Representation of the Spatial Relationship Among Object Parts by Neurons in Macaque Inferotemporal Cortex

Yukako Yamane,[1,2] Kazushige Tsunoda,[1,3] Madoka Matsumoto,[1] Adam N. Phillips,[1] and Manabu Tanifuji[1]

[1]*Laboratory for Integrative Neural Systems, RIKEN Brain Science Institute, Saitama;* [2]*Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo;* [3]*Laboratory of Visual Physiology, National Institute of Sensory Organs, Tokyo, Japan*

Yamane, Yukako, Kazushige Tsunoda, Madoka Matsumoto, Adam N. Phillips, and Manabu Tanifuji. Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *J Neurophysiol* 96: 3147–3156, 2006. First published August 30, 2006; doi:10.1152/jn.01224.2005. We investigated object representation in area TE, the anterior part of monkey inferotemporal (IT) cortex, with a combination of optical and extracellular recordings in anesthetized monkeys. We found neurons that respond to visual stimuli composed of naturally distinguishable parts. These neurons were sensitive to a particular spatial arrangement of parts but less sensitive to differences in local features within individual parts. Thus these neurons were activated when arbitrary local features were arranged in a particular spatial configuration, suggesting that they may be responsible for representing the spatial configuration of object images. Previously it has been reported that many neurons in area TE respond to visual features less complex than natural objects, but it has remained unclear whether these features are related to local features of object images or to more global features. These results indicate that TE neurons represent not only local features but also global features such as the spatial relationship among object parts.

INTRODUCTION

Visual information about object images is conveyed from early visual area V1 to inferior temporal (IT) cortex through areas V2 and V4 in macaque monkeys (for review, see Logothetis and Sheinberg 1996). Area TE is a ventral part of IT cortex and is the site that represents object images necessary for visual recognition (Gross 1994; Logothetis and Sheinberg 1996).

Early studies on visual responses of TE neurons showed that these neurons respond to various visual stimuli including natural object images (Bruce et al. 1981; Desimone et al. 1984; Gross et al. 1979; Perrett et al. 1982; Schwartz et al. 1983). More recently, a number of studies have attempted to identify the simplest visual features that activate individual neurons in area TE (Kobatake and Tanaka 1994; Tanaka et al. 1991). These studies have revealed that essential stimuli for TE neurons are visual features that are geometrically less complex than natural objects. Thus combinations of visual features are necessary for neural representation unique to individual object images in area TE.

As in the primary visual cortex, neurons in area TE with similar response properties are reported to be clustered into columns (Fujita et al. 1992; Gochin et al. 1991). The columns responding to visual stimuli have been visualized with intrinsic signal imaging as darkened spots scattered across the cortical surface (Tsunoda et al. 2001; Wang et al. 1996, 1998). In particular, Tsunoda and colleagues used this technique together with conventional extracellular recordings and showed that an object image activates multiple spots, each of which represents a particular visual feature of the object image (Tsunoda et al. 2001). They reported that some of the visual features represented by activated spots were local features of object images, that is, features that appear in a spatially localized part of the object image. Thus it remains unknown how spatial arrangements of these local features in an object image are specified. Neurons in area TE may represent global features, such as spatial configuration of local features, in addition to spatially localized features. In this paper, we address this question by combining data from intrinsic signal imaging and extracellular recordings.

METHODS

Anesthesia and the general recording condition

Four rhesus monkeys were artificially ventilated with a mixture of N_2O, O_2, and isoflurane for anesthesia and paralyzed with pancuronium bromide or vecuronium bromide (Tsunoda et al. 2001). The visual stimuli were presented on a 20-in CRT display placed 57 cm from the eye contralateral to the recording hemisphere. The pupil of the eye was dilated by local application of 0.5% tropicamide 0.5% phenylephrine, and the cornea was covered with a contact lens of appropriate power to focus the visual stimuli onto the retina. The fovea was identified with a custom-made ophthalmoscope, and the position of the fovea was back-projected onto the center of the CRT screen. Except for three-dimensional (3D) objects for manual presentations, the visual stimuli were presented at the center of the CRT display. Electroencephalography (EEG), electrocardiography (ECG), expired CO_2 concentration, and rectal temperature were monitored throughout the experiments. The experimental protocol was approved by the Experimental Animal Committee of the RIKEN Institute. All experimental procedures were done in accordance with the guidelines of the RIKEN Institute and the National Institute of Health.

Intrinsic signal imaging

The dorsal part of area TE was exposed and illuminated by light with a wavelength of 605 nm through a glass cover slip window attached to a titanium chamber centered 15.0–17.5 mm anterior to the ear bar position (Tsunoda et al. 2001). Reflected light from the cortex was detected by a low-noise video camera (frame rate, 1/30 frames/s; S/N ratio, 60 dB; CS8310, Teli, Japan) and digitized by a 10-bit video capture board (Pulsar, Matrox). The light was focused to a depth of 500 μm below the cortical surface. The imaged area was 6.5×4.9
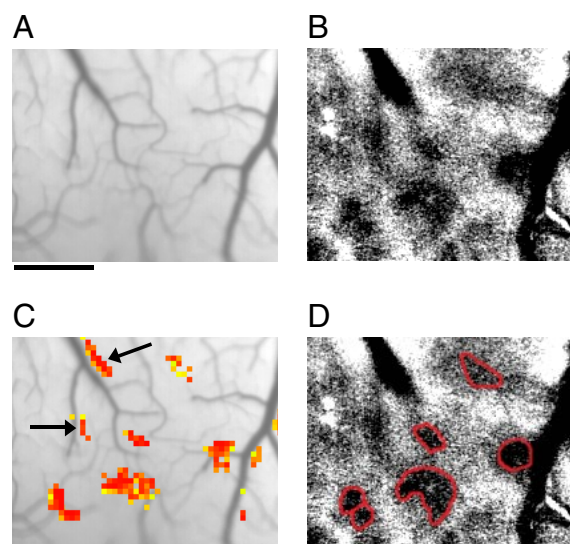
FIG. 1. Intrinsic signal imaging detected local modulation of light absorption changes in area TE. *A*: surface view of the exposed portion of the cortex. *B*: differential image showing local increase in light absorption. *C*: regions that showed statistically significant darkening after presentation of a visual stimulus are indicated in red ($P < 0.05$). Apparent artifacts that appeared along the thick vessels (arrows) were eliminated from the analysis. *D*: extracted active spots outlined by connecting pixels with half the peak absorption value. Scale bar = 1 mm.

mm and contained 320×240 pixels. We presented a visual stimulus to the monkey for 2.0 s, and sequential images were acquired for 4.0 s (starting from 1.0 s before the stimulus onset). During the 2-s stimulus presentation period, a stimulus image appeared and moved in a circular path (with a radius of 0.4° at the rate of 1 cycle/s). The imaging experiments consisted of two sessions. In the first session, the visual stimuli were 10–20 object images together with two blank images as control. Then on the basis on these results, we selected several stimuli that activated a large number of spots in the imaged region. In the second session, the selected stimuli ("the original"), their modifications and two controls were used as visual stimuli. Each stimulus was randomly presented 15–30 times in one session. The same imaging session as the second session was repeated at least twice on different days to confirm the consistency of the observed spots.

