IBRO Neuroscience Reports 10 (2021) 171–177

Contents lists available at ScienceDirect

IBRO Neuroscience Reports

journal homepage: www.sciencedirect.com/journal/IBRO-Neuroscience-Reports

Research Paper

Object discrimination performance and dynamics evaluated by inferotemporal cell population activity

Ridey H. Wang, Lulin Dai, Jun-ya Okamura, Takayasu Fuchida, Gang Wang *

Dept. of Bioengineering, Graduate School of Science and Engineering, Kagoshima University, Kagoshima 890-0065, Japan

ARTICLE INFO

Keywords:
Object recognition
View-invariance
Monkey
Inferotemporal cortex
Discrimination
Learning

ABSTRACT

We have previously reported an increase in response tolerance of inferotemporal cells around trained views. However, an inferotemporal cell usually displays different response patterns in an initial response phase immediately after the stimulus onset and in a late phase from approximately 260 ms after stimulus onset. This study aimed to understand the difference between the two time periods and their involvement in the view-invariant object recognition. Responses to object images with and without prior experience of object discrimination across views, recorded by microelectrodes, were pooled together from our previous experiments. With a machine learning algorithm, we trained to build classifiers for object discrimination. In the early phase, the performance of classifiers created based on data of responses to the object images with prior training of object discrimination across views did not significantly differ from that based on data of responses to the object images without prior experience of object discrimination across views. However, the performance was significantly better in the late phase. Furthermore, compared to the preferred stimulus image in the early phase, we found 2/3 of cells changed their preference in the late phase. For object images with prior experience of training with object discrimination across views, a significant higher percentage of cells responded in the late phase to the same objects as in the early phase, but under different views. The results demonstrate the dynamics of selectivity changes and suggest the involvement of the late phase in the view-invariant object recognition rather than that of the early phase.

Introduction

It is easy to distinguish between different objects. However, understanding the underlying neural basis for object recognition is one of the most challenging tasks for neuroscientists. If unique features represent an object, its recognition is instantaneous regardless of the viewing angle changes (Biederman, 1987). However, an unfamiliar object cannot be discriminated from similar ones with changes in viewing angle (Bülthoff and Edelman, 1992; Logothetis et al., 1994; Tarr, 1995). An additional learning in object recognition is necessary. It is assumed that the capability of view-invariant recognition develops as different views of an object become associated while seeing rotating objects either through active learning or the passive experience of successive object views (Foldiak, 1991; Wiskott and Sejnowski, 2002; Wyss et al., 2006; Masquelier and Thorpe, 2007). Such an association across views is thought to be the underlying neural mechanism of object recognition (Wallis and Rolls, 1997; Wallis and Bülthoff, 1999; Riesenhuber and Poggio, 2000; Palmeri and Gauthier, 2004; Connor et al., 2007; DiCarlo et al., 2012).

Object information is represented and processed in the ventral cortical stream. The inferotemporal cortex is the final cortical area along this stream for pure visual information (Kravitz et al., 2013). Inferotemporal (IT) cells have stimulus selectivity to relatively complex object features compared to cells in the early visual areas of the ventral cortical stream (see Tanaka (1996) for review). Dynamic changes in the stimulus selectivity of inferotemporal cells accompany changes to the visual environment. Discrimination training across similar shapes increases the number of cells responding to the trained shapes (Logothetis et al., 1995; Kobatake et al., 1998). Cells selectively respond to the trained shapes (Baker et al., 2002) or to the stimulus dimension relevant to the discrimination (Sigala and Logothetis, 2002; De Baene et al., 2008). The positional consistency of neuronal stimulus selectivity in the inferotemporal cortex has been shown to be capable of deformation within a quarter of an hour by experiencing the successive appearance of

* Correspondence to: Department of Bioengineering, Graduate School of Science and Engineering, Kagoshima University, 1-21-40 Korimoto, Kagoshima 890-0065, Japan.
E-mail address: gwang@ibe.kagoshima-u.ac.jp (G. Wang).

https://doi.org/10.1016/j.ibneur.2021.02.008
Received 4 November 2020; Accepted 24 February 2021
Available online 25 February 2021

different stimuli at two different retinal positions (Li et al., 2009; Wallis and Bülthoff, 2001; Cox et al., 2005), whereas the temporal property of the response was first accessed using information value (Optican and Richmond, 1987). By computing the amount of information in small time segments, several studies have focused on the temporal property of information encoding. Reports on inferotemporal cells demonstrated that the global information is represented first, followed by local information for hierarchical stimuli (Sripati and Olson, 2009) and faces (Sugase et al., 1999). Information about multipart configurations is conveyed later than is information about a single part (Brincat and Connor, 2006). The stimulus selectivity in the initial phase of responses of IT cells to the object or face stimuli showed broader tuning than that in the late phase (Tamura and Tanaka, 2001).

View-invariant object recognition development involves associating representations of the same object at different viewing angles. Single-cell recordings from the inferotemporal cortex in monkeys showed a response tolerance within a range of viewing angles around the experienced viewing angle for each cell (Wang et al., 2005; Okamura et al., 2014; Zhao et al., 2018), helping understand the underlying neuronal mechanism of object recognition across viewing angles. The similarity in the response patterns at the cell population level differed significantly for views of the same object and different objects (Yamaguchi et al., 2016). Temporally, the difference between the neural distance for views of the same objects and that for views of different ones was initially small but became significantly different gradually from the viewing angle separation of 30°, then 60° and 90°. Our previous reports used the averaged spike rate in a several hundred milliseconds time period immediately after stimulus onset. However, the response of a cell usually significantly differed between the early and late response phases separated by 260 ms after stimulus onset. This study was designed to further understand the difference between the two response phases and its involvement in the view-invariant object recognition.

