Chapter 5
How Images of Objects Are Represented in Macaque Inferotemporal Cortex

Manabu Tanifuji, Takayuki Sato, Go Uchida, Yukako Yamane, and Kazushige Tsunoda

Abstract Visual object recognition is a simple and easy task in our daily life. However, the mechanisms for recognizing objects are not at all simple nor easy. To understand neural mechanisms of object recognition, we have investigated representation of object images in macaque inferior temporal cortex that is the area essential for object recognition. Optical intrinsic signal imaging has revealed that object images are represented by the combinatorial code at the columnar level, where each column represents a visual feature of object images. The visual features represented by columns include local features as well as global features representing spatial arrangements of local features. Here, columns are supposed to be functional units for object representation. However, difference in object selectivity among nearby cells does not support the concept of columns as the functional units. Quantitative analysis of object responses of single cells and population activity revealed that each cell in a columnar region is characterized by cell specific property and property common across the cells in the columnar region, suggesting two different levels (single cell and columnar level) of object representation. Possible role of these two levels of object representation will be discussed.

5.1 Introduction

In order to understand neural mechanisms of object recognition, we have investigated representation of object images in inferotemporal (IT) cortex of macaque monkeys. This area of brain is essential for the recognition of objects by their visual images, and neurons in this area are known to respond to images of complex objects (Gross 1994; Gross et al. 1979; Desimone et al. 1984; Bruce et al. 1981; Perrett

M. Tanifuji (✉)
Laboratory for Integrative Neural Systems, RIKEN Brain Science Institute, Wako, Saitama, 351-0198, Japan
e-mail: tanifuji@riken.go.jp

A.W. Roe (ed.), *Imaging the Brain with Optical Methods*,
DOI 10.1007/978-1-4419-0452-2_5, © Springer Science+Business Media, LLC 2009

et al. 1982; Tanaka et al. 1991). Although some of these neurons specifically respond to certain objects such as faces, many IT neurons also well respond to visual features that are less complex than object images (Desimone et al. 1984; Bruce et al. 1981; Perrett et al. 1982; Tanaka et al. 1991). Since none of these visual features are specific enough, object representation takes the combined activation of multiple neurons to represent a particular object image in IT cortex. Investigations of object representation require analyses of activities of multiple neurons evoked by an object image.

Neurons responding to particular objects or visual features are not randomly distributed in IT cortex. Tanaka and colleagues found that neurons tuned to similar features were clustered together and formed a column in IT cortex (Feature column) (Fujita et al. 1992). Optical intrinsic signal imaging (OISI), that has columnar level resolution, is an appropriate technique to investigate object representation with multiple neurons (Grinvald et al. 1999). In fact, OISI revealed that an object image activates multiple spots (activity spots) in IT cortex, and that distribution patterns of activity spots are different from object to object (Tsunoda et al. 2001; Yamane et al. 2006). In this chapter, we present our studies with OISI that revealed what and how activity spots (corresponding to feature columns) are used to represent object images in IT cortex (Tsunoda et al. 2001; Yamane et al. 2006; Wang et al. 1996; Wang et al. 1998).

OISI is still limited in spatial resolution. It does not allow us to resolve visual responses at a level of single cells. Strictly speaking, object representation revealed by OISI is at the level of columns, and, in principle, columnar level representation could be different from representation at the level of single cells. In addition, recent studies have reported that object selectivity of nearby cells is largely different; these observations seemingly contradict to the column hypothesis (Tamura et al. 2005; Kreiman et al. 2006; Sato et al. 2009). Our recent study suggests that this apparent contradiction is due to the response property specific to neurons that obscure the response property common across the cells in a columnar region (Sato et al. 2009). The critical question is how to relate object representation at columnar level and a possible representation by combinations of activities of single cells in close vicinity. In the last part of the chapter, we discuss this issue. The above study also suggests that, in contrast to the columnar organizations in V1, the columnar structure in IT cortex may not uniformly cover entire IT (Sato et al. 2009).

IT cortex is subdivided into multiple areas. Our investigations focus on dorsal part of area TEa, that is the most anterior part of IT cortex where OISI is still applicable. Furthermore, the results shown here were based on the experiments with anesthetized monkeys. Thus, we cannot directly relate object representation observed in our studies to recognition in waking animals, but studies with anesthetized animals have an advantage that they can reveal object representation unbiased toward particular behavioral contexts.

To begin with, basic characteristics of intrinsic signals in IT cortex and physiological studies related to the columnar organizations are briefly reviewed, since studies on these two issues provide basis for our studies with OISI.

5.2 Optical Intrinsic Signal Imaging (OISI) in IT Cortex

Activation of neurons elicits changes in reflection of light from neural tissues. The changes (intrinsic signals) reflect three secondary physiological mechanisms associated with neural activation: deoxygenation level of hemoglobin (Vanzetta and Grinvald 1999), blood volume changes in capillaries (Fukuda et al. 2005; Vanzetta et al. 2004), and microscopic changes in tissue structure that cause changes in light scattering (MacVicar and Hockman 1991; Tsunoda et al. 2004). Relative contribution of these three components to intrinsic signals depends on the wavelength of light used for illuminating the cortex. In OISI from area TE, we used wavelengths at around 610 nm, where deoxygenation of hemoglobin is the dominant component. The typical time courses of the signals were not very different from those in V1: the signals reached a peak of increase in absorption at about 1–2 s after the stimulus onset, crossed the baseline absorption level at around 2–4 s. after the stimulus onset, and then showed absorption decreases that lasted for several seconds (Fig. 5.1). The initial increase in absorption corresponds to deoxyhemoglobin increase due to oxygen consumption by neurons activated by the stimulus and late decrease in absorption corresponds to replacement of deoxyhemoglobin with oxyhemoglobin due to increased blood flow triggered by the neural activation.