Identification of the active spots

The active spots were extracted as follows (Tsunoda et al. 2001): *1*) images acquired during the 0.5- to 3.0-s period after the onset of stimulus presentation were divided by an average of images during the 1-s period just before the stimulus onset. *2*) Gaussian spatial filtering was used to eliminate the global (stimulus nonspecific) darkening and high-frequency noise (cut-off frequencies: $\sigma = 0.04$ mm^{-1} for high cut and $\sigma = 2.1$ mm^{-1} for low cut). *3*) The *t*-values were calculated by pixel-by-pixel comparison of signal intensity between the filtered images for the trials with a particular stimulus and those for the control trials. The filtered images with a stimulus were averaged for all the trials and a differential image was created by subtracting the averaged image for control trials. Localized dark regions of the differential image, which showed significant darkening (*t*-test, $P < 0.05$), were defined as active spots. *4*) The contour of active spots was demarcated at the half-value of the peak absorption value. Representative images for each step are shown in Fig. 1.

Extracellular recording

The exposed cortex used for intrinsic signal imaging was covered with a transparent artificial dura made of silicon rubber (Arieli et al.

2002). Tungsten microelectrodes were inserted into the spots through the artificial dura. The surface blood vessel pattern was used as a mapping reference to identify the position of the spots. Extracellular action potentials were recorded for 3 s in each trial. Visual stimulus presentation started 1 s after the onset of a trial and lasted for 1 s. During the 1-s stimulus-presentation period, a stimulus image appeared and moved in a circular path (with a radius of 0.4° at the rate of 1 cycle/s). No intertrial interval was inserted, so that a blank period between two stimuli was 2 s. The different stimuli were presented in pseudo-random order, and the number of trials for each stimulus was between 10 and 20. For each stimulus, we applied the Wilcoxon test to the difference in the mean firing rate during and before the stimulus presentation. The amplitude of evoked responses for each stimulus was calculated by subtracting the mean firing rate during the 1-s period before the stimulus onset from the mean firing rate during the 1-s stimulus-presentation period, and by averaging for all the trials.

To characterize individual cells, we determined visual features critical for the cells according to previous studies (Fujita 1993; Fujita et al. 1992; Tanaka et al. 1991) (Fig. 2): *1*) we manually searched for the most effective visual stimulus among 96 hand-held 3D objects (Fig. 3), *2*) we simplified the best stimulus by removing or modifying a particular visual feature of the stimulus, and *3*) if the simplified image elicited significant responses (Wilcoxon test, $P < 0.05$) and also if the response amplitude for the simplified image exceeded a certain threshold, we used this image as the best stimulus in the next step. This procedure was repeated until further simplification failed to produce any response that exceeded the threshold. The threshold was set to 70% of the response elicited by the stimulus before simplification because there was no significant difference in evoked responses at this threshold. Typically, we started with the examination of a
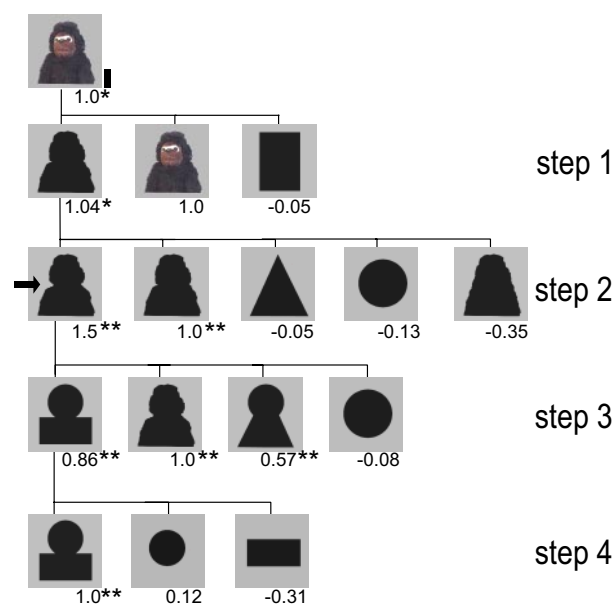

FIG. 2. Systematic simplification of an object image. The stimulus was simplified step by step. The stimulus that evoked the strongest response in one step was examined in the next step as the reference stimulus. The numbers below each picture indicate the response amplitudes normalized to the response to the reference stimulus together with statistical significance (Wilcoxon test, *$P < 0.05$, **$P < 0.01$). Step 1 showed that neuronal activities elicited by the best object and the silhouette were the same. Step 2 examined the effect of the 'sharpness' of the corner at the junction of upper and lower parts (arrow) and showed that the silhouette with the sharpest corners (leftmost picture) was the most effective stimulus. Step 3 showed that activities elicited by the leftmost two stimuli were not significantly different. Step 4 showed that neither the upper nor lower part activated the cell. In this case, the critical feature was determined as a combination of a circle and a rectangle (leftmost picture at step 4). Scale bar, 5°.
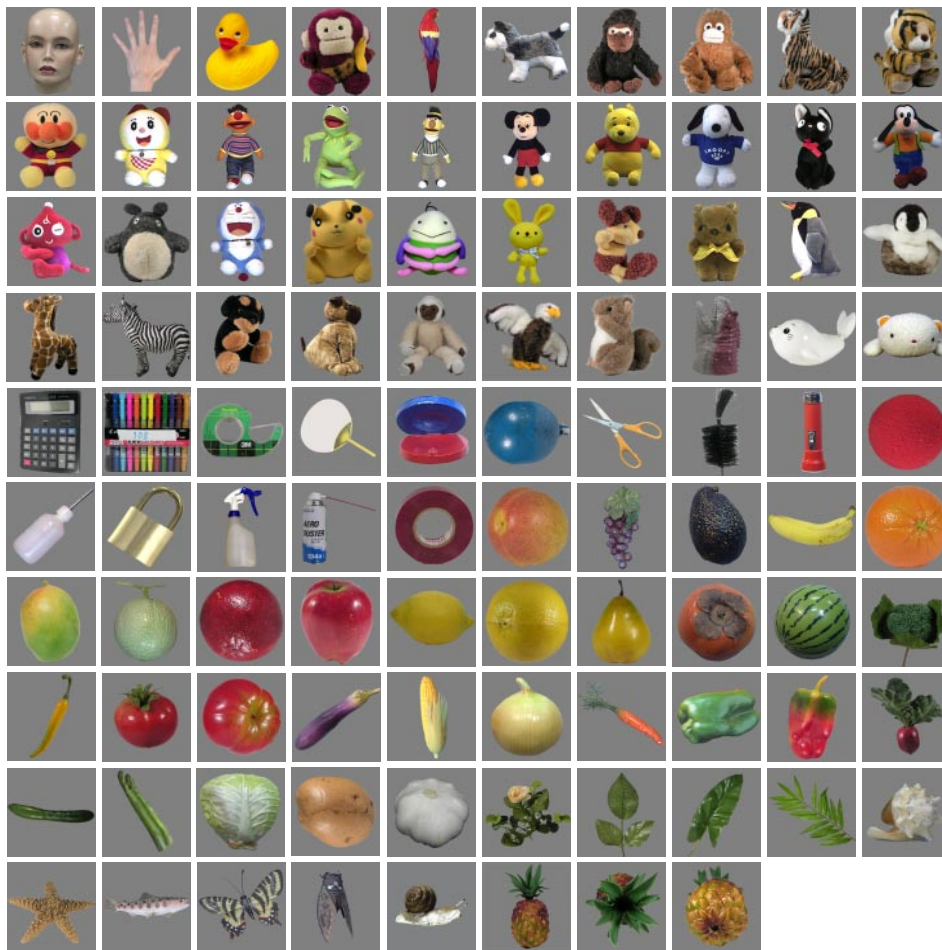
FIG. 3. The 3-dimensional (3D) objects used as visual stimuli to search for effective objects for individual cells. These objects were presented to the animals from various perspectives, as shown in the last three pineapple images in the *bottom row*.