Methods

We re-analyzed the data from a series of our previous experiments (Okamura et al., 2014; Yamaguchi et al., 2016; Zhao et al., 2018; Okamura et al., 2018). We trained monkeys in months before electrode recordings to be familiar with the training object images for object discrimination tasks. Monkeys experienced the images of different object sets in different ways with different tasks. After saturation of the behavioral performance, the electrophysiological activities of single cells were recorded from the inferotemporal cortex.

Data from prior work

We pooled and re-analyzed the data previously obtained from the monkeys' inferotemporal cortex (Okamura et al., 2014; Yamaguchi et al., 2016; Zhao et al., 2018; Okamura et al., 2018). In total, 1032 individual cells responding to images with prior training experiences were pooled, including 223, 241, 198, 203, and 167 cells from five macaque monkeys, respectively. Details on the experimental procedure have been previously described (Okamura et al., 2014). Extracellular single-cell recordings with tungsten electrodes (FHC, Bowdoinham, ME, USA), were conducted after the training session for prior experiences of the object views. Recordings were performed around the ventrolateral region of the inferotemporal cortex, lateral to the anterior middle temporal sulcus, in the posterior/anterior range, between 16 mm and 26 mm anterior to the ear bar position in the monkeys.

To provide prior experience of the object images to the monkeys, we created object sets using three-dimensional graphics software, and trained the monkeys for several months before electrophysiological recording. Details of stimulus creation have been described previously (Wang et al., 2005). For each object set, we first designed a prototype object defined by seven parameters (e.g., length, angle). To avoid the possibility that the difference between a pair of objects was limited to only one or two features (parameters), the parameters were combined into three groups. By changing the parameters in different ways, four artificial objects were created. Four views of each object were created by rotating the object in 30° intervals around an axis perpendicular to the visual axis connecting the viewer's eyes and the object. One object set consisted of 16 images (4 views × 4 objects; an example set is shown in Fig. 1). A number of such sets were created from distinctly different prototypes for various prior experiences with the monkeys. The similarity was evaluated by the Euclidean distance between coefficients of wavelet image transformations of the images. The similarity between a pair of object images across sets was significantly smaller than that for any pair of objects in the same set. We used human psychophysics to make the difficulty of discrimination comparable among stimulus sets at 80% correct responses. The discrimination among images across object sets was perfect for all subjects already at the beginning; no any learning was required.

Before electrophysiological recording, we trained the monkeys for 2–3 months so that they were familiar with the training object images. During the training session, we exposed the monkeys to objects' images in two different ways: in an object task and an across-set image task (Okamura et al., 2014; Yamaguchi et al., 2016; Zhao et al., 2018; Okamura et al., 2018). As shown in Fig. 1A, in both tasks, the monkey could start a trial by pressing the lever at a time of his own choosing. After the appearance of the first stimulus image, two to five stimuli randomly chosen from the same object set were successively presented in each trial. To be rewarded, in each task, the monkey had to release the lever within 1 s when the object changed, but not when only the object view changed. There was a 33.3% chance of an object change on the second, third, and fourth presentations. In both of the tasks, the monkeys had to detect object change but ignore view change of the same object (Fig. 1B). The object task required an association to be formed across different views of each object. Different views of the same object were repeated randomly 1–4 times, and an image of another object in the same set appeared subsequently. In the across-set image task, after an identical image was repeated 1–4 times, an object image from a different set appeared. No discrimination was required between the images in the same set. In the Object task, the hit rate became larger than the false alarm with a difference between the two being greater than 0.5 usually within 10 days. In the Across-set Image task, the performance was close to perfect (with hit rate of ~1.0 and false alarm rate of ~0.0) from the beginning. Although the saturation of performance in the Across-set image task from beginning, the training was continued so that the total number of times each image was presented across object sets became equal. We trained each monkey to be familiar to the object images of different object sets with the two tasks. Multiple object sets were created. They were swapped across the tasks and monkeys. An object set was used only one time for either the object task or the across-set image task in each monkey.

Object discrimination performance evaluated by neuronal activity

All recordings were conducted while the monkey was performing the across-set image task. We analyzed neuronal responses to the first stimulus presentation in each trial. Only correct responses were included. The magnitude of the responses was determined as the mean firing rate during stimulus presentation minus the spontaneous firing rate immediately preceding the stimulus presentation. For each neuron, the significance of the response was tested using a Wilcoxon signed-rank test with Bonferroni correction. Data with $p < 0.05$ were considered statistically significant.