To obtain spatial patterns of intrinsic signals in area TE, we took images of cortical surface during 2.5 s period starting from 0.5 s after the stimulus onset by a CCD

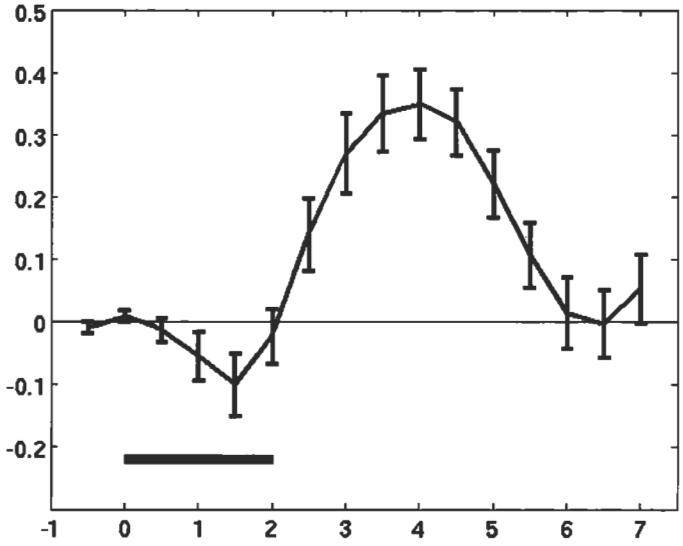

Fig. 5.1 The time course of intrinsic signals in area TE. Horizontal and vertical axes represent time (s) and reflection changes (%), respectively. *Horizontal bar*, the period for visual stimulation

camera, averaged them, and subtracted an image taken just before the stimulus onset from the averaged image (Tsunoda et al. 2001; Yamane et al. 2006). Thus, spatial patterns of intrinsic signals in area TE mainly reflect deoxygenation of hemoglobin induced by increased neural activity. Though OISI is typically used to visualize activation at the level of columns, raw intrinsic signals were not confined to columnar regions (global signals) (Fig. 5.2c). The global signals were, however, locally modulated. The local peaks of modulation (activity spots) were extracted by removing a low spatial frequency component of intrinsic signals (filtered image, Fig. 5.2d). Activity spots were then demarcated by drawing contours crossing half height between peak absorption of the spot and background in the filtered image (Fig. 5.2b). The magnitudes of the absorption changes in the filtered intrinsic signals were well correlated with the average of multi-unit activities (avgMUA) recorded from the same spot. For eleven activity spots, for example, the mean and standard deviation of correlation coefficient was 0.61 ± 0.17, and this value is statistically significant ($p < 0.05$). As shown later, avgMUA gives a good estimate of the neural activity at the columnar level. Thus, the high spatial frequency component

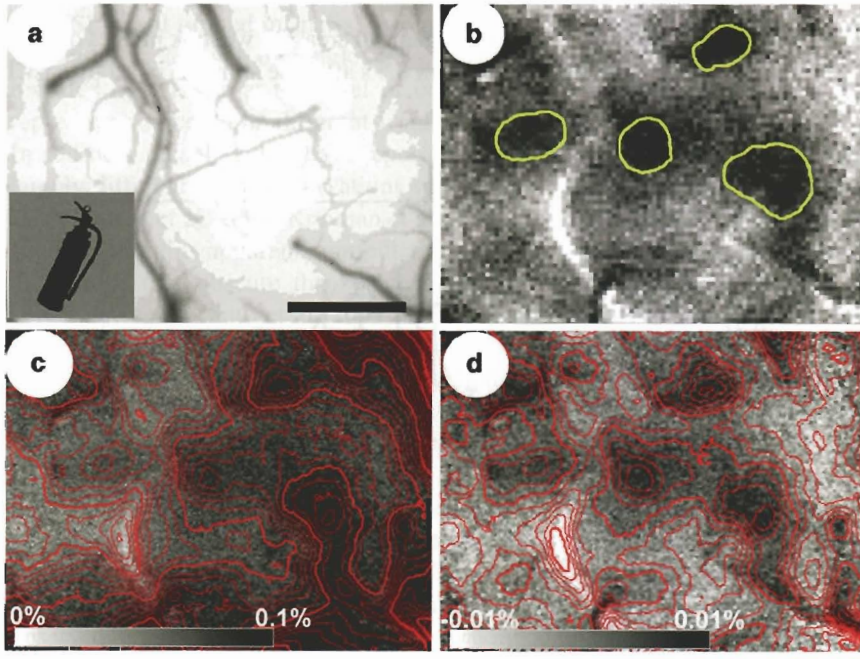

Fig. 5.2 A spatial pattern of intrinsic signals in IT. (**a**) Surface view of the exposed portion of IT cortex. Scale bar, 1 mm. (**b**) Activity spots outlined at the level of half the peak absorption value. (**c**) the spatial pattern of raw intrinsic signals without filtering. (**d**) spatial pattern of activity spots obtained by filtering raw intrinsic signals Scale bars in (**c**) and (**d**), percent changes in absorption increase. (modified from Tsunoda et al. (2001))

of intrinsic signals as in Fig 5.2b,d is in a good agreement with firing activity of population of neurons as seen in intrinsic signals in V1.

What about the low spatial frequency component of the raw intrinsic signals indicates? One possibility is that this low spatial frequency component (global signals) reflects synaptic potentials elicited by visual stimuli. To examine this possibility, we stained the cortex with the voltage sensitive dye RH1692 and compared the dye signal with intrinsic signals recorded from the same region (Homma and Tanifuji 2003). It is known that the voltage sensitive dye stains cellular membrane and that changes in flourescence of the dye reflect synaptic potentials rather than spiking activities (Grinvald et al. 1999). We found that the dye signals also revealed low and high spatial components, and local peaks of the dye signal coincide well with activity spots revealed by OISI. Since the low spatial frequency component of voltage sensitive dye signals was similar to that of the intrinsic signals, we think that raw intrinsic signals may also reflect synaptic potentials, and that intrinsic signals outside of the regions of activity spots reflect synaptic potentials below the threshold of neuronal firing. Similar argument was made for the global and local modulation of intrinsic signals observed in cat striate cortex (Das and Gilbert 1995). Interestingly, the low spatial frequency component is not uniformly distributed in space. For example, more absorption is seen in the right half than in the left half (Fig. 5.2c). Similar regional specificity was also observed in activation patterns recorded by voltage sensitive dye imaging with various object images. In this way, the distribution of the low frequency components was stimulus dependent. Some representations of object images may exist at the level of synaptic potentials as well as those at columnar levels.