monochrome image and silhouette of the original as in Fig. 2. However, image simplifications in the intermediate levels were different from case to case even if the original object was the same. The average numbers of simplification steps before reaching the simplest visual feature was 4.8 ± 2.0 (mean $\pm$ SD).

We examined receptive field size of each cell manually with the most effective 3D objects. We presented these objects not through the CRT display but directly to the animals because the size of receptive fields was larger than the size of the CRT display in most of the cells. On average, edges of receptive fields were $18.82 \pm 8.17°$ above, $23.14 \pm 5.30°$ below, $23.86 \pm 8.41°$ contralateral, and $22.60 \pm 8.62°$ ipsilateral to the recording hemisphere (means $\pm$ SD, $n = 35$). These values were measured from the center of the fovea. These values represent distances between the center of the CRT screen and the center of objects at the edges of receptive fields. Two exceptional cell having smaller receptive field size was eliminated from the analysis. The size of the object images used for quantitative analyses was on average $13.30 \pm 3.08°$ along the vertical axis and was $10.88 \pm 4.02°$ along the horizontal axis.

Definition of object parts

Throughout this manuscript, we took the simplest definition of object parts as the ones naturally distinguishable by discontinuities at minima of negative curvature of the object shape. For example, the minimum of negative curvature of stimulus 1 in Fig. 4*A* is the joint connecting head and body, and accordingly the stimulus is segmented into "head" (stimulus 3) and "body" (stimulus 2). Although this definition of parts is conventionally used in the field of object vision, there is no a priori reason to define object parts according to this

definition for TE neurons. Our intention, however, was not to explore the optimal segmentation for TE neurons, but rather to search for a possible mechanism for representing the spatial relationship among local features.

RESULTS

Intrinsic signal imaging

First, using intrinsic signal imaging, we identified one or two visual stimuli that activated a large number of spots in an imaged region of area TE. Each of these "original" visual stimuli was segmented into parts containing local features. We then conducted another intrinsic signal imaging session with a stimulus set consisting of the original (stimulus 1), two individual parts of the original (stimuli 2 and 3), and the original with a gap between the two parts (stimulus 4; Fig. 4). The results revealed that each spot was activated differently by these four stimuli. For example, one spot (indicated by an arrowhead in Fig. 4*A*) was activated by stimuli 1, 3, and 4 but not by stimulus 2. We interpret activity in this spot as being related to local features in the upper part of the original image (Fig. 4*A*). Because we were interested in identifying spots related to the spatial arrangement of local features in parts, we restricted our analysis to spots that were activated by combinations of the two parts but not by individual parts. Among four examined hemispheres, we found three spots, A–C, that satisfied this criterion: these spots were activated by the original image (stimulus 1) and the original with a gap (stimulus 4)
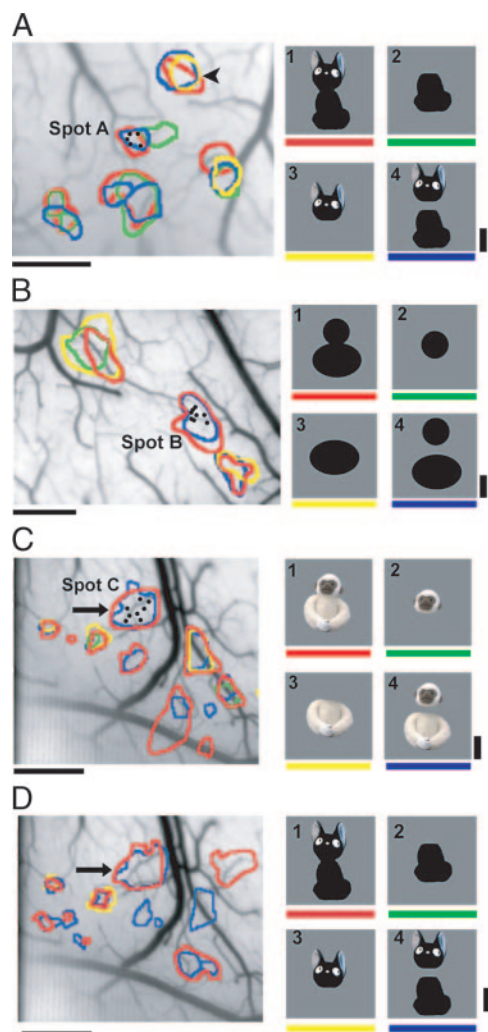
FIG. 4. Activation patterns of intrinsic signals evoked by sets of visual stimuli. *A–C*: activation patterns obtained from 3 different hemispheres, in which spots A, B, and C [spatial relationship relevant spots (SRR spots)] had specific response selectivity to sets of visual stimuli. *D*: activation pattern obtained from the same area of cortex shown in *C* but with a different set of stimuli. In each panel, activation patterns evoked by a set of the stimuli 1–4 (right) are indicated by colored contours superimposed on the surface image of the cortex. The bars below each stimulus and the outlines of spots activated by the stimulus match in color. The arrowhead in *A* indicates the spot related to local features in stimulus 3. The arrow in *D* indicates the region corresponding to spot C. The dots in the spots A–C indicate electrode penetration sites for subsequent extracellular recordings. Horizontal scale bar, 1 mm. Vertical scale bar, 5°.

but not by either part alone (stimuli 2 and 3; Fig. 4). Because the original with a gap (stimulus 4) activated these spots, activity in these spots could not be caused by specific responses to particular local features at the junction between the parts, such as a sharp negative curvature. We therefore considered these spots to be spatial relationship relevant spots (SRR spots), where we would likely find neurons representing the spatial relationship between two parts or between features within these parts.

Responses of the cells in SRR spots to spatial arrangements of object parts

We then conducted extracellular recordings from 49 cells located within SRR spots to characterize responsiveness of individual cells (13, 14, and 22 cells in spots A–C, respectively). First, we examined visual responses of each cell with 96 real object stimuli including faces, hands, imitations of living animals, stuffed animals, tools, and plastic fruits and vegetables (Fig. 3). These objects were presented in various sizes, orientations, and views so that the actual number of two-dimensional images used as visual stimuli was three or four times larger than the number of real objects. The stimuli that elicited significant responses (Wilcoxon test, $P < 0.05$) were diverse in color, texture, and local shapes (Fig. 5, *A, C,* and *E*). We could not explain this visual diversity in effective stimuli by preferred stimuli being different from cell to cell in a spot because individual cells in these spots responded to different stimuli and the response amplitudes did not significantly differ from each other (1-way ANOVA, $P > 0.25$; Fig. 5, *B, D,* and *F*). One common aspect of these effective visual stimuli was that the objects tend to consist of at least two distinguishable parts (Fig. 5, *B, D,* and *F*). These results from extracellular recordings were in accordance with the observation that with optical imaging, stimulus selectivity of a spot was the same for stimulus sets that originated from different object images (Fig. 4, *C* and *D*).