In the present study, a support vector machine (SVM) was used as the algorithm to create a classifier for object discrimination. We first generated a cell population response vector for each image by collecting the responses of individual cells to the image. Vectors for the 16 object images of the same object set were grouped. During electrophysiological recordings, one object image was repeatedly presented > 10 times. An

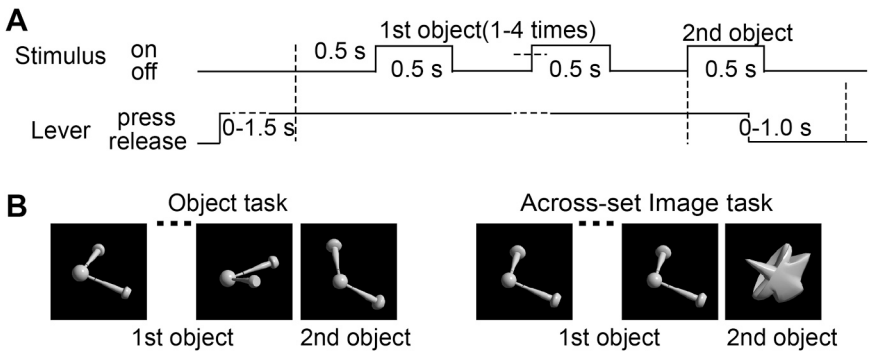

Fig. 1. Tasks used for prior experiences. (A) Time sequence of events for the tasks. The basic time sequence of events was consistent. In each trial, while the monkey pressed a lever and maintained fixation on a point, 2–5 object images were presented sequentially: one to four presentations of a first object were followed by one view of a second object. The monkey had to release the lever when the second object appeared. (B) Examples of the stimulus images presented in the object task and the across-set image task. While different views of the first object were presented in the object task, the view of the first object did not change in the across-set image task. Additionally, the second object was selected from the same set as the first object in the object task and from a different object set in the across-set image task.

object classifier was created for each object set by the SVM algorithm based on the 16 × 10 vectors. The responses to the 16 images were normalized, and Z-scores were calculated according to the following formula:

$$r_{normalization} = \frac{r - r_{mean}}{r_{std}}$$

The normalized response, $r_{normalization}$, was obtained by subtracting the mean spike rate, r_{mean}, from each individual spike rate, r, and then dividing by the standard deviation, r_{std}. Data were randomly divided between training and testing data with a ratio of 90–10%. Cross-validation was used to validate the model's accuracy. Among the 16 images of four objects, we labeled the images of an object different from

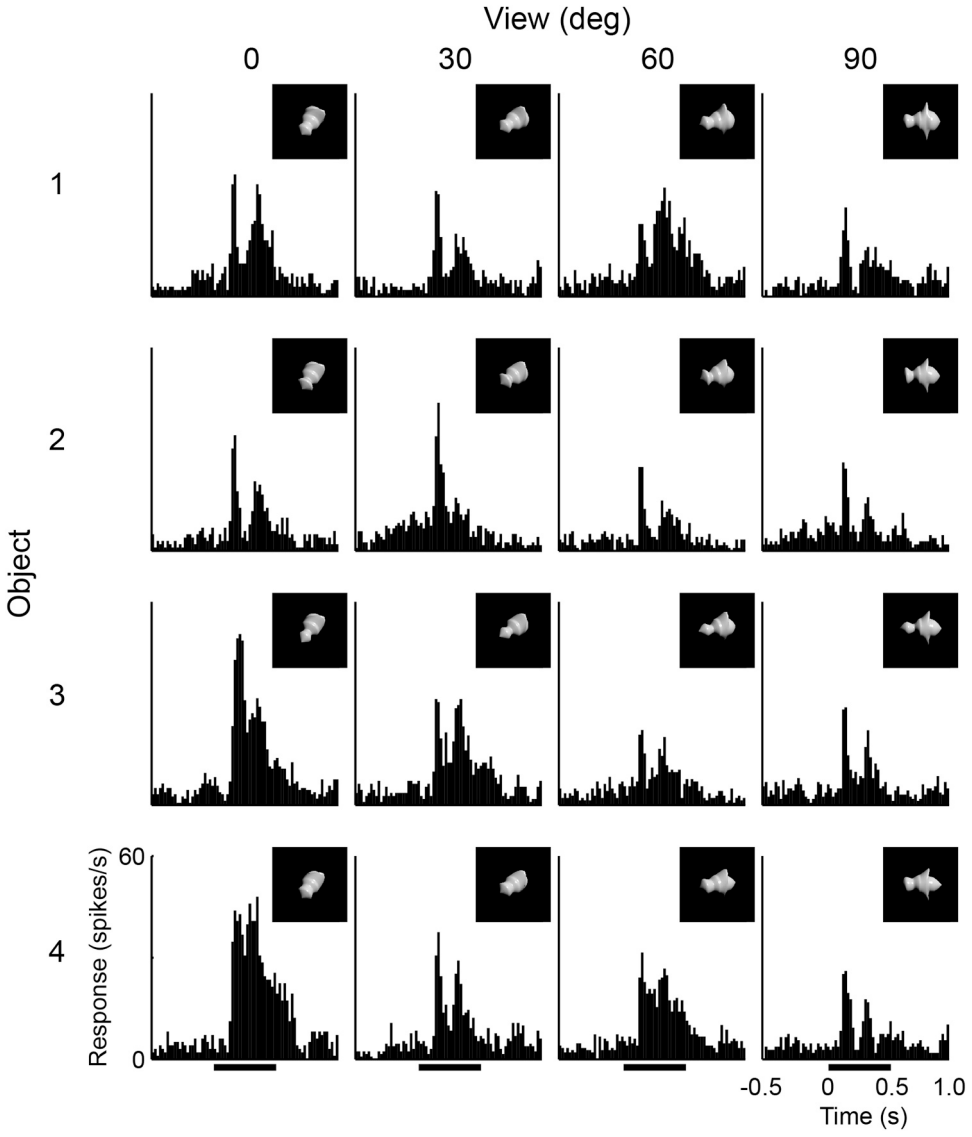

Fig. 2. Peri-stimulus time histograms (PSTHs) of a sample IT cell in response to the 16 images in an object set. Images on the upper-right corner of each plot represent stimulus images. The horizontal bar indicates stimulus presentation (0.5 s).

the other three objects, regardless of the difference in viewing angle. This was repeated four times for all four objects. The classifier's accuracy was evaluated by the percentage of the correct object identification.

