via the weighted summation of phase features.

In the shuffling analysis to characterize informative correlations, correlation features were calculated with ECoG data in which relative timing/phase differences between electrodes within each trial were destroyed by a trial-shuffling procedure (Averbeck et al., 2006; Yamashita et al., 2008). In each electrode, we randomly permuted the data across trials for the same visual stimulus, while preserving the ECoG time series within each electrode and trial. In these shuffled data, if ECoG responses to each category were constant across trials (time-locked to the stimulus onset), the original correlations would be preserved (Fig. 3A). However, if the original correlations were from the relative time series in each trial (not time-locked to the stimulus onset) and the variability of response latencies across trials was sufficiently large, the correlations would be removed by the shuffling (Fig. 3B).

Feature vectors were created from individual trials for a given time window in each subject. A feature vector consisted of the features described above or the concatenated features of powers and correlations, each calculated from a time window of each trial. Before training, we conducted a feature normalization procedure and a feature selection procedure (see the next section).

**Decoding procedure**

We constructed a linear classifier (decoder) to predict the categories of presented stimuli from ECoG feature vectors on a trial-by-trial basis. Feature vectors labeled by the stimulus categories of individual trials in each subject were divided into training and test datasets, and a linear support vector machine (SVM; Vapnik, 1998) algorithm determined the parameters of the decoder using the training dataset. The decoder calculated the linearly weighted sum of the features plus a bias for each category (class) given a feature vector from the test dataset. The
category with the maximum value was chosen as the predicted category (Kamitani and Tong, 2005). The values of each feature were normalized using the sample mean and standard deviation calculated with the training dataset. The dimensionality of the feature vector was reduced by selecting informative features based on a univariate analysis (F-statistics) applied to the training dataset. We ranked the features by the F-value that indicated differential responses to the categories, and the top \( N \) features were used as inputs to the decoder. The number of used features (\( N \)) was decided by a cross-validation analysis within the training dataset, where \( N \) was varied from 50 to 1000 in increments of 50, and the \( N \) with the highest accuracy was chosen (nested cross-validation). The average numbers of selected features for the analysis with the 0–300-ms time window were 70 ± 63 (mean ± SD across subjects) for power features, 694 ± 404 for correlation features, and 778 ± 330 for the combined features. To see the dependence on the number of features, we also performed cross-validation analysis with a fixed \( N \) (without the nested cross-validation to optimize \( N \)), and decoding performance was calculated for each feature type while \( N \) was varied (from 10 to 1000 increased by 10).

To evaluate generalization performance for category classification across different exemplars, we ensured that trials corresponding to the same visual stimuli were not included in both the training and test datasets (Vindiola and Wolmetz, 2011). We divided the 120 stimuli into five groups, each of which contained 24 stimuli from the 24 different categories and divided the corresponding trials into five groups. Four groups were then used to train a decoder and the remaining group was used for evaluating the trained classifier. This procedure was repeated until the trials from all five groups were tested (5-fold cross-validation), and the percentage of correct classification was calculated. The cross-validation for determining the number of features was performed using the four groups in the training dataset (4-fold cross-validation).

To compare decoding accuracy between conditions, we used a chi-square test for within-subject analysis, and a paired t-test for group analysis. For the group analysis, we calculated the logit-transformed accuracy of each subject, \( \log \left( \frac{a}{1 - a} \right) \) where \( a \) is an accuracy, and then applied a paired t-test to those values. We used logit-transformed accuracies rather than the original ones because a normal distribution has infinite support while accuracies are bounded from zero to one, and the assumption that obtained accuracies follow a normal distribution is
models to take into account category-specific time. In addition, we also tested modifications by each model, we constructed decoders using either spectral power or recording condition in the experiments. Using simulated data produced signals were (24 × 50 = 1200 trials in total), to match the empirical data. Electrode number of categories was set to 24 and each category had 50 trials spatial sums of the sources with a lead intervals. The signals of the ECoG electrodes were expressed by the ECoG electrodes were placed 1 mm above those sources with 10 mm intervals. The phase values were then reset to values specific to the stimulus category at 130 ms. The amplitudes after 130 ms were multiplied by category-specific values randomly chosen from one to (M + 1). The parameter M was manipulated to control the degree of dominance of amplitude coding in the model. When M = 0, both models had no amplitude modulation. Power and correlation features were calculated from a 100-ms sliding time window. The correlations for all electrode pairs and the power values for all electrodes and the five frequency bands were used as input features with no feature selection.