To address the question of whether cells in these spots could represent spatial relationships between object parts, we generated a set of visual stimuli for each cell where the upper part of the best object stimulus was rotated by various angles relative to the lower part of it. We then examined selectivity of each cell to this set of visual stimuli (Fig. 6). We found that cells were selectively activated when the parts were aligned vertically (Fig. 6, *A* and *B*). The selectivity could not be simply explained by changes in retinotopic positions of the upper parts that occurred incidentally during the spatial rearrangements of parts because of the large receptive field sizes of cells in area TE (Gross et al. 1969; Ito et al. 1995). In fact, evoked responses to the best object stimulus at different positions in space did not significantly differ from each other (Fig. 6*A,* symbols on the vertical axis). Of 30 cells examined for spatial arrangements of object parts, the responses of 20 cells (67%) significantly depended on the spatial configurations of two parts (1-way ANOVA, $P < 0.05$; Fig. 6, *C–F*). Thirteen cells had a single peak at 0 or 45° (Fig. 6, *C* and *D*), 4 cells had a single peak at other positions (Fig. 6*E*), and the remaining 3 cells had two peaks (Fig. 6*F*). Receptive field sizes of these cells were larger than the range of the shift of upper parts in retinotopic position incidental to the spatial rearrangement of parts (see METHODS). These results suggest that the cells in SRR spots are sensitive to particular spatial arrangements of objects' parts.

Response properties of the cells in SRR spots for the simplest visual features that activated the cells

The sensitivity of the cells to a particular spatial arrangement of parts could be due to the changes in local shapes of either part that occurred incidentally during spatial rearrangements of parts (for example, see Fig. 6, *A* and *B*). We conjectured that this was not the case because SRR spots were less sensitive to variations in local shapes (Fig. 5). To confirm this point, however, we determined the simplest visual feature that produced maximal activation ("critical feature"), and examined the cell's sensitivity to modifications of the critical feature for
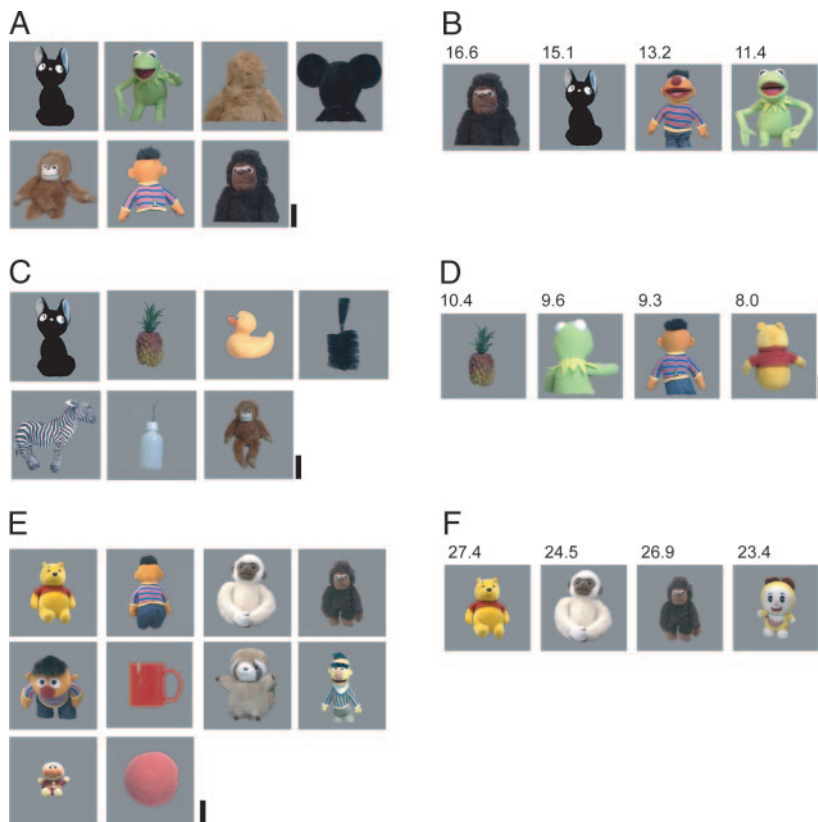
FIG. 5. Representative object stimuli that elicited significant responses of cells in SRR spots. *A, C,* and *E*: each object image represents the stimulus in a specific orientation that elicited the strongest significant responses out of the 96 objects in various orientations and views tested (*t*-test, $P < 0.05$) for a cell recorded in spots A (*A*), B (*C*), and C (*E*). *B, D,* and *F*: best 4 stimuli of 96 objects that elicited significant responses for a representative cell in spots A (*B*), B (*D*), and C (*F*). One-way ANOVA confirms no significant difference in responses to these 4 stimuli in both cases ($P > 0.25$). Evoked responses (spikes/s) are indicated above each stimulus image. Scale bar, 5°.

each cell in spots A and B. We systematically simplified the best object stimulus step by step to find critical features for 27 cells, following procedures from previous studies (Fujita 1993; Fujita et al. 1992; Kobatake and Tanaka 1994; Tanaka et al. 1991) (Fig. 2). Figure 7*A* shows the responses of a representative cell in spot A to its critical feature and to modifications of the critical feature. The critical feature was a combination of a circle and a rectangle (Fig. 7*A,* stimulus 1); the presentation of the upper or lower part alone caused significant decrease in the evoked responses (*t*-test, $P < 0.05$; Fig. 7*A,* stimuli 2 and 3). The cell responded equally well to the original colored object image and a silhouette of the original, indicating that color and texture of the stimulus were not essential (for example, see Fig. 2). The cell was not sensitive to changes in the shape of individual parts as long as the combination was preserved (Fig. 7*A,* stimuli 4 and 5). The existence of two parts was required, but the cell was not sensitive to local features at the junction between the two parts. For example, evoked responses to the stimulus with a gap (stimulus 7) and to the original (stimulus 1) were not significantly different (Fig. 7*A*). If, however, the "two parts" distinction was made less evident by smoothing the sharp joints (Fig. 7*A,* stimulus 8), the response was significantly reduced. Thus stimulus 6 but not stimulus 5 caused significant decrease in the evoked responses (Fig. 7*A*). The representative cell in spot B shows results consistent with the cell in spot A (Fig. 7*C*). In addition, we found that sharp joints between the two parts presented in isolation (Fig. 7*C,* stimulus 10) significantly reduced the responses, indicating that a sharp joint by itself was not sufficient.