Results

Data recorded from five monkeys were included in this study. In each monkey, there were two tasks for two different types of prior experience and two different object sets in each task. Depending on the prior experience from either the object or exposure tasks, cells recorded from each animal were divided into populations based on the object set. Cells responding to any of the 16 images in the same object set were pooled as a population. Therefore, there were 20 cell populations. The results below are based on these populations.

Single cell selectivity and its dynamics

Once we encountered a responsive cell, the responses of the cell to all the 16 images in object set were recorded. Due to the distinct difference in shape across images in different object sets, a cell usually responded to images in only one object set but not to others. During the electrophysiological recording session, the full set of images was always presented to a responsive cell to measure the selectivity among the 16 object-view images. Each cell usually responded differently to the 16 images in an object set. As in the example cell shown in Fig. 2, view 0 of object 4 evoked the largest response. The responses remained relatively large to some of the 16 images, but decreased significantly to other images. We previously discussed the stimulus selectivity change caused by object discrimination learning (Okamura et al., 2014). In addition to the response differences to the 16 images, we could also observe spike rate changes along the time axis. To investigate the time course of the change, we averaged the response histograms for all recorded cells (Fig. 3). In response to the presentation of a stimulus image, cells responded with a sharp increase in spike rate, peaking on average at 107 ms and then remaining relatively high until several tens of milliseconds after stimulus removal. The averaged response time course displayed two clear phases. The early response phase was in the time period of 100–260 ms, where spike rates showed a rapid increase and decrease, forming a sharp peak at 140 ms after stimulus onset. The late phase immediately followed the early phase from 260 to 660 ms. Changes in this phase were, by contrast, significantly smooth with a moderate peak. In the present study, we divided the response period into

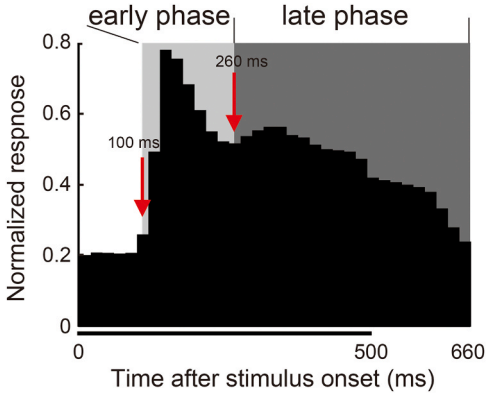

Fig. 3. Averaged peri-stimulus time histogram (PSTH) for normalized responses in all cells from one monkey. Rate at each bin was first subtracted by the averaged spike rate of a 400 ms window right before the stimulus onset, and then normalized by the maximum spike rate in the 16 images of an object set. Arrows indicate the turning points in the change of the normalized response. The early phase (light gray) and the late phase (dark gray) of response were defined as the time period of 100–260 ms and 260–660 ms, respectively. The horizontal bar indicates stimulus presentations (0.5 s).

an early phase of 100–260 ms and a late phase of 260–660 ms and separately investigated their response properties.

Stimulus selectivity and its change

In the following several hundred milliseconds after stimulus onset, a cell's stimulus selectivity changed among images of an object set. To better understand this property, we independently investigated the preferred stimulus image in each object set of the 16 images, in the early and late phases. Depending on the difference of preferred image in the late phase from that in the early phase, we classified cells into four types, based on whether the preferred images of the early and late phases had the same view point or belonged to the same object. Fig. 4 shows a sample cell for each of the four cell types. The type I cell responded optimally to a 30° view of object 4 in both early and late phases. The type II cell showed a maximal response to the 90° view of object 1 in the early phase, and in the late phase the image evoked the largest response was 60° view of object 1, a different view of the same object. The type III cell showed a preference to the 90° view of object 1 in the early phase, and then the preference shifted to the same 90° view of object 2. The preferred images of the type IV cell in the early and the late phases were different for both viewing angles and objects.

Type I cells demonstrated the same stimulus preference among the 16 images included in the object set in both the early and late phases, while type II cells preferred the same objects but in different views between early and late phases, and type III cells showed a preference for the same views but in different objects. The remaining cells were categorized as type IV, which showed the largest responses to different objects in different views between the early and late phases. We counted the number of cells for each of the four types. Fig. 5 displays the distribution of the four cell types separated by training task for prior experience. For the images experienced in the object task, type I cells constituted 34% of all the cells, and type II cells 32%, while the percentages for type III and type IV cells were 12% and 22%, respectively. For the across-set image task, the percentage of type I cells was 33%, comparable to the object task. Interestingly, type II cells constituted only 13%, which was significantly lower than that for the object task. The percentage of type III cells was 20%, significantly larger than that for the object task, and that of type IV cells was 34%. Comparing to the case with prior experience in across-set image task, the object task demonstrated a significantly different distribution in cell types ($p < 0.0001$, Chi-square test). A significant increase in the percentage of type II cells and, at the same time, a significant decrease in the percentage of type III cells were confirmed.