Simulation analysis

To perform a simulation analysis, we assumed 200 neural signal sources arranged in a one-dimensional array with 1-mm intervals. Ten ECoG electrodes were placed 1 mm above those sources with 10 mm intervals. The signals of the ECoG electrodes were expressed by the spatial sums of the sources with a lead field matrix (Nunez, 1981). In the latency coding model, each source had a fixed waveform produced by sampling from an i.i.d. Gaussian distribution \( N(0, 3^2) \). Given a stimulus, each source produced a waveform with a latency specific to the stimulus category. Those latency values were uniformly distributed from 130 to 180 ms over stimulus categories. In each trial, the latencies of all sources were jittered by the same amount randomly chosen from 0 to 50 ms (while preserving relative timings across trials), and independent Gaussian white noise \( N(0, 3^2) \) was added. In the phase reset model, the sources were oscillators at a frequency of 40 Hz with an amplitude of one. In each trial, the initial phase values of all sources were randomly chosen. The phase values were then reset to values specific to the stimulus category at the reset time, which was randomly chosen for each trial between 130 and 180 ms. Independent Gaussian white noise \( N(0, 3^2) \) was added to all sources in each trial. The number of categories was set to 24 and each category had 50 trials \((24 \times 50 = 1200\) trials in total), to match the empirical data. Electrode signals were filtered between 0.55 and 150 Hz, which was similar to the recording condition in the experiments. Using simulated data produced by each model, we constructed decoders using either spectral power or correlation features, and calculated the decoding performance as a function of time. In addition, we also tested modified versions of the two models to take into account category-specific amplitude modulation of sources. It is commonly assumed that the amplitudes of neuronal sources change depending on the stimulus (Naruse et al., 2010; Palva and Palva, 2007). In the latency coding model, each source signal before noise addition was multiplied by a value specific to the given stimulus category, which was randomly chosen from one to \((M + 1)\) in each category, for each electrode. In the phase reset model, the amplitude of each source was changed from one to a value specific to the given stimulus category at 130 ms. The amplitudes after 130 ms were multiplied by category-specific values randomly chosen from one to \((M + 1)\). The parameter M was manipulated to control the degree of dominance of amplitude coding in the model. When M = 0, both models had no amplitude modulation. Power and correlation features were calculated from a 100-ms sliding time window. The correlations for all electrode pairs and the power values for all electrodes and the five frequency bands were used as input features with no feature selection.

Results

We measured ECoG data from implanted electrodes located predominantly in the temporal cortex; data from five subjects were recorded while they sequentially viewed natural images (Fig. 2C). We confirmed that all data were acquired during periods without seizure events. Electrodes on the IT cortex that consists of the middle and inferior temporal gyri were selected and used for the analysis of this study (Fig. 1, Table 1).

To illustrate the category selectivity of spectral power and temporal correlations, we calculated the trial-averaged time courses of power values in the 1–10 Hz spectral band and that of temporal correlations responding to each stimulus category. We plotted the time courses of the power values at two representative electrodes for the face category and the letterstring category, and those of the temporal correlation between the electrodes (Fig. 4). In this example, the temporal correlation showed category-selective time courses whereas the modulation of power at the two electrodes is not distinctive.

To evaluate information encoded in each type of feature, we performed decoding analysis using the power, correlation and combined features calculated from the ECoG time series from 0 to 300 ms relative to the stimulus onset (24 categories; chance level, 4.16%). Fig. 5A shows the cross-validated performances for the three feature types in each subject, using the number of features optimized by a nested cross-validation (see Materials and methods). Performance exceeded the chance level for all feature types and all subjects \((P < 10^{-5}, \text{binomial test})\). By combining power and correlation features, the decoding performance was substantially improved compared with the performance using power alone \((P < 0.05, \text{paired t-test across subjects}; P < 0.05 \text{ in all individual subjects, chi-square test})\). Even correlation features alone significantly outperformed power features \((P < 0.05, \text{paired t-test across subjects}; P < 0.05 \text{ in four out of five subjects, chi-square test})\).