5.3 Evidence for the Columnar Organization with Respect to the Critical Features in Area TE

Columnar organizations in the area TE were systematically investigated by Tanaka and colleagues (Tanaka et al. 1991; Fujita et al. 1992). They first determined a visual feature critical for each cell (Tanaka et al. 1991), and then investigated columnar organizations with respect to the critical features (Fujita et al. 1992). Here, the critical feature for a cell means the simplest visual feature that activates the cell equally well as the best object stimulus. To find such a critical feature, first, they searched for the most effective stimulus for each cell among more than a hundred of three-dimensional object stimuli. These stimuli included stuffed animals, plastic fruits and vegetables, and experimenter's hand and body. These stimuli were presented to the monkeys from various viewpoints to maximize the chance of finding the most effective image. Then, they generated modifications of the most effective stimulus and examined responses evoked by the simplified stimuli. If the cell responded to one of the simplified stimuli equally well as the original object stimulus, this new stimulus was further simplified. This procedure was repeated until the experimenter could not simplify the stimulus without significant reduction of the

evoked responses. Figure 5.3 shows a representative case of the stimulus simplification procedure according to their method, where we found a combination of the circle and rectangle was essential for the maximal activation of a cell. This cell was recorded from the activity spot that was expected to represent spatial relationship among object parts (discussed later in detail).

Tanaka and colleagues examined many TE neurons in this way and identified visual features essential for activating individual cells (Tanaka et al. 1991; Kobatake and Tanaka 1994). Many of these features are the combinations of simple shapes,

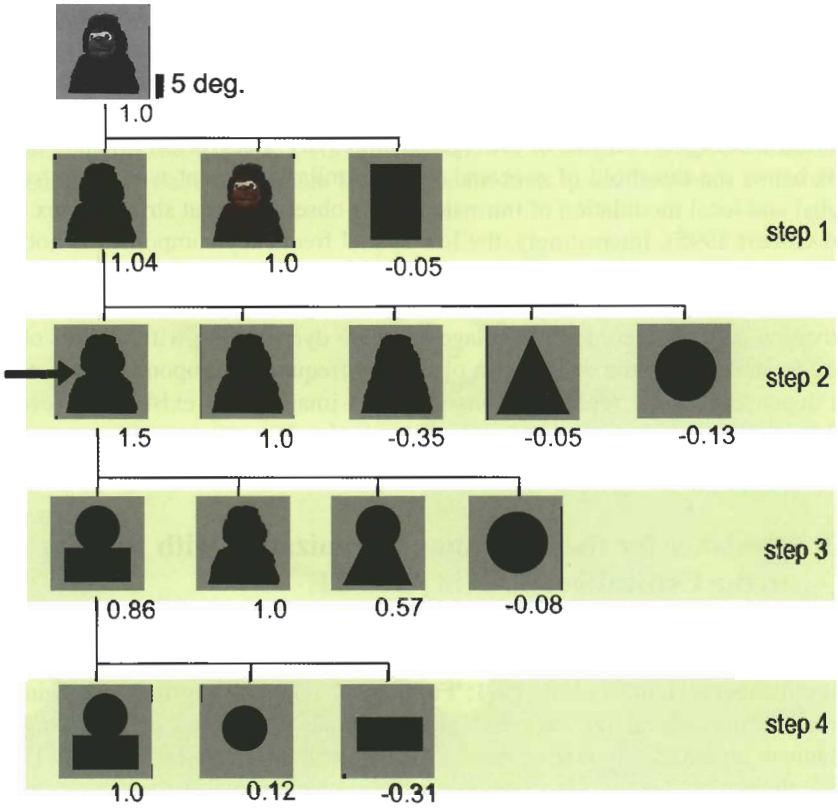

Fig. 5.3 Systematic stimulus simplification of an object image. The most effective object stimulus (*top* picture) was simplified step by step to determine the simplest stimulus that maximally activates the cell. Step 1 shows that neuronal activity elicited by the most effective object and its silhouettes are the same. The number below each picture indicates the response amplitudes normalized to the response to the reference stimulus. Step 2 examines the effect of the "sharpness" of the corner at the junction of the *upper* and *lower* parts (*arrow*) and shows that the silhouette with the sharpest corners (*leftmost* picture) is the most effective stimulus. Step 3 shows that activities elicited by the two *leftmost* stimuli are not significantly different. Step 4 shows that neither the *upper* nor *lower* part activates the cell. In this case, the critical feature was determined as the combination of a circle and a rectangle (*leftmost* picture in step 4). (Modified from Yamane et al. (2006))

colors, luminance gradient/contrast, and textures. These features are more complex than the optimal visual stimuli for cells in areas V1, V2, and V4, but still less complex than natural objects (Kobatake and Tanaka 1994).

Fujita et al. (1992) systematically examined the columnar organization in the area TE with respect to the critical features. First, they penetrated electrodes perpendicular to the cortical surface and determined a critical feature of one neuron within the track. Then, they prepared a set of visual stimuli including this critical feature and its modifications and examined the response selectivity of other neurons within the track using the set of visual stimuli. This experiment showed that the most effective stimulus for the cells in the track is the critical feature or visual features similar to it. They also examined neuronal responses along a tangential electrode penetration and found that neurons with similar responsiveness were localized within the range about 0.4 mm along the tangential track. These results suggest a columnar organization in the IT cortex with respect to the critical features of the cells. However, we often encounter nearby cells spaced only for about 150 μm apart but having different selectivity for object images. We will discuss later in this chapter what makes this difference in the selectivity of the critical features and object images.