Evaluation of object discrimination by individual IT cell responses

Based on the responses of individual cells to the same image set, an object classifier was created using the support vector machine (SVM) algorithm. We trained the model for object identification regardless of the difference in view-point by using the averaged spike rates over a time window of 100–660 ms. One classifier was created based on the data obtained from each object set for each monkey. There were two types of prior experiences provided by the object task and the across-set image task. We always included two object sets in each condition. Data from five monkeys were used for analysis. For each data set, the object discrimination performance for the object set with prior experience of the object task was plotted against the performance for the set with prior experience of the across-set image task (Fig. 6, right). There was no significant difference among the performances among animals, and we could not confirm any significant difference in the performance between two object sets for the same condition, consistent with the results showing that the performance for the object set with prior experience of the object task was always better than that with prior experience of the across-set image task (Fig. 6, left). On average, the object discrimination performance for the object set with prior experience of the object task

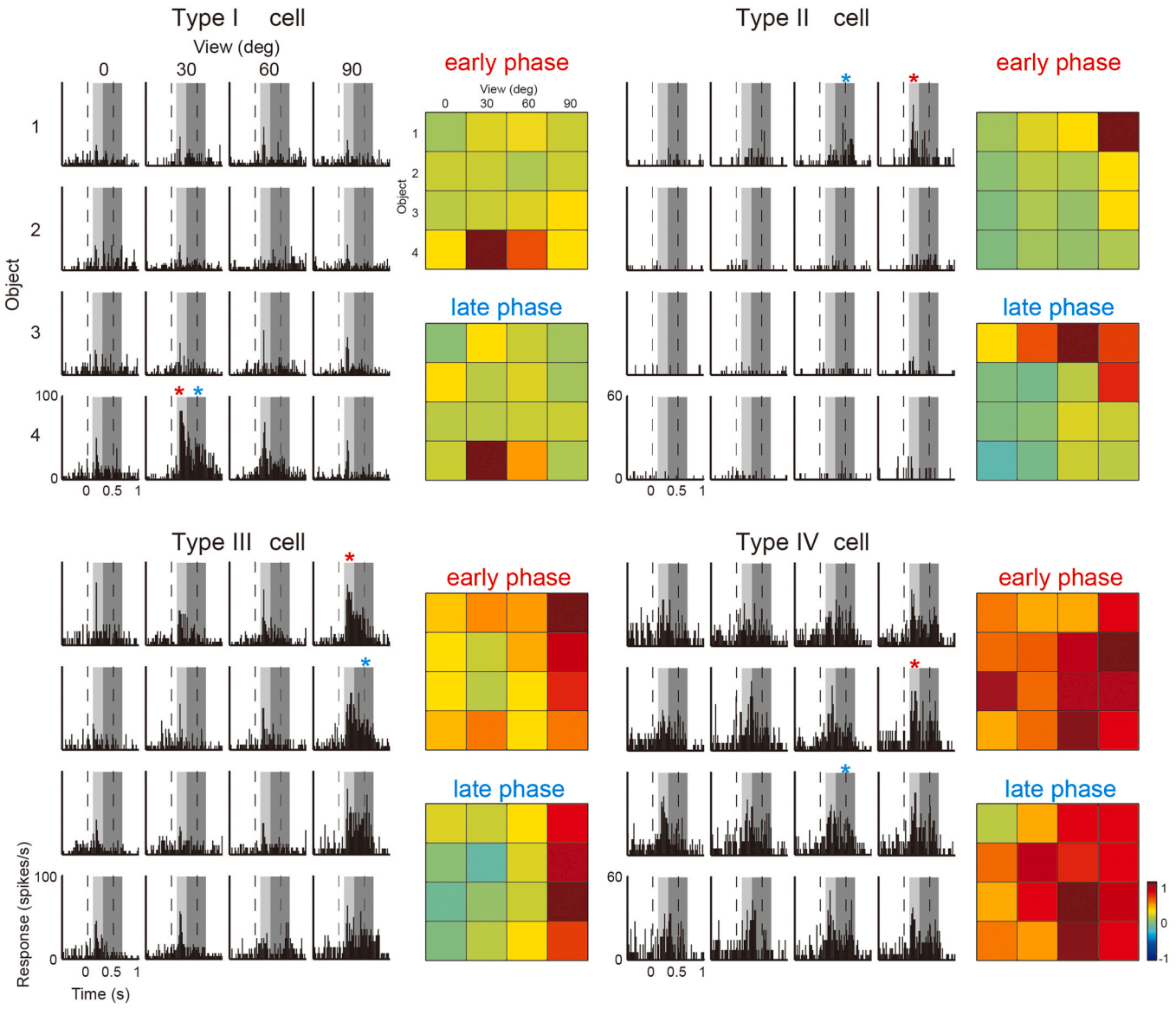

Fig. 4. Four cell types samples. PSTHs in response to 16 images in an object set as well as color codes of normalized spike rates in the early and late phases are plotted. In each plot, the 4 objects in the object set are aligned in different rows, the 4 view points are in different columns. Light grey: time period for the early phase. Dark grey: time period for the late phase. The red and blue asterisks denote the largest responses among the images in a set in the early and late phase respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

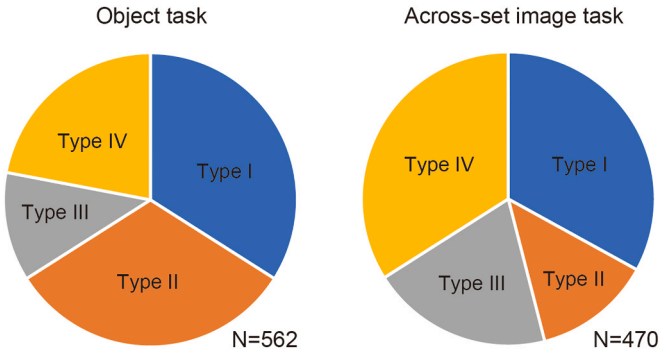

Fig. 5. Cell type distribution. Distribution for the object task and for the across-set image task were separately plotted.

was 80.9 ± 4.4%, significantly larger than that for the across-set image task 63.8 ± 6.0% ($t = 6.705$, $df = 9$, $p < 0.0001$).