Fig. 5B shows the mean performances as a function of the number of used features. In addition to power, correlation, and combined features, we plotted the results for the products of power values (product-of-power features), which were introduced to control for the effect of higher-order feature extraction and the number of available features (see Materials and methods). The performances with power and product-of-power features peaked around 50 features, while the performances with correlation and combined features continued to improve, even with >500 features. When compared with the same number of features used, correlation and combined features outperformed power features even at small numbers of features (around 50–250 features). Furthermore, the decoding performance with product-of-power features remained lower than those with the other features, even though the number of available features was the largest. Thus, the performance improvement by adding correlation features (Fig. 5A) cannot be attributed to the larger number of used or available features. Temporal correlations appear to carry additional information about object categories.
that is not represented by power or higher-order features generated from power.

Correlation features reflect phase or timing differences of ECoG time series between electrodes. To see whether phase information in individual electrodes is sufficient to achieve the high decoding performance obtained with correlations, we tested decoding accuracies using phase features (the number of features optimized by a nested cross-validation; see Materials and methods). Fig. 6 shows the mean performances obtained with power, phase, and correlation features. The decoding performance obtained with correlation features was better than the others \((P < 0.05,\) paired \(t\)-test). Although phase features could implicitly encode phase or timing differences between electrodes, the results suggest that a more explicit representation of the differences by correlations was critical for achieving the high decoding performance. Informative correlations could be generated by category specific ECoG responses with constant latencies across trials (stimulus-locked responses), or alternatively by relative time series of ECoG signals with variable latencies across trials. To examine which type of response produced informative correlations, we created trial-shuffled data (see Materials and methods) and compared the decoding performance between the original and shuffled datasets. Fig. 7 shows the decoding performance using the original and shuffled data for each frequency band. Using the data before the band-pass filtering (“All”), the shuffling degraded the performance in most subjects \((P < 0.05,\) chi-square test). Among the five frequency bands, the 30–80 Hz and the 80–150 Hz bands showed performance degradation \((P < 0.05\) in 5/5 and 2/5 subjects, respectively). At the group level, the decoding performance in the 30–80 Hz was significantly degraded by shuffling \((P < 0.05,\) paired \(t\)-test, corrected for multiple comparison). While the overall performance was highest in the 1–10 Hz band, the performance was comparable between the original and shuffled data. The 10–20 Hz and 20–30 Hz bands showed poor performance in both of the original and shuffled data. These results suggest that stimulus-unlocked, relative time series may contribute to category coding in the high frequency bands. Informative correlations in the low frequency band may arise in a stimulus-locked manner, but it is also possible that the variability...
of response latencies across trials was too small compared to the low frequency cycle to destroy correlations from relative time series.

To characterize the time course of decoding performance, we used a sliding time window and calculated the decoding performance as a function of time. The time window was shifted by 25 ms, and decoding performance was plotted for the power and correlation features (Fig. 8, left). In this analysis, power values from five frequency bands and correlations for the raw ECoG signals were used as input features, respectively. The decoding performance with the 300-ms time window began to rise when the center of the time window was earlier than 100 ms after the stimulus onset and remained above chance level even after the disappearance of the stimulus (300 ms). To evaluate when decoding performance began to rise from chance level, we defined the onset time as the first point at which the mean decoding performance over subjects exceeded the 99% percentile (% correct = 5.4%) in the performance distribution at 0 ms. The decoding performance with correlations rose earlier than that with power (25 ms and 100 ms for correlation and power features, respectively). To see the dependence on the width of the used time window, we calculated decoding accuracies with a 100-ms sliding time window (Fig. 8, right). Even with the shorter time window, the onset time for correlations was earlier than that for power. The peak performance for correlations was slightly lower than that for power with this short time window, presumably because correlations could be better estimated using a longer time series. The results indicate the capacity of temporal correlations between electrodes to encode category information in early responses.

To examine what mechanism could explain the early onset of the good decoding performance with correlations, we tested two representative models: a latency coding model and a phase reset model, where category-specific temporal patterns were produced by latency differences and by phase differences across sources, respectively (Figs. 9A, B; see Materials and methods; Eckhorn et al., 1988; Engel et al., 1991; Gautrais and Thorpe, 1998; Gray et al., 1989; Thorpe et al., 2001; Van Rullen et al., 1998).