Many investigations suggest that neurons in IT cortex were plastic. For example, the above group trained monkeys for a specific set of visual stimuli with a delayed match-to-sample task (Kobatake et al. 1998). They found that, compared with naïve monkeys, IT cortex of the trained monkeys contained more neurons tuned to a trained stimulus: the response to the best of the trained stimuli was higher than the responses to any other object images. Because we cannot keep tracking responses of a single neuron during training that takes more than a month, these results may not indicate that training changed visual responses of IT cells. Neurons that were previously not visually driven may turn out to be responsive to the trained stimulus. In either case, the columnar organization in IT cortex, if it exists, seems not static but could be modified by extensive experience of the monkeys.

5.4 Object Representation by Combinations of Activity Spots in Area TE

OISI revealed that an object image activates multiple activity spots (Figs. 5.2 and 5.4). The spatial patterns of these activity spots are different from object to object. Some of the spots were co-activated by different objects, while other spots were activated only by one object (Fig. 5.4a, b). Assuming that each spot represents a visual feature, activation of the spots specific to a single object is likely to indicate that the other objects lack the visual feature represented by these spots. For example, the spots only activated by stimulus 1 may be related to horizontal red/blue stripes only seen in stimulus 1 (Fig. 5.4b). Comparison between the distribution patterns of activity spots and those produced by systematically simplified stimuli revealed that this was indeed the case (Fig. 5.4c). Here, a "black cat" (c-**1**) was

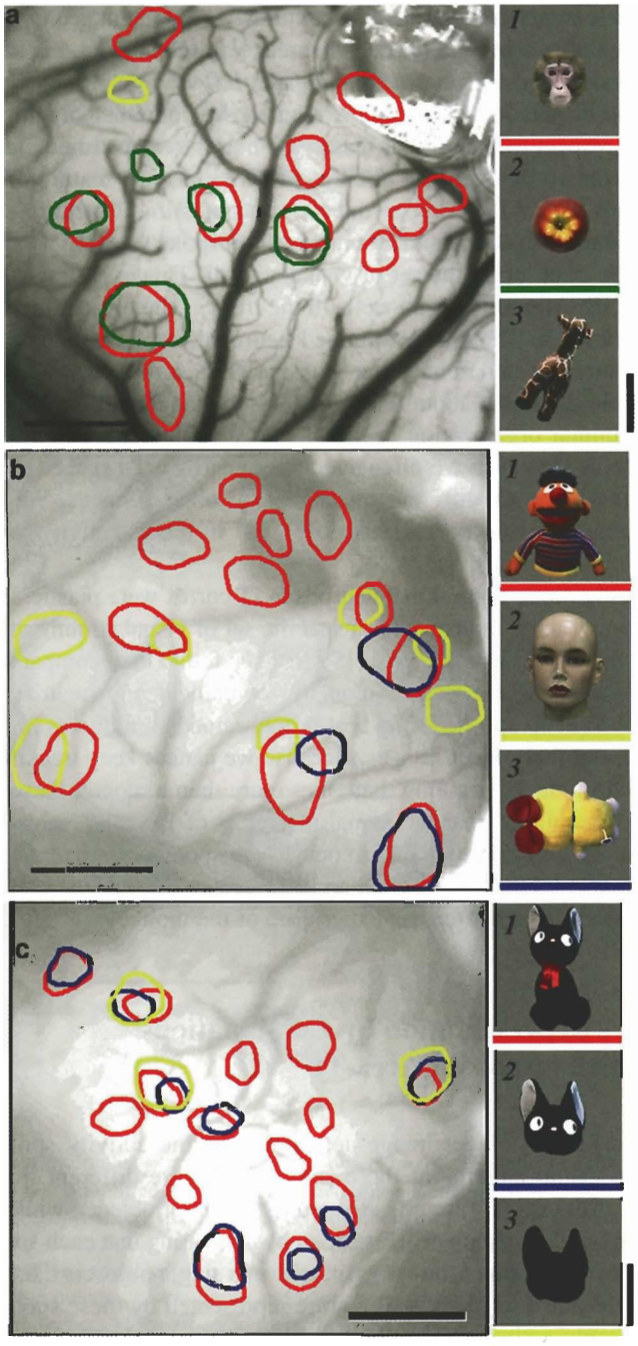

Fig. 5.4 Distribution of activity spots elicited by various object images (**a**,**b**) and simplified images (**c**). Contours of the activity spots were drawn as in Fig. 5.2. Horizontal scale bars, 1.0 mm. Vertical scale bars, 10°. (Modified from Tsunoda et al. (2001))

simplified to the "head" (c-**2**), and then to the "silhouette of the head" (c-**3**). The original image (c-**1**) elicited 14 spots, but presenting the "head" (c-**2**) elicited only eight spots of the original fourteen. The "silhouette" (c-**3**) only activated three (arrows) of the eight spots elicited by the "head" (c-**2**). Simplified stimuli lacking some of the visual features of the original image activated only a subset of the spots elicited by the stimuli before simplification. We examined 12 pairs of activation patterns elicited by the original and the simplified stimuli, and found that five pairs (42%) showed similar results.

Interestingly, in addition to the disappearance of spots, there were also cases where new spots emerged by the simplification of an object (Fig. 5.5). For example, in Fig. 5.5a, spots A and B disappeared but spot C appeared when stimulus **1** was simplified to stimulus **3**. Similarly in Fig. 5.5b, spot A was only activated by the simplified stimulus. Among the twelve pairs, the emergence of spots by simplification was observed in seven pairs (58%) (two cases showed only emergence and five cases showed both emergence and disappearance).

To identify the visual feature represented by each spot and also to find reasons for emergence of activity spots with stimulus simplification, we recorded single cellular activities from neurons in spots shown in Fig. 5.5a (Fig. 5.6). Cells in spots A and B were significantly activated by the "handle and hose" in isolation (Fig. 5.6a-1, b-1). This result is consistent with optical response patterns elicited by stimuli 1 and 3 (Fig. 5.5a). In addition, the cells in spot A were activated by the "handle" (Fig. 5.6a-2) having protrusions, but not by the "hose" (Fig. 5.6a-3). Furthermore, other stimuli with sharp protrusions, such as a "hand" (Fig. 5.6a-4) and "cat's head" (Fig. 5.6a-5), also activated the cells. These cells seemed to require a generic visual feature, "sharp protrusions," for activation. In contrast, the cells in spot B were activated by the "hose" (Fig. 5.6b-3), but neither by the "handle" (Fig. 5.6b-2) nor by a "line segment" (Fig. 5.6b-4). We also found no activation was elicited by a segment of a circle (Data not shown). The critical feature of these cells was, thus, an "asymmetric arc."