Of 27 examined cells, the stimulus simplification revealed that 25 cells (93%) were not selective for color, luminance, or texture (exceptions are given in Fig. 7*B,* stimuli 5 and 6), and the critical features of 22 cells (81%) consisted of two parts (Fig. 7, *B* and *D;* exceptions are stimulus 8 in Fig. 7*B,* and stimuli 7 and 8 in Fig. 7*D*). The sensitivity of these neurons to modifications of critical features are summarized in the scatter plots, where diagonal lines indicate that evoked responses to the critical features and those to the modifications were the same (Fig. 8). Evoked responses to isolated parts were typically smaller than those to the critical features except for in a few cells (Fig. 8, *A* and *B*). The cells did not respond differently after changes in shape of either part or by inserting a gap between the two parts (Fig. 8, *C* and *D*). In 37% of the cells (filled symbols; 6 of 16 cells), significant reduction was observed when the stimuli had smoothed edges between the parts, but there was no significant reduction for the rest of the cells (open symbols; 10 of 16 cells) probably because two parts of the critical features were still distinguishable even when the joint was smoothed (Fig. 8*E*). Elongated shapes and a sharp joint between the two parts presented in isolation caused significant reductions in responses (Fig. 8, *F* and *G*). Thus sensitivity to these modifications of the critical features obtained from a population of neurons in these spots generally agreed with that obtained from representative cells.

In addition to the preceding results, as expected from the sensitivity of these cells to spatial arrangement of the object parts (Fig. 6), we found these cells were also sensitive to modifications of critical features in spatial arrangement of parts (Fig. 9). It should be noted that among the cells that were selective to a particular spatial arrangement of parts, 73% did
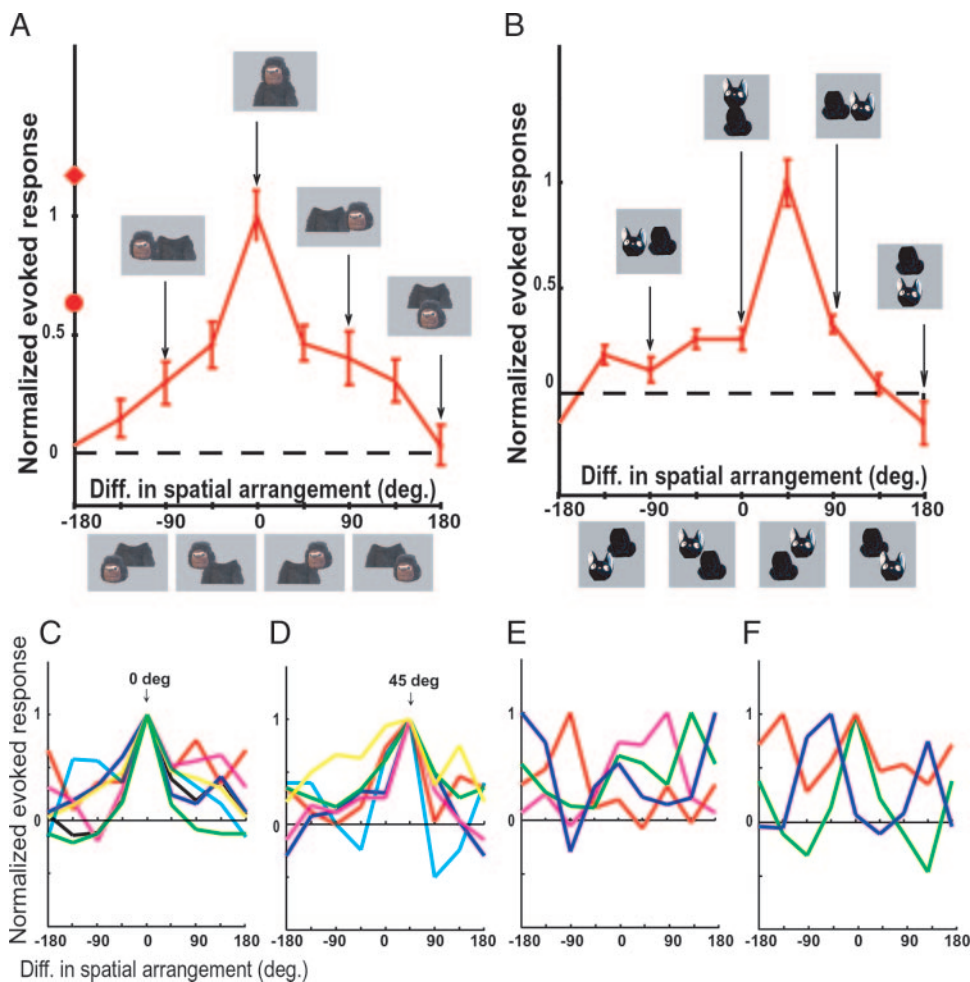
FIG. 6. Selectivity of cells in SRR spots to different spatial arrangements of the upper and the lower parts of object images. The normalized evoked responses (vertical axis) were plotted against the difference in the spatial arrangement of the parts (horizontal axis). The difference in spatial arrangement is defined by the angle between a line connecting the centers of the 2 parts of the best object stimuli and that of each rearranged stimulus. The pictures of stimuli corresponding to each angle are shown below the plot and also in the insets. A and B: responses of representative cells in spots A and B, respectively. Error bars indicate SE ($n = 20$). One-way ANOVA confirmed that the evoked responses significantly differed depending on the spatial arrangement of the parts ($P < 0.05$). Normalized responses to the original stimuli presented at different retinotopic positions are indicated by symbols along the vertical axis in A. The distance between the leftmost (●) and rightmost (◆) stimuli was 12.8°. C–F: tuning curves for other cells in spots A–C with a single peak at 0° (C), 45° (D), and other angles (E), and tuning curves with multiple peaks (F). For simplicity, only the mean values of responses are plotted.
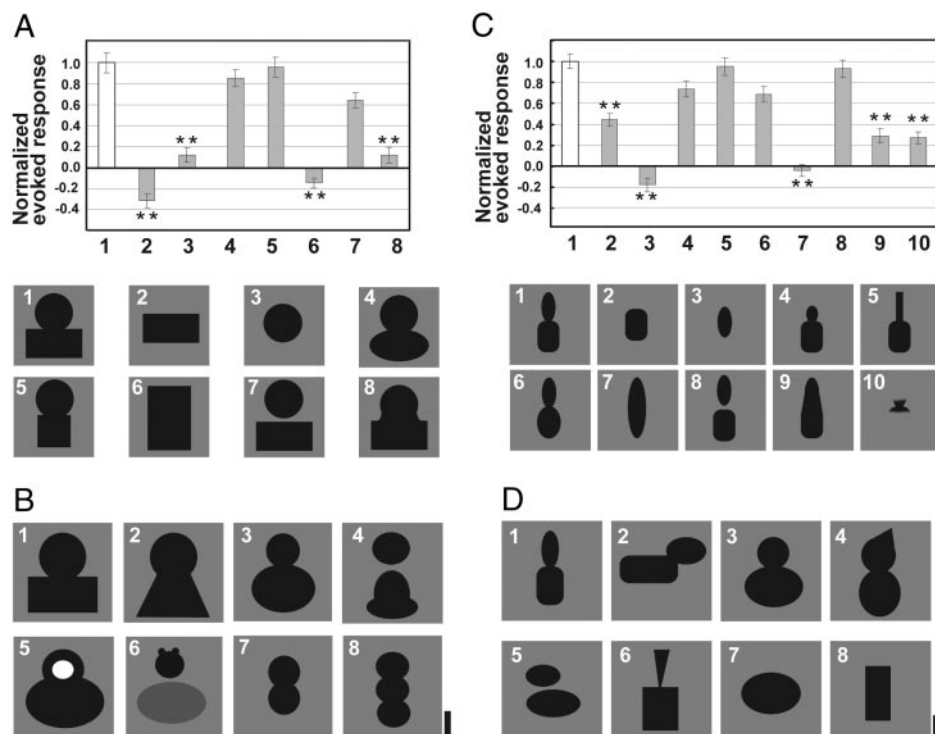