In the latency coding model, decoding performance began to rise around the same time for both power and correlation features in all conditions (Fig. 9C), failing to account for the dissociation between power and correlation features. This is presumably because changes in power and correlation are tightly coupled via response latencies. In contrast, for the phase reset model, the performance with correlations began to rise earlier compared with power, except when the amplitude modulation was strong (M = 10; Fig. 9D, bottom). Phase resetting may cause only subtle differences in power via the summation of synchronized/desynchronized neuronal sources, resulting in a slower onset of decoding performance with power features.

### Discussion

In the present study, we have shown that temporal correlations of ECoG signals between electrodes provide additional information on visual object categories compared to information represented by spectral power or phase in individual electrodes (Figs. 5 and 6). The trial-shuffling procedure degraded decoding performance obtained with correlations for raw and gamma band ECoG signals (Fig. 7), indicating that relative time series between electrodes contain information about categories. Time course analysis using a sliding time window revealed that decoding performance obtained with correlations began to rise earlier compared with power (Fig. 8). In the simulation analysis, this difference between power and correlation features was reproduced in a model where neuronal electric sources encode information using their phase differences (Fig. 9). These results suggest that not only response amplitudes but also temporal correlations over multiple brain areas carry information on visual object categories, and that informative correlations can be derived from category-specific, relative time series of neural activity. The simulation results suggest that temporal correlations may reflect phase differences of neuronal electric sources in the temporal cortex at least in the early response period.

Although we focused on correlations of ECoG time series between electrodes without a time lag, the use of different features that take into account interactions between two different points in space and time could improve decoding performance. In an EEG study, it was reported that estimated off-diagonal coefficients in a multivariate autoregressive model are more informative than diagonal ones in classifying sleep stages (Penny and Roberts, 2002). It remains to be seen whether such sophisticated features are efficient for the decoding of neural object representations.

Several previous studies have proposed the possibility that visual information is encoded in temporal relations of spikes or field potentials between multiple brain locations, and more specifically, in the order of spike latencies (Gautrais and Thorpe, 1998; Thorpe et al., 2001; Van Rullen et al., 1998), spike sequences with millisecond precision (Abeles, 1991; Abeles et al., 1993; Lestienne and Strehler, 1987; Oram et al., 1999), and gamma band synchronization (Eckhorn et al., 1988; Engel et al., 1991; Gray et al., 1989). Some experimental results have shown that such patterns encode low-level visual information, such as ...
light intensity in the retina (Gollisch and Meister, 2008), orientation in the primary visual cortex (Celebrini et al., 1993; Gawne et al., 1996; Shriki et al., 2012), and co-occurrence of edges (Eckhorn et al., 1988; Engel et al., 1991; Gray et al., 1989). Although our experimental data are not at the level of single neurons, ECoG signals have been assumed to provide an aggregate signature of transmembrane currents within a cortical area of several millimeters (Buzsáki et al., 2012; Reimann et al., 2013; Varela et al., 2001). Thus, informative correlations across electrodes in the IT cortex shown in our present study may derive from such temporal patterns encoding visual object categories.

Our shuffling procedure degraded decoding performance obtained with correlation features for the raw ECoG signals and the signals from the two highest frequency bands (Fig. 7). These results suggest that informative correlations partly arise from relative time series that are specific to each category and not time-locked to the stimulus onset across trials. On the other hand, the shuffling procedure did not degrade decoding performance in the three lower frequency bands. This may imply that informative correlations in those bands arise in a stimulus-locked manner. However, it is also possible that even if relative time series contributed to information coding in the lower frequency bands, the fluctuations of response delays across trials were too small relative to the low frequency cycles to have a substantial effect from shuffling.