Object discrimination in the early and late phases

Due to the distinct difference between responses in the early and late phases, we further trained object classifiers using the data averaged in the early phase and late phase separately. The classifier performance for the early phase was compared to the performance for the late phase. In Fig. 7, the performances for the prior experience of the object task were plotted against the performances for those with prior experience in the across-set image task. In the early phase of the response, the performance for responses to the images prior experienced in the object task was 61.6 ± 9.8%, while the performance for the responses to the images prior experienced in the across-set image task was 64.1 ± 7.8%, not being statistically different ($t = 0.708$, $df = 9$, $p = 0.497$). In the late phase, the performance for responses to the images prior experienced in the object task was as high as 79.6 ± 7.4%, significantly different from

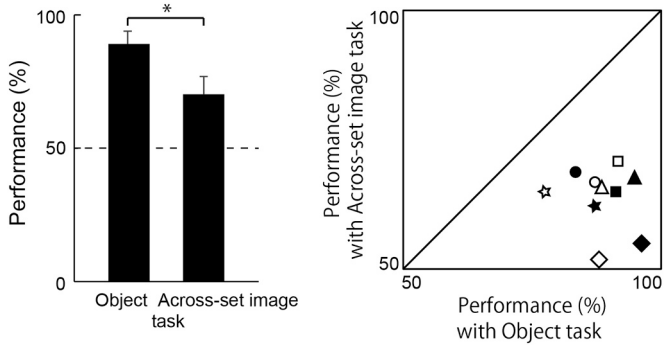

Fig. 6. View-invariant object discrimination performance averaged (left) and for individual animals (right). Computation was based on averaged spike rates over the whole time window of 100–660 ms, the early phase, and the late phase respectively. Each dot in the right panel represents the data from an object set. Two object sets for each monkey are marked in the same shape. Opened shapes represent data from one object set; filled shapes represent data from the other set. Error bars: standard deviation. *$p < 0.0001$.

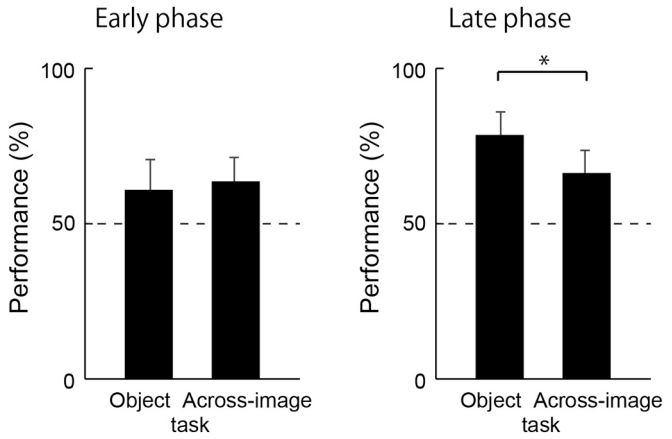

Fig. 7. View-invariant object discrimination performances in the early and late phases. Error bars: standard deviation. *$p < 0.005$.

that for the responses in the across-set image task ($66.5 \pm 7.4\%$, $t = 4.392$, $df = 9$, $p < 0.005$). In summary, in the early phase, the performances for the object task and the across-set image task were comparable, whereas the performance for the object task became significantly higher in the late phase of the response.

Discussion

In addition to the averaged spike rate during the whole response time period immediately after stimulus onset, we separated the response period based on the response histogram shape into two parts: an early response phase with a sharp peak immediately after stimulus onset and a late response phase, which appeared after the sharp peak in the early phase with a much flatter protrusion. In the present study, we demonstrated significant differences in the involvement of the view-invariant object recognition between the two response phases, based on the neuronal responses induced by prior training tasks. Responses to object images with prior experience of the object task, which required object recognition across viewing angles, were compared to the responses to object images prior experienced in the across-set image task, which worked as a passive exposure to the object images with no need of association across views. With the classifier trained by the respective responses, we found that in the early phase, the performance of the classifier created by the response data to the images experienced in the object task was comparable to that created for the across-set image task. By contrast, in the late phase, the performance for the object task was significantly higher than that for the across-set image task, thus, implying that in the late phase, the activity of inferotemporal cells may reflect the neural processing necessarily to achieve generalization across views of the same object. As in the early phase, considering previous findings on the selectivity of inferotemporal cells (Tanaka, 1996; Wang et al., 1996; Okamura et al., 2014), cell activity of cells may be more involved in the discrimination of the two-dimensional image shape. Comparisons between the signals present in different response time periods have been previously discussed (Sugase et al., 1999; Brincat and Connor, 2006; Tamura and Tanaka, 2001; Matsumoto et al., 2005; Brincat and Connor, 2004). Global categorical information such as monkey faces, human faces, or shapes was conveyed in the early time period, and fine information such as identity or facial expression, in the late time period (Sugase et al., 1999; Matsumoto et al., 2005). For shape, information about individual parts was provided in the early time period, and information about specific multipart configurations was given in the late time period (Brincat and Connor, 2004; Brincat and Connor, 2006). The current study provides a perspective to understand the underlying neuronal processing in view-invariant object recognition. By computing the amount of information in small time segments, several studies have been conducted on the temporal properties of information encoding. The information carried by single units was calculated every 50 ms, and global information was found to be conveyed faster than fine information (Sugase et al., 1999), as such information about multipart configuration is conveyed later than is information about a single part (Brincat and Connor, 2006). The information index has a negative correlation with the sharpness of stimulus selectivity. The stimulus selectivity in the initial phase of responses of IT cells to the object or face stimuli demonstrated broader tuning than that of the late phase (Tamura and Tanaka, 2001). We previously analyzed the temporal change in the correlation coefficient (r) between the population activities in response to two different stimulus images, and evaluated the neural distance by subtracting r from 1. We found that after discrimination of similar objects across viewing angles, the neural distance of IT cell populations between the same objects were significantly smaller than those between the different objects at a viewing angle differences of up to 90° (Yamaguchi et al., 2016). We propose here a new perspective by analyzing the dynamics of stimulus selectivity and the change of the optimal stimulus evoking the largest spike rates among stimulus sets, in different time periods.