FIG. 7. Visual features critical for the cells in SRR spots. A and C: single-cell responses in spots A (A) and B (C). Evoked responses normalized to those by the critical features (stimulus 1) were plotted against the stimulus numbers. The pictures of stimuli with the stimulus number are shown below. Asterisks indicate a significant decrease in evoked responses compared with responses to the critical features (stimulus 1) (t-test, $P < 0.01$). Error bars indicate SE ($n = 20$). B and D: representative critical features for different cells in spots A (B) and B (D). The stimulus was filled with black if color, luminance, and texture were not essential for activation. Only the cells with critical features 5 and 6 (B) were sensitive to luminance. Scale bar, 5°.
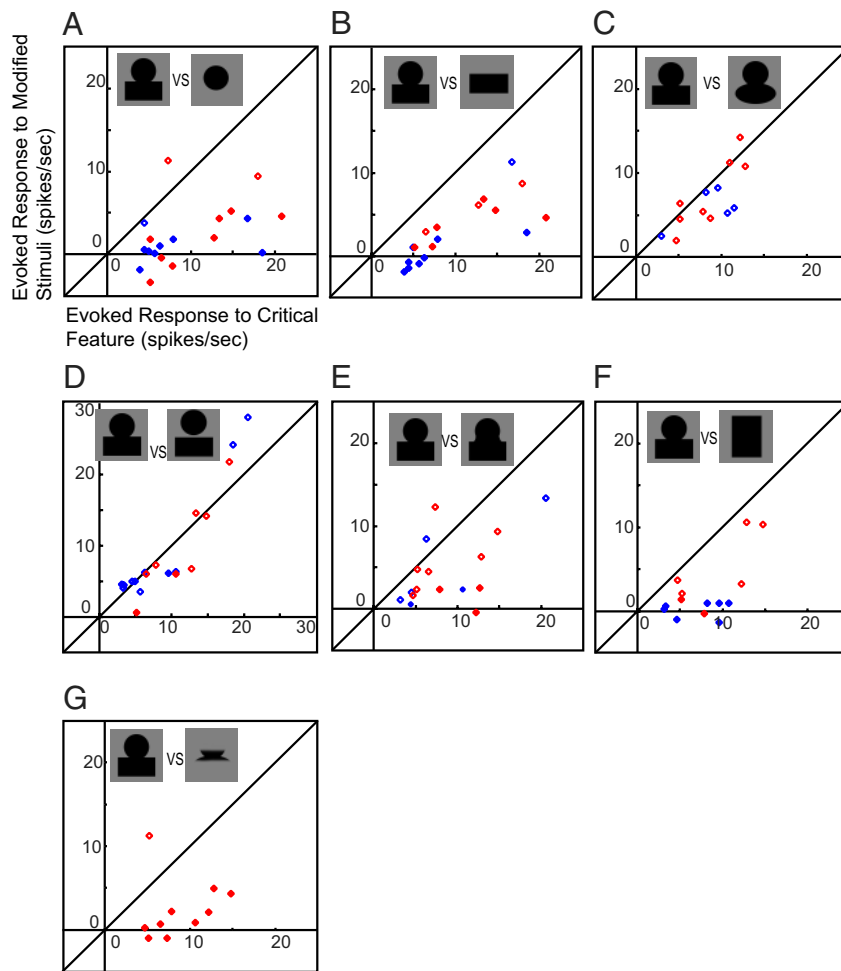
FIG. 8. Comparison between responses to the critical features and those to the modifications of the critical features. The modifications were isolated parts (A, *top,* and B, *bottom*), changes in shape of parts (*C*), inserting a gap between parts (*D*), smoothing the sharp joint (*E*), elongated shape (*F*), and isolated sharp joint (*G*). These modifications are signified by icons at the top of each graph, but these icons were not exactly the ones used in the experiment. The actual stimuli were created by modifying the critical features of individual cells. Each point represents the evoked response of single cell to its critical feature (horizontal axis) and that to modifications (vertical axis). Points below (or above) the diagonal line indicate that the response to the modification was smaller (or larger) than the response to the original critical feature. Filled symbols represent the cells that showed a significant difference in response to critical features and the modifications. These single cellular responses were obtained from spots A (red symbols) and B (blue symbols).

not respond to the combination of parts when the upper part was rotated 180° (Figs. 6 and 9). This means that, although neurons in these spots are less sensitive to local features, they still possess the capability to distinguish upper parts from lower parts by certain features existing within the parts. Because the upper and lower parts of each critical feature were different in size for most of these cells (Fig. 7, *B* and *D*), relative size may be one determining factor.

Taken together, at the level of the simplest visual features that maximally activate individual cells, we found that cells in SRR spots required two distinguishable parts for activation. For maximal activation of these cells, these two parts had to be arranged in particular spatial configurations, but it was not necessary to have particular local features embedded in parts.

DISCUSSION

It has been shown that object images are represented by combinations of neurons representing component visual features that are less complex than object images in IT cortex (Desimone et al. 1984; Fujita 1993; Fujita et al. 1992; Ito et al. 1995; Kobatake and Tanaka 1994; Tanaka et al. 1991; Tsunoda et al. 2001; Wang et al. 1996, 1998). As for the spot indicated by the arrowhead in Fig. 4*A,* some of these neurons represent local features of object images (see also Tsunoda et al. 2001). Thus for complete reconstruction of an object image from these features, neural mechanisms that specify spatial configuration

of these local features are needed. A general framework of object representation is the structural description where object images are composed of parts and the spatial relation among parts (Biederman 1987; Marr and Nishihara 1978). Although parts in this general framework and visual features represented in IT cortex are not necessarily the same, the representation of spatial configuration of elementary components is the common central issue.

Recently, Brincat and Connor examined visual responses of IT neurons with variations of two-dimensional (2D) silhouettes consisting of multiple curvatures and found that optimal features of these neurons were a combination of specific local curvatures arranged in particular positions in space (Brincat and Connor 2004). In this study, they examined the cells with 2D silhouettes but not with object images. Because representation of spatial configuration requires the cell to be insensitive to the visual attributes specific to particular parts, their results were not conclusive with respect to representation of spatial configuration of object parts.