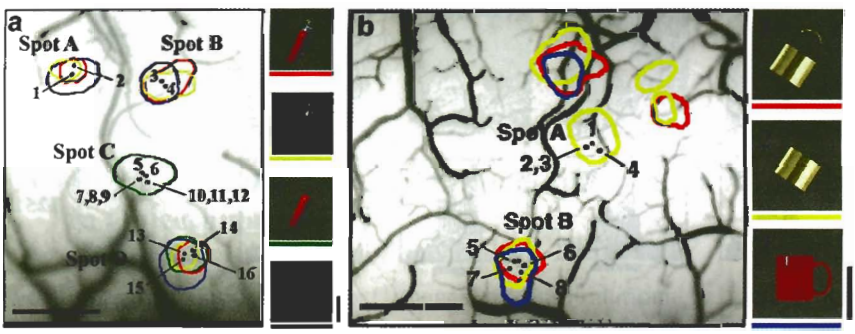

Fig. 5.5 Stimulus simplification causes appearance of new spots. The numbers indicate electrode penetration sites. (see Tsunoda et al. 2001 for electrode recordings in b) (Modified from Tsunoda et al. (2001))

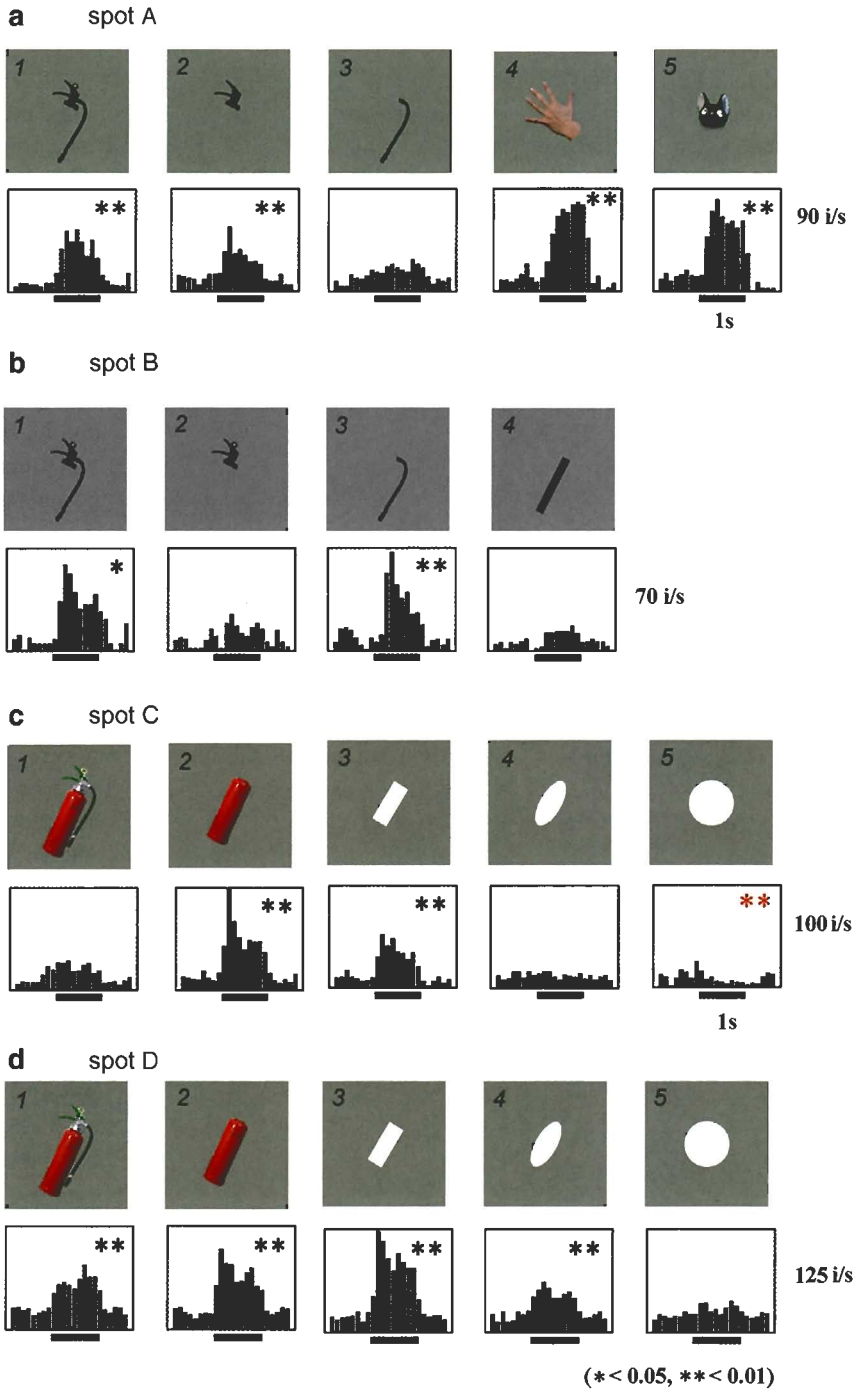

Fig. 5.6 Visual responsiveness of representative cells in spots A–D in Fig. 5.5a. Each histogram (PSTH) shows neuronal responses elicited by the above stimulus. Red asterisks indicate significant inhibition ($p<0.01$). *p <0.05, **$p<0.01$. Scale bars, 1-s periods of stimulation. (Modified from Tsunoda et al. (2001))