By comparing the preferred object images in the early and late phases, we defined four types of cells. Training of association across object views did not significantly change the percentage of type I cells. Regardless of the training task for prior experience of stimulus images, approximately 30% of cells in IT did not change their stimulus selectivity during the response period immediately after stimulus onset. The approximately 70% cells that remained changed their stimulus images in the object set. Comparing with the across-set image task, the object task required the association across view images of the same objects but differentiation of the images of different objects. Such experience of the object task had significantly more of the remaining cells starting to respond to different views of the same objects in the early and late response phases. The prior experience of the association of the same object views in the object task may lead to the IT starting to respond to different views of the same objects, as pairing learning (Sakai and Miyashita, 1991). Such changes among different views of the same objects may work as an underlying neuronal basis for the behavioral association of the same object views. At the same time, the experience of the object task decreased the percentage of cells responding to the same view images of different objects. Such neuronal changes may be involved in object differentiation. The brain could increase the number of cells to respond to same object views so as to complete the computation to achieve view-invariance, and decrease the number of cells to respond to the same view of different objects to achieve differentiation across objects. Compared to the control, the object task led to more cells

changing from type IV to other types of cells. The decreased percentage of cells contributed to the increase in type II cells in the object task. A finding in line with those of Kobatake et al. (1998) who found that more cells started to respond to training stimuli.

Based on the single cell responses in the immediate period of 100–660 ms after stimulus onset to the object images previously experienced in the object task, the discrimination model created using the machine learning algorithm demonstrated significantly better performance than that using the responses with the across-set image task. This is reasonable because of object discrimination learning in the object task. Even by the use of spike rates, we were able to separate the objects in some extent if with prior experience in the object task (Okamura et al., 2014), whereas with the use of population activity, we could separate by the neural distance index (Yamaguchi et al., 2016; Zhao et al., 2018). One more important finding is the difference between the early and late phases during a period of 100–660 ms. The significantly higher performance for the object task in the period of 100–660 ms is mainly due to the activity of the late phase. In the early phase, we failed to find any significant difference between tasks. Instead, in the late phase, the performance for the object task was significantly better than the across-set image task. The early phase may mainly reflect the initial activity of the sensory response to the presentation of visual stimuli, as has been repeatedly demonstrated by previous studies. By contrast, the late phase of response may involve more in view-invariant computation. In addition to the view-invariance, we can also recognize object despite changes in their position and size. The development of such invariance was reported in a fixed temporal order. Size and position invariance developed first, followed by rotation and viewpoint invariance (Murty and Arun, 2017).

Ethics Statement

The present study did not start any new animal experiment, but re-analyzed the data from our previous animal experiments.

We, the authors, certify that all our previous animal studies were carried out in accordance with the National Institute of Health Guide for the Care and Use of Laboratory Animals (NIH Publications No. 80-23) revised 1996 or the UK Animals (Scientific Procedures) Act 1986 and associated guidelines, or the European Communities Council Directive of 24 November 1986 (86/609/EEC).

All the animal experiments were approved by the animal subjects review board of Kagoshima University, Japan. All efforts were made to minimize the number of animals used and their suffering.

CRediT authorship contribution statement

Ridey H. Wang: Methodology, Software, Formal analysis, Investigation, Writing - original draft. **Lulin Dai:** Analysis, Investigation. **Takayasu Fuchida:** Writing - original draft. **Jun-ya Okamura:** Investigation. **Gang Wang:** Conceptualization, Writing - review & editing, Supervision.

Conflict of interest

none

Acknowledgments

This research was partly supported by a Grant-in-Aid for Scientific Research on Priority Areas (20020021) from the Ministry of Education, Culture, Sports, Science and Technology of Japan to G.W.