We found that decoding performance with temporal correlations increased earlier than that with power values, and that the difference in the timing of the performance increase was reproduced by our phase reset model, where object categories were coded by phase differences rather than latency differences across neuronal sources. While latency differences can cause robust patterns in both power and correlation features, phase resetting may produce distinct patterns in correlation features but not in power features. Even in the phase reset model, spectral power could be modulated via the synchronization or desynchronization of neuronal sources near each electrode. However, the detection of such small changes in power in the presence of noise may require a broader temporal summation, resulting in a delayed increase in decoding performance. The results suggest that at least in the early response period,
informative correlations originate from subtle temporal patterns like phase differences. In our experiment, we did not rigorously match lower-level visual features across stimulus exemplars. The spatial frequency of an image has been known to be an important feature used by the visual system to assign category membership (Crouzet and Thorpe, 2011). Single neurons in the IT cortex have been shown to show selectivity to specific complex shapes, which can be considered as components of objects (Tanaka, 1996). Our results are based only on population-level selectivity to object categories, and further analysis will be necessary to understand the exact nature of neural representation in IT that underlies object decoding.

A critical advantage of the use of ECoG is that it allows recording of electric signals simultaneously from multiple sites in the temporal cortex at high temporal resolution. In our shuffling analysis, we examined whether informative correlations derived from category-specific relative time series in each trial. This kind of analysis requires simultaneous recording of neural signals from multiple brain locations with high temporal resolution. In some studies using single-unit recordings, the signal from each electrode was recorded separately because of the difficulty of simultaneous recording from multiple sites (Georgopoulos and Massey, 1988; Gochin et al., 1994; Hung et al., 2005; Kiani et al., 2007; McAdams and Maunsell, 1999; Rolls et al., 1997). However, with this method, stimulus-unlocked patterns over multiple brain locations may be overlooked (Averbeck et al., 2006). Recording with a microelectrode array might be a promising method for such analysis, but the coverage of current state-of-the-art in vivo arrays is limited to an area of several millimeters and is not sufficient to cover the whole temporal cortex (Fejtl et al., 2006). Although other neuroimaging techniques such as fMRI, scalp EEG, and MEG provide simultaneous recordings from wide brain regions, the temporal resolution of fMRI is limited to the timescale of seconds, and the spatial resolution of scalp EEG and MEG is insufficient to reveal synchronization within two or less centimeters (Varela et al., 2001). ECoG provides better spatiotemporal resolution than EEG, MEG, and fMRI, making ECoG a useful tool to examine theoretical hypotheses about fine temporal coding. The combination of ECoG recordings and neural decoding techniques has become a powerful approach to the investigation of neural representations in the last decade (Chao et al., 2010; Hammer et al., 2013; Liu et al., 2009; Pasley et al., 2012; Tsuchiya et al., 2008; Yanagisawa et al., 2009). In addition, new, high-density ECoG electrode arrays are being developed (Hollenberg et al., 2006; Matsu et al., 2011; Rubeh et al., 2009; Toda et al., 2011; Watanabe et al., 2012; Yeager et al., 2008). The use of these electrode arrays in monkeys and other animals may reveal even more detailed neural representations.

Acknowledgments
We thank Tomoyasu Horikawa, Paul Sukhanov, Atsushi Matsui, Nancy Wang, Makoto Takeymiya, Satoshi Hirose, Makoto Fukushima, Kazushi Ikeda, and Yoichi Miyawaki for helpful discussion and comments. This work was supported by JSPS KAKENHI (11)08024 and 21390405, SRPBS from MEXT, Grant for Promotion of Niigata University Research Project, Grant for Comprehensive Research on Disability, Health and Welfare (H23–Nervous and Muscular-General-003) from MHLW, and 2008 Specified Research grant from Takeda Science Foundation.

Conflict of interest
The authors declare no competing financial interests.

References
Abeles, M., 1991. Corticonics: Neural circuits of the cerebral cortex. Cambridge UP, New York.

Abeles, M., Bergman, H., Margalit, E., Vaadia, E., 1993. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J. Neurophysiol. 70, 1629–1638.

Averbeck, B.B., Latham, P.E., Pouget, A., 2006. Neural correlations, population coding and computation. Nat. Rev. Neurosci. 7, 358–366. 

Buzsáki, G., Anastassiou, C.A., Koch, C., 2012. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. 13, 407–420.