OISI revealed that the "cylinder" produced significant responses in both spots C and D (Figs. 5.5a-3), while other stimuli having "the cylinder" as a part (Figs. 5.5a-1, a-2 and a-4) only activated spot D but not spot C. The neural responses of cells in these spots were consistent with the imaging results: the cells in spot C was activated by the "cylinder" but not by the original "fire extinguisher" (Figs. 5.6c-1 and c-2), and a cell in spot D was significantly activated by both stimuli (Fig. 5.6d-1, d-2). The feature critical for the cells in spot D was a "rectangular shape" (Fig. 5.6d-3), but the cells also responded significantly to an "ellipse" (Fig. 5.6d-4). Since there were no responses to a "circle" (Fig. 5.6d-5), an "elongated structure" seemed to be necessary for activation. The simplest visual feature for the cells in spot C was also a "rectangular shape" (Fig. 5.6c-3). In contrast to the cells in spot D, however, none of the cells was activated by an "ellipse" (Fig. 5.6c-4). In addition, all of these cells were inhibited by the "circle" (Fig. 5.6c-5). These results suggest that the response properties of the cells in spot C (Fig. 5.5a) were determined by the balance between excitatory and inhibitory inputs, that is, the excitatory inputs were given by a feature related to a rectangular shape and the inhibitory inputs were given by a feature related to a circular shape. This explanation would account for the lack of activation by "the fire extinguisher," where the hose having a circular shape was attached with the cylinder having a rectangular shape. In general, these results suggest that some of the spots representing a particular feature were inactive when other features were presented together with the feature. This could explain that active spots appeared following simplification of stimuli in some cases.

So far, we have revealed two factors involved in representation of object images in the area TE. First, some of visual features represented by activity spots are local features such as protrusion and asymmetric arc. We refer "local features" to as visual features that occupy part of an object image and are distinguishable from other parts of an object image by their particular shapes, colors, or textures. Second, specific object representation is not made by combinations only of active spots but also of inactive spots. In the case of representation of original fire extinguisher, activation of spots A and B represents the presence of protrusion and curvature in the object, respectively. Activation of spot D indicates that entire structure is elongated, but no activation in spot C further indicates the structure has to be relatively elliptic. In such a way, combinations of inactive as well as active columns help representation of object images to be more specific.

5.5 Representation of Configurational Information Appeared in Object Images

Specific representation of object images by combinations of local features requires mechanisms to represent information about the spatial arrangement of "local features" or about spatial arrangement of parts including local features ("configurational information"). In the previous section, we showed that spot C in Fig. 5.5a was not activated when the hose is attached to the side of the cyl-

inder and makes the entire shape elliptical. This spot, however, would be activated if the hose were secured above the handle where the rectangular shape of cylinder was exposed. From the viewpoint of representing "configurational information," we may consider that activity in spot C carries information about the position of the hose relative to the cylinder, although the way to represent "configurational information" about the relative position of hose and cylinder is indirect.

We searched for the neural substrates that explicitly represent "configurational information." Particularly, we explored representation of a particular spatial relationship, that is, "on top of." This is a typical spatial relationship appeared in object images. For example, the head is above the body, the lampshade is on top of the base, and a pineapple is separated into upper part leaves and lower part fruit. To find candidate spots related to representation of the spatial relationship, we searched for the spots that showed a specific pattern of activation: activation by the object consisting of two parts and the same object with a gap introduced between parts of the object, but no activation by a part alone (Yamane et al. 2006). Among four hemispheres, we found three spots showing this specific pattern of activation (Fig. 5.7). These spots should not simply represent local features of an upper or lower part of the objects because either part is not essential for activation. Moreover, activation by the stimulus with an introduced gap indicates that local features appearing at the junction of two parts, such as sharp connecting corners, are also not essential. Thus, these spots were good candidates that could represent spatial relationship, "on top of." For further characterization of these spots, we made single cellular recordings from these spots, and found that three unique response properties. First, the cells in these spots preferred object images consisting of vertically aligned two parts with a small number of exception (Fig. 5.8). The critical features for these cells determined by the stimulus simplification procedure were also the combinations of vertically aligned two parts except two (Fig. 5.9b-7, b-8). Second, these cells were less sensitive to color, texture, and local shapes of either part. Thus, there were no changes in the responses after removing color and texture during the stimulus simplification procedure (Figs. 5.3 and 5.9). The changes in shapes of the parts did not significantly alter responses of these cells. For example, a neuron, whose critical feature was determined as a combination of a circle and a rectangle, was also significantly activated by a combination of a circle and an ellipse. Third, these cells were highly selective to a particular spatial arrangement of the upper part and the lower part (Fig. 5.10). Most of the cells were maximally activated when the upper and lower parts were vertically aligned or tilted only for 45° from vertical arrangement. We cannot explain the selectivity to a particular spatial alignment of upper and lower parts by changes in retinotopic positions of the upper parts that occurred incidentally during the spatial rearrangements of the parts because the receptive field of these cells covered even larger area in the visual field.

These response properties enable these spots to respond to two-part objects regardless of the local features embedded in either part, but only when the parts are aligned vertically (Fig. 5.8). These results, as well as neurons in spot C in