References

Baker, C.I., Behrmann, M., Olson, C.R., 2002. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nature Neurosci. 5, 1210–1216.
Biederman, I., 1987. Recognition by components: a theory of human image understanding. Psychol. Rev. 94, 115–147.
Brincat, S.L., Connor, C.E., 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat. Neurosci. 7, 880–886.
Brincat, S.L., Connor, C.E., 2006. Dynamic shape synthesis in posterior inferotemporal cortex. Neuron 49, 17–24.
Bülthoff, H.H., Edelman, S., 1992. Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proc. Natl. Acad. Sci. USA 89, 60–64.
Connor, C.E., Brincat, S.L., Pasupathy, A., 2007. Transformation of shape information in the ventral pathway. Curr. Opin. Neurobiol. 17, 140–147.
Cox, D.D., Meier, P., Oertelt, N., DiCarlo, J.J., 2005. 'Breaking' position-invariant object recognition. Nature Neurosci. 8, 1145–1147.
De Baene, W., Ons, B., Wagemans, J., Vogels, R., 2008. Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. Learn. Mem. 15, 717–727.
DiCarlo, J.J., Zoccolan, D., Rust, N.C., 2012. How does the brain solve visual object recognition? Neuron 73, 415–434.
Foldiak, P., 1991. Learning invariance from transformation sequences. Neural Comput. 3, 194–200.
Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G., Mishkin, M., 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. Trends Cogn. Sci. 17 (1), 26–49.
Kobatake, E., Wang, G., Tanaka, K., 1998. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. J. Neurophysiol. 80, 324–330.
Li, N., Cox, D.D., Zoccolan, D., DiCarlo, J.J., 2009. What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? J. Neurophysiol. 102, 360–376.
Logothetis, N.K., Pauls, J., Bülthoff, H.H., Poggio, T., 1994. View-dependent object recognition by monkeys. Curr. Biol. 4, 401–414.
Logothetis, N.K., Pauls, J., Poggio, T., 1995. Shape representation in the inferior temporal cortex of monkeys. Curr. Biol. 5, 552–563.
Masquelier, T., Thorpe, S.J., 2007. Unsupervised learning of visual features through spike timing dependent plasticity. PLoS Comput. Biol. 3, e31.
Matsumoto, N., Okada, M., Sugase-Miyamoto, Y., Yamane, S., Kawano, K., 2005. Population dynamics of face-responsive neurons in the inferior temporal cortex. Cereb. Cortex 15, 1103–1112.
Murty, N.A.R., Arun, S.P., 2017. A balanced comparison of object invariances in monkey IT neurons. eNeuro 4 (2), e0333–16.
Okamura, J.Y., Yamaguchi, R., Honda, K., Wang, G., Tanaka, K., 2014. Neural substrates of view-invariant object recognition developed without experiencing rotations of the objects. J. Neurosci. 34, 15047–15059.
Okamura, J.Y., Uemura, K., Saruwatari, S., Wang, G., 2018. Difference in the generalization of response tolerance across views between the anterior and posterior part of the inferotemporal cortex. Eur. J. Neurosci. 48, 3552–3566.
Optican, L.M., Richmond, B.J., 1987. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. J. Neurophysiol. 57, 162–178.
Palmeri, T.J., Gauthier, I., 2004. Visual object understanding. Nature Rev. Neurosci. 5, 291–303.
Riesenhuber, M., Poggio, T., 2000. Models of object recognition. Nat. Neurosci. 3, 1199–1204 (Suppl).
Sakai, K., Miyashita, Y., 1991. Neural organization for the long-term memory of paired associates. Nature 354, 152–155.
Sigala, N., Logothetis, N.K., 2002. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415, 318–320.
Sripati, A.P., Olson, C.R., 2009. Representing the forest before the trees: a global advantage effect in monkey inferotemporal cortex. J. Neurosci. 29, 7788–7796.
Sugase, Y., Yamane, S., Ueno, S., Kawano, K., 1999. Global and fine information coded by single neurons in the temporal visual cortex. Nature 400, 869–873.
Tarr, M.J., 1995. Rotating objects to recognize them: a case study of the role of viewpoint dependency in the recognition of three-dimensional objects. Psychol. Bull. Rev. 2, 55–82.
Tanaka, K., 1996. Inferotemporal cortex and object vision. Annu. Rev. Neurosci. 19, 109–139.
Tamura, H., Tanaka, K., 2001. Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. Cereb. Cortex 11, 384–399.
Wallis, G., Rolls, E.T., 1997. Invariant face and object recognition in the visual system. Prog. Neurobiol. 51, 167–194.
Wallis, G., Bülthoff, H., 1999. Learning to recognize objects. Trends Cogn. Sci. 3, 22–31.
Wallis, G., Bülthoff, H., 2001. Effects of temporal association on recognition memory. Proc. Natl. Acad. Sci. USA 98, 4800–4804.
Wang, G., Tanaka, K., Tanifuji, M., 1996. Optical imaging of functional organization in the monkey inferotemporal cortex. Science 272, 1665–1668.
Wang, G., Obama, S., Yamashita, W., Sugihara, T., Tanaka, K., 2005. Prior experience of rotation is not required for recognizing objects seen from different angles. Nature Neurosci. 8, 1768–1775.
Wiskott, L., Sejnowski, T.J., 2002. Slow feature analysis: unsupervised learning of invariances. Neural Comput. 14, 715–770.
Wyss, R., König, P., Verschure, P.F., 2006. A model of the ventral visual system based on temporal stability and local memory. Plos. Biol. 4, e120.
Yamaguchi, R., Okamura, J.Y., Wang, G., 2016. Dynamics of population coding for object views following object discrimination training. Neuroscience 330, 109–120.
Zhao, C., Wang, R.H., Wang, G., 2018. Long-term object discrimination at several viewpoints develops neural substrates of view-invariant object recognition in inferotemporal cortex. Neuroscience 392, 190–202.