From our findings in the present study, we suggest that neurons in area TE could represent a particular spatial arrangement of object parts based on two observations: *1*) with intrinsic signal imaging, we found activity spots that responded to a combination of two parts but not to either part shown in isolation (Fig. 4) and *2*) neurons recorded in these SRR spots were selectively activated by stimuli in which the parts were
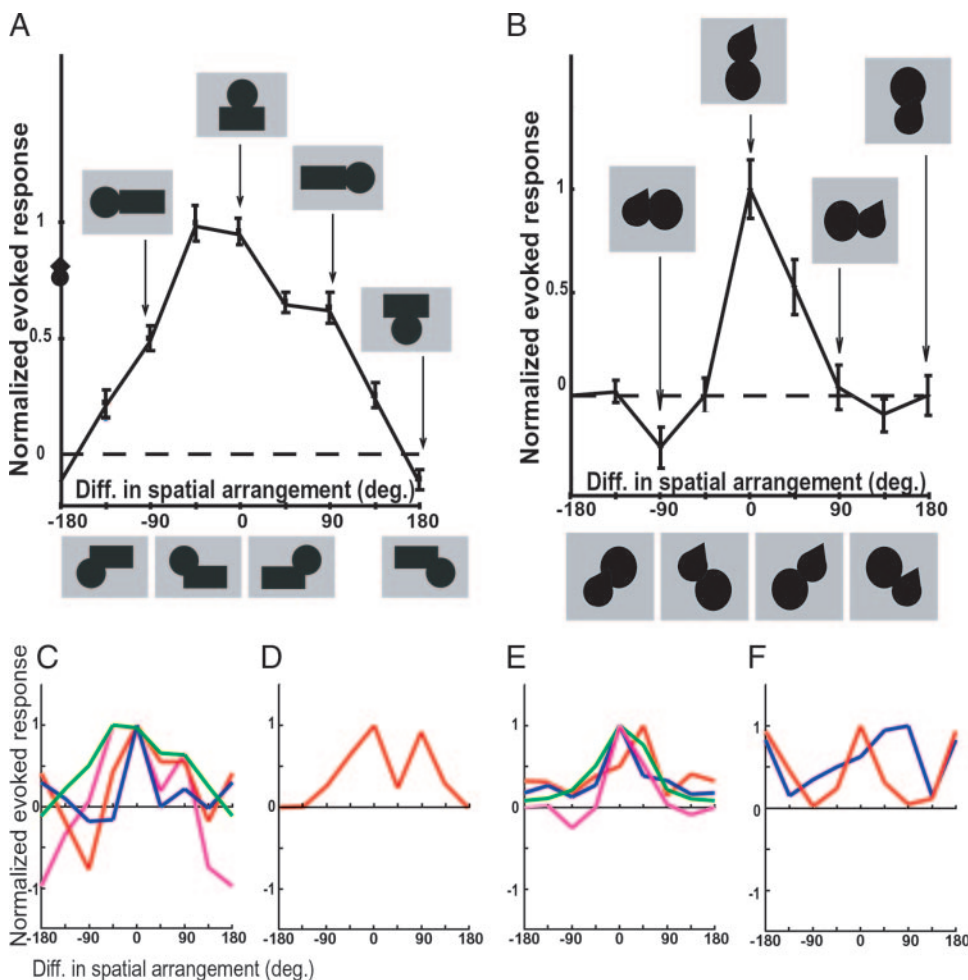
FIG. 9. Selectivity of cells in SRR spots to different spatial arrangements of the upper and the lower parts of their critical features. Normalized responses of the representative cells in Fig. 6, *A* and *B,* are represented in *A* and *B,* respectively. The stimuli were generated from the critical feature of each cell instead of the best object stimulus. Other conventions are the same as those in Fig. 6. One-way ANOVA confirmed that the evoked responses significantly differed depending on the spatial arrangement of the parts ($P < 0.05$). Normalized responses to the critical feature presented at different retinotopic positions are indicated by symbols along the vertical axis in *A.* The distance between the leftmost (●) and rightmost (◆) stimuli was 12.8° (*C–F*). Tuning curves for other cells with single (*C* and *E*) and multiple peaks (*D* and *F*) in spots A (*C* and *D*) and B (*E* and *F*).

arranged in specific spatial relationship (Figs. 6 and 9). In addition, our data show that cells in these spots are less sensitive to changes in visual attributes that are essential to characterize local features, such as color, texture, and local shape: *1*) activity spots showed the same response selectivity for stimulus sets derived from different object images (Fig. 4, *C* and *D*), *2*) neurons in these spots responded equally well to the stimuli including different colors, textures, and local shapes (Fig. 5), and *3*) the critical features of these neurons did not include particular local features (Fig. 7, *B* and *D;* see also Fig. 2). These two sets of results suggest that neurons in these spots were activated when arbitrary local features were arranged in a particular spatial configuration. Further evidence supporting this view is provided by direct comparison between responses to variations in color, texture, and local shape of parts and those to variations in the spatial arrangements of parts: the cells in SRR spots were more selective to particular spatial arrangements of parts than to examined variations in color, texture, and shape (Fig. 10). Therefore in terms of representation of object images, the neurons characterized in this study could play a role in specifying spatial relationships between parts. Altogether we found only four SRR spots among 26 activity spots elicited by "original" object images (16.7%). This relatively small proportion indicates that an object image consists of multiple local features and different types of spatial configurations. In the present study, we only investigated neural representation of one particular type of spatial configuration:

two parts aligned vertically. It should be noted that these spots did not respond to the combination of parts when the upper part was rotated 180° (Figs. 6, *C* and *D,* and 9, *C* and *E*), indicating that these neurons are capable of differentiating two parts. Sensitivity of the cells to some unidentified local cues could be essential for differentiating two parts. Further investigations will be necessary to fully understand the neural representation of spatial relationships among parts or local features in general.

Multiple studies with anesthetized as well as alert monkeys have shown that on average, cells in area TE have large receptive fields (Ito et al. 1995; Kobatake and Tanaka 1994; Op De Beeck and Vogels 2000). However, some reports shows that the size of receptive fields of IT cells could be a small portion of the visual field (DiCarlo and Maunsell 2002; also see Op De Beeck and Vogels 2000). Thus one may consider that the sensitivity to a particular spatial arrangement of parts (Figs. 6, 9, and 10) is simply due to upper parts of stimuli falling out of the receptive fields when upper parts were rotated relative to lower parts. One determining factor of receptive field size is the size of stimuli used in these investigations; it has been reported that the size of receptive fields increases when the size of stimuli increases (Ito et al. 1995; Op De Beeck and Vogels 2000). DiCarlo and Maunsell (2002) used stimuli as small as 0.6° but Kobatake and Tanaka (1994) used the stimuli as large as 10°. In the present study, we hand-plotted a receptive field for each cell; on average, receptive fields were as large as 42 × 46°. The stimulus size in our study was 13.3 × 10.9° on

average. Thus it is less possible that our results (Figs. 6, 9, and 10) are due to a decrease of responses when the stimuli fell out of the small receptive field, although quantitative analysis of the receptive field for neurons in SRR spots would be necessary in the future.

Finally, although it has been reported that many neurons in area TE respond to visual features less complex than natural objects, it has remained unclear whether these features are related to local features of object images or to more global features (Fujita 1993; Fujita et al. 1992; Ito et al. 1995; Kobatake and Tanaka 1994; Tanaka et al. 1991). Here, by global features we mean the combination of elementary components such as combinations of color and shape and local features. In particular, specification of spatial relationship among parts is one such global features. One important contribution of the present study is that it provides concrete evidence that critical features can be such global features of object images.