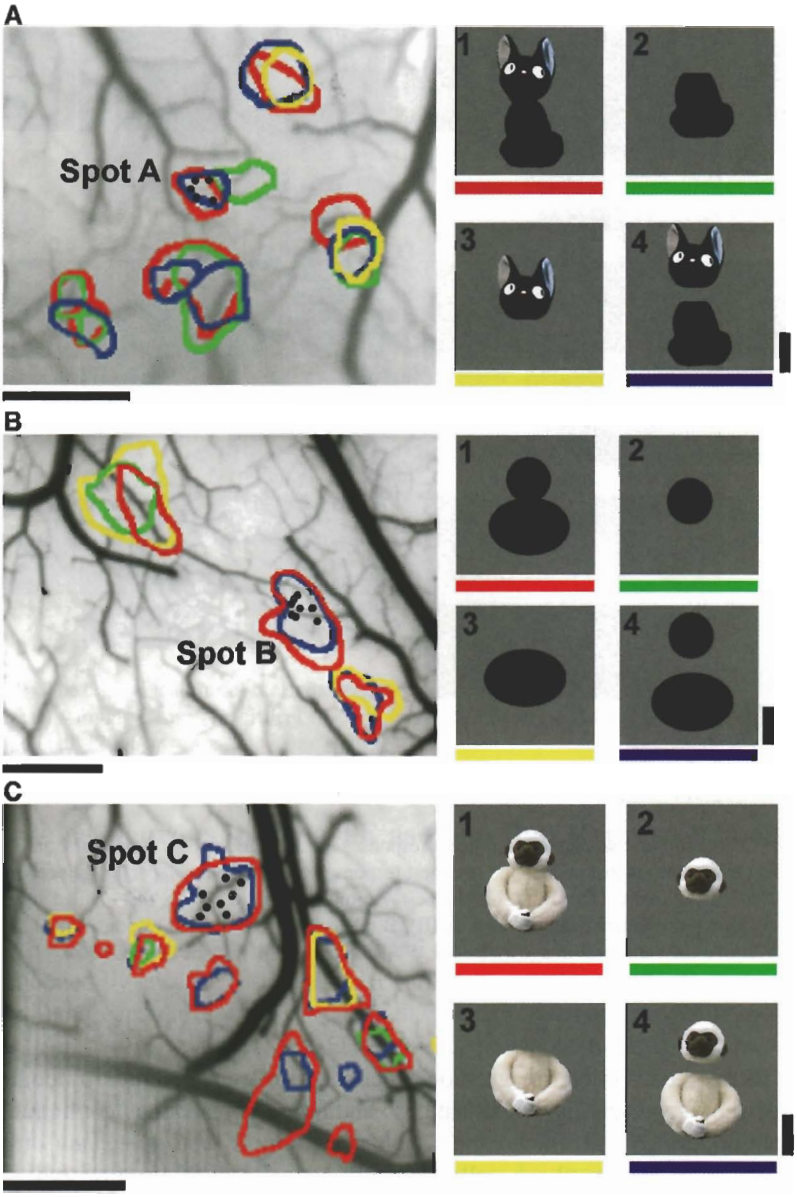

Fig. 5.7 Distribution of activity spots elicited by the stimulus set designed for searching spots related to spatial arrangement between *upper* and *lower* parts. Three spots A, B, and C were identified. *Black dots*, electrode penetration sites (Modified from Yamane et al. (2006))

Fig. 5.5a, suggest that neurons in area TE do not necessarily represent local features but also "configurational information" of the object images. As shown below, we consider that face neurons also represent configurational information

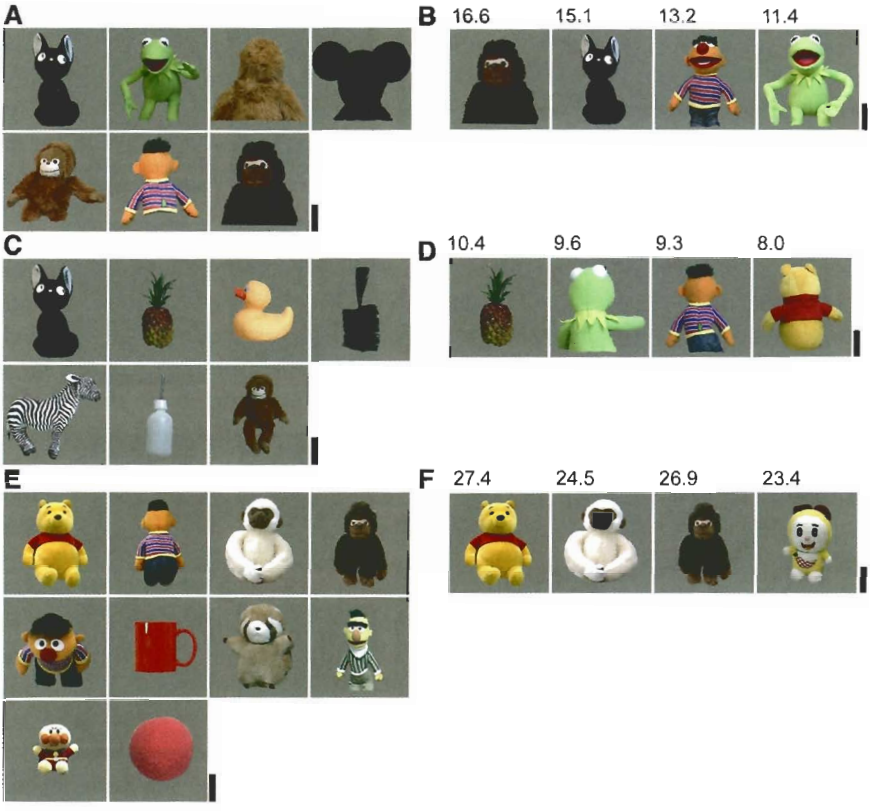

Fig. 5.8 Object stimuli that elicited significant responses for the cells in spots A, B, and C in Fig. 5.7. (**A, C, E**) Object images that elicited the strongest responses out of the 96 objects for cells recorded in spots A (**A**), B (**C**) and C (**E**). (**B, D, F**) The best 4 stimuli out of 96 objects that elicited significant responses for a representative cell in spots A (**B**), B (**D**), and C (**F**). Evoked responses (spikes/s) are indicated above each stimulus image. Scale bar, 5° (Modified from Yamane et al. (2006))

about facial parts. Object images could be specifically represented by combinations of spots representing "local features" and those representing "configurational information."

In the above study, these cells responded to the configuration where the upper part was above the lower part, but not to the configuration where the upper part is below the lower part (Fig. 5.10). This result means that, in some ways, the cells distinguished the upper part and the lower part of objects. In this respect, the cells were not entirely insensitive to local features. At least, the cells could distinguish upper and lower parts. These cells may be sensitive to difference of area of parts. Alternatively, some combination of a curvature in the upper part and in the lower part may be the critical factor (Brincat and Connor 2004).