Celebrini, S., Thorpe, S., Trotter, Y., Imbert, M., 1993. Dynamics of orientation coding in in vivo arrays is limited to an area of several millimeters and is not sufficient to cover the whole temporal cortex (Fejtl et al., 2006). Although other neuroimaging techniques such as fMRI, scalp EEG, and MEG provide simultaneous recordings from wide brain regions, the temporal resolution of fMRI is limited to the timescale of seconds, and the spatial resolution of scalp EEG and MEG is insufficient to reveal synchronization within two or less centimeters (Varela et al., 2001). ECoG provides better spatiotemporal resolution than EEG, MEG, and fMRI, making ECoG a useful tool to examine theoretical hypotheses about fine temporal coding. The combination of ECoG recordings and neural decoding techniques has become a powerful approach to the investigation of neural representations in the last decade (Chao et al., 2010; Hammer et al., 2013; Liu et al., 2009; Pasley et al., 2012; Tsuchiya et al., 2008; Yanagisawa et al., 2009). In addition, new, high-density ECoG electrode arrays are being developed (Hollenberg et al., 2006; Matsu et al., 2011; Rubeh et al., 2009; Toda et al., 2011; Watanabe et al., 2012; Yeager et al., 2008). The use of these electrode arrays in monkeys and other animals may reveal even more detailed neural representations.

Acknowledgments
We thank Tomoyasu Horikawa, Paul Sukhanov, Atsushi Matsui, Nancy Wang, Makoto Takeymiya, Satoshi Hirose, Makoto Fukushima, Kazushi Ikeda, and Yoichi Miyawaki for helpful discussion and comments. This work was supported by JSPS KAKENHI (11)08024 and 21390405, SRPBS from MEXT, Grant for Promotion of Niigata University Research Project, Grant for Comprehensive Research on Disability, Health and Welfare (H23–Nervous and Muscular-General-003) from MHLW, and 2008 Specified Research grant from Takeda Science Foundation.

Conflict of interest
The authors declare no competing financial interests.

References
Abeles, M., 1991. Corticonics: Neural circuits of the cerebral cortex. Cambridge UP, New York.

Abeles, M., Bergman, H., Margalit, E., Vaadia, E., 1993. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J. Neurophysiol. 70, 1629–1638.

Averbeck, B.B., Latham, P.E., Pouget, A., 2006. Neural correlations, population coding and computation. Nat. Rev. Neurosci. 7, 358–366. 

Buzsáki, G., Anastassiou, C.A., Koch, C., 2012. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. 13, 407–420.

Celebrini, S., Thorpe, S., Trotter, Y., Imbert, M., 1993. Dynamics of orientation coding in in vivo arrays is limited to an area of several millimeters and is not sufficient to cover the whole temporal cortex (Fejtl et al., 2006). Although other neuroimaging techniques such as fMRI, scalp EEG, and MEG provide simultaneous recordings from wide brain regions, the temporal resolution of fMRI is limited to the timescale of seconds, and the spatial resolution of scalp EEG and MEG is insufficient to reveal synchronization within two or less centimeters (Varela et al., 2001). ECoG provides better spatiotemporal resolution than EEG, MEG, and fMRI, making ECoG a useful tool to examine theoretical hypotheses about fine temporal coding. The combination of ECoG recordings and neural decoding techniques has become a powerful approach to the investigation of neural representations in the last decade (Chao et al., 2010; Hammer et al., 2013; Liu et al., 2009; Pasley et al., 2012; Tsuchiya et al., 2008; Yanagisawa et al., 2009). In addition, new, high-density ECoG electrode arrays are being developed (Hollenberg et al., 2006; Matsu et al., 2011; Rubeh et al., 2009; Toda et al., 2011; Watanabe et al., 2012; Yeager et al., 2008). The use of these electrode arrays in monkeys and other animals may reveal even more detailed neural representations.

Acknowledgments
We thank Tomoyasu Horikawa, Paul Sukhanov, Atsushi Matsui, Nancy Wang, Makoto Takeymiya, Satoshi Hirose, Makoto Fukushima, Kazushi Ikeda, and Yoichi Miyawaki for helpful discussion and comments. This work was supported by JSPS KAKENHI (11)08024 and 21390405, SRPBS from MEXT, Grant for Promotion of Niigata University Research Project, Grant for Comprehensive Research on Disability, Health and Welfare (H23–Nervous and Muscular-General-003) from MHLW, and 2008 Specified Research grant from Takeda Science Foundation.

Conflict of interest
The authors declare no competing financial interests.

References
Abeles, M., 1991. Corticonics: Neural circuits of the cerebral cortex. Cambridge UP, New York.