REFERENCES

Arieli A, Grinvald A, and Slovin H. Dural substitute for long-term imaging of cortical activity in behaving monkeys and its clinical implications. *J Neurosci Methods* 114: 119–133, 2002.

Biederman I. Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94: 115–147, 1987.

Brincat SL and Connor CE. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7: 880–886, 2004.

Bruce C, Desimone R, and Gross CG. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46: 369–384, 1981.

Desimone R, Albright TD, Gross CG, and Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062, 1984.

DiCarlo JJ and Maunsell JHR. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol* 89: 3264–3278, 2002.

Fujita I. Columns in the inferotemporal cortex: machinery for visual representation of objects. *Biomed Res* 14: 21, 1993.

Fujita I, Tanaka K, Ito M, and Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360: 343–346, 1992.

Gochin PM, Miller EK, Gross CG, and Gerstein GL. Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Exp Brain Res* 84: 505–516, 1991.

Gross CG. How inferior temporal cortex became a visual area. *Cereb Cortex* 5: 455–469, 1994.

Gross CG, Bender DB, and Rocha-Miranda CE. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* 166: 1303–1306, 1969.

Gross CG, Bender DB, and Gerstein GL. Activity of inferior temporal neurons in behaving monkeys. *Neuropsychology* 17: 215–229, 1979.

Ito M, Tamura H, Fujita I, and Tanaka K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73: 218–226, 1995.

Kobatake E and Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867, 1994.
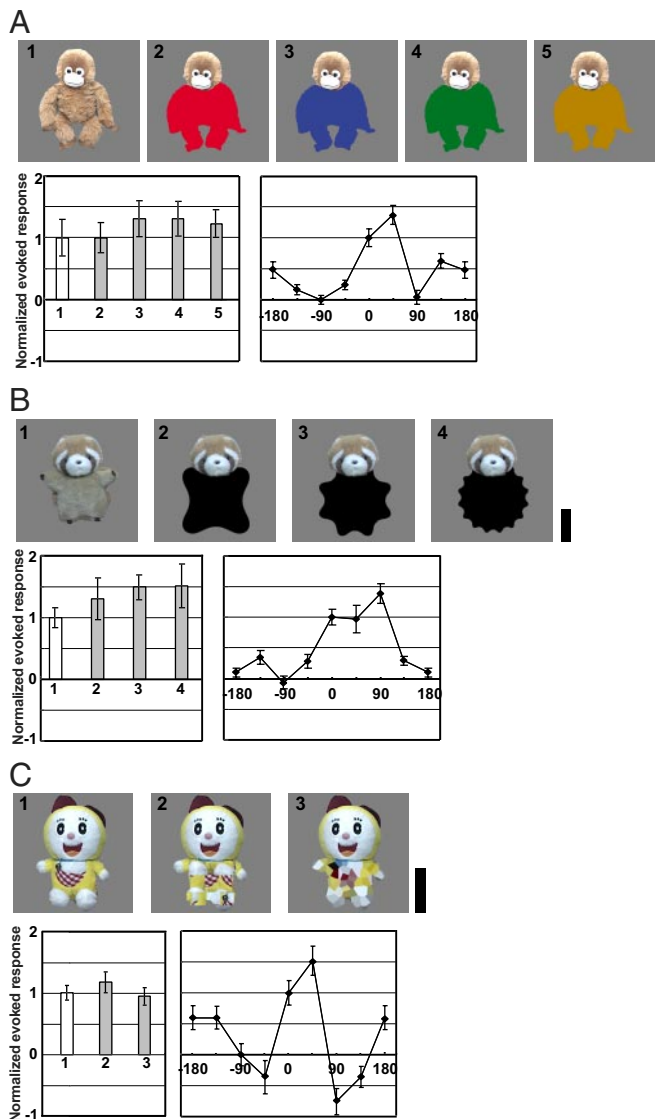
FIG. 10. Tuning specificity of representative neurons to stimuli in which lower parts were filled with different colors (*A*), had different shapes (*B*) or had altered textures (*C*). *A*: tuning specificity for color was examined for red (17.6, 0.506, 0.366), blue (11.7, 0.187, 0.213), green (19.3, 0.278, 0.518), and yellow (32.5, 0.399, 0.491). *Top*: [the numbers in the parentheses give the values for luminance (*Y*), and chromaticity coordinate (*x, y*) in CIE color space]. Among 10 cells (6 and 4 cells tested on changes to the *bottom* and *top,* respectively), no cell showed evoked responses that significantly depended on different colors, except for 3 cells tested on changes to the *top* (1-way ANOVA, $P > 0.1$). *B*: tuning specificity for shape was examined with Fourier descriptor stimuli (frequency, 4, 8, and 16; amplitude, 0.8) (*top*). Among 5 cells (4 and 1 cells tested on changes to the *bottom* and *top*, respectively), no cell showed evoked responses that significantly depended on shape (1-way ANOVA, $P > 0.1$). *C*: tuning specificity for texture was examined using stimuli in which the texture within one of the parts was modified. To create coarse texture modification, we partitioned either part of the stimulus into nine rectangular regions and shuffled them (stimulus 2). To create fine texture modification, we first made a Voronoi diagram based on randomly chosen pixels that were at most separated by 1.1° of visual angle. Then each area of the Voronoi region was filled with the color of the pixel which was its own base point (stimulus 3). Among 8 cells (5 and 3 cells tested on modifications to the *bottom* and *top*, respectively), only 2 cells showed evoked responses that significantly depended on differences in texture in either part (1-way ANOVA, $P > 0.1$). *A—C,* recorded from different neurons. In all figures, upper panel shows the stimuli, the *bottom left panel* shows evoked responses to these stimuli normalized to the responses elicited by the original best object image and *bottom right*: the tuning specificity of the same neuron to different spatial arrangements of parts of the best object for that cell. All of these cells were sensitive to particular spatial arrangements of parts. Scale bar, 5°.

Logothetis NK and Sheinberg DL. Visual object recognition. *Annu Rev Neurosci* 19: 577–621, 1996.

Marr D and Nishihara HK. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc R Soc Lond B Biol Sci* 200: 269–294, 1978.

Op De Beeck H and Vogels R. Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426: 505–518, 2000.

Perrett DI, Rolls ET, and Caan W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47: 329–342, 1982.

Schwartz EL, Desimone R, Albright TD, and Gross CG. Shape recognition and inferior temporal neurons. *Proc Natl Acad Sci USA* 80: 5776–5778, 1983.

Tanaka K, Saito H, Fukada Y, and Moriya M. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66: 170–189, 1991.

Tsunoda K, Yamane Y, Nishizaki M, and Tanifuji M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci* 4: 832–838, 2001.

Wang G, Tanaka K, and Tanifuji M. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272: 1665–1668, 1996.

Wang G, Tanifuji M, and Tanaka K. Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neurosci Res* 32: 33–46, 1998.

Zahn CT and Roskies RZL. Fourier descriptors for plane closed curves. *IEEE Trans Comput* 21: 269–281, 1972.