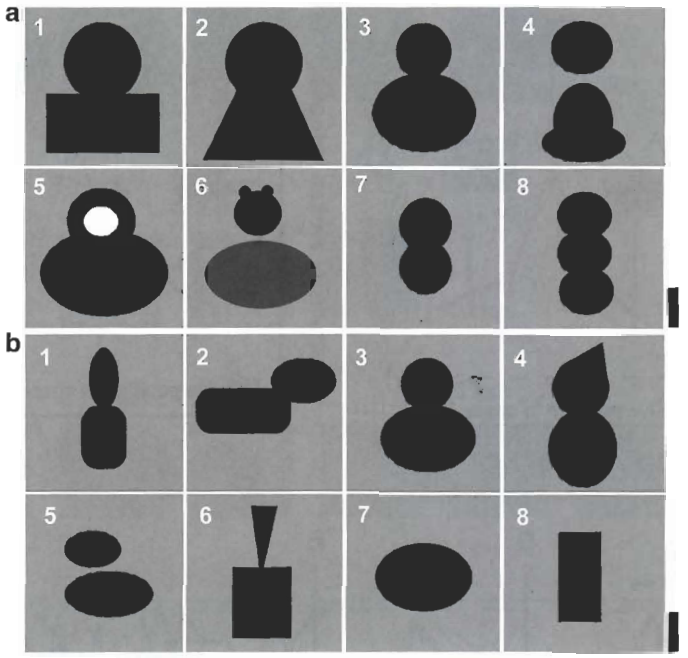

Fig. 5.9 The critical features of the cells in spots A(a) and B(b) related to representation of spatial relationship among parts. The critical features were not investigated for spot C. (Modified from Yamane et al. (2006))

5.6 Face Neurons in Area TE as Ones that Represent Facial Configuration

"Face neurons" is the neurons that respond to "faces," but these responses cannot be explained by specific responses to a part of "face" (Fig. 5.11). For example, a face without eyes did not activate the cell, but there was no activation by "eyes" alone (Fig. 5.11b right most, c). Furthermore, previous studies have shown that "face" with scrambled facial parts do not activate these neurons (Bruce et al. 1981). There are two characteristic properties of "face neurons." First, many of them are tuned to images of faces from a particular vantage point (Fig. 5.11a)(Perrett et al. 1982; Perrett et al. 1991). Second, these cells are less sensitive to difference of individual faces (Perrett et al. 1984; Baylis et al. 1985; Yamane et al. 1988; Young and Yamane 1992). These response properties suggest that the face neurons represent not specific faces but facial configuration.

Intrinsic signal imaging showed that there are spots specifically activated by faces (Fig. 5.12) (Wang et al. 1996; Wang et al. 1998). Thus, face neurons, as well as neurons specifically responding to visual features, are clustered together. Furthermore, activation patterns produced by images of faces from different vantage points revealed that the peaks of activity spots shift along the cortical surface as the

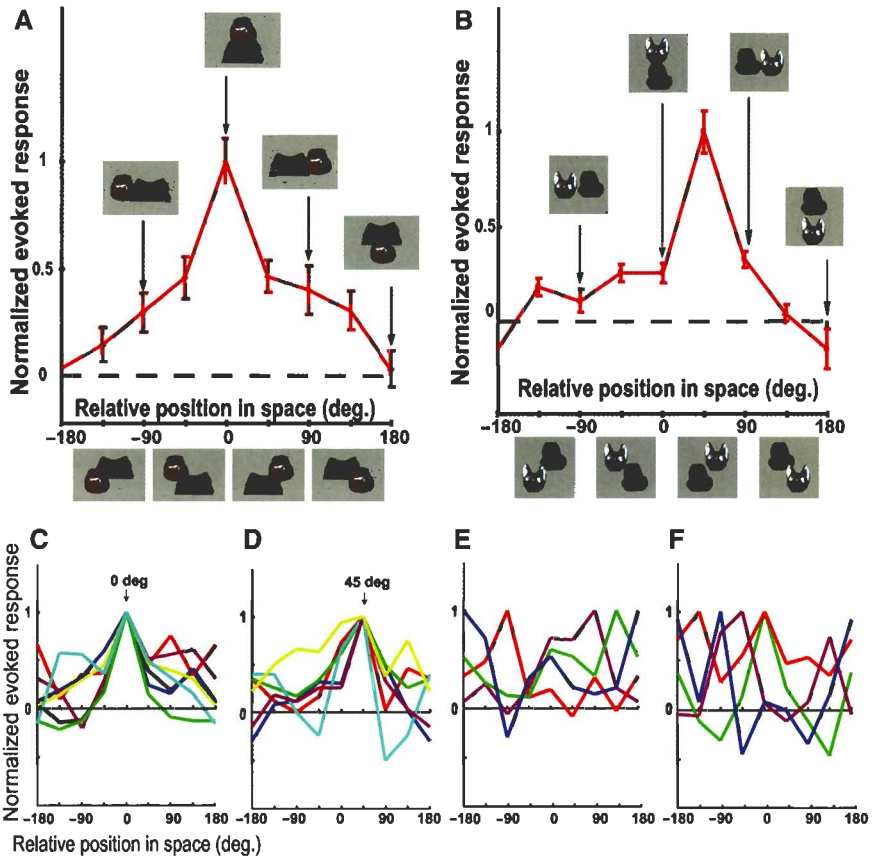

Fig. 5.10 Selectivity of cells to different spatial arrangements of the *upper* and the *lower* parts of object images. The normalized evoked responses (vertical axis) were plotted against the difference in the spatial arrangement of the parts (horizontal axis). The difference in spatial arrangement is defined by the angle between a line connecting the centers of the two 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. The panels (**A**) and (**B**) show responses of representative cells in spots A and B in Fig. 5.7, respectively. (**C-F**) Tuning curves for other cells in others in spots A, B, and 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. (Yamane et al. (2006))

face rotates from the left profile to the right profile through the front face. This representation of faces from different vantage points in close vicinity may be important for view-independent recognition of faces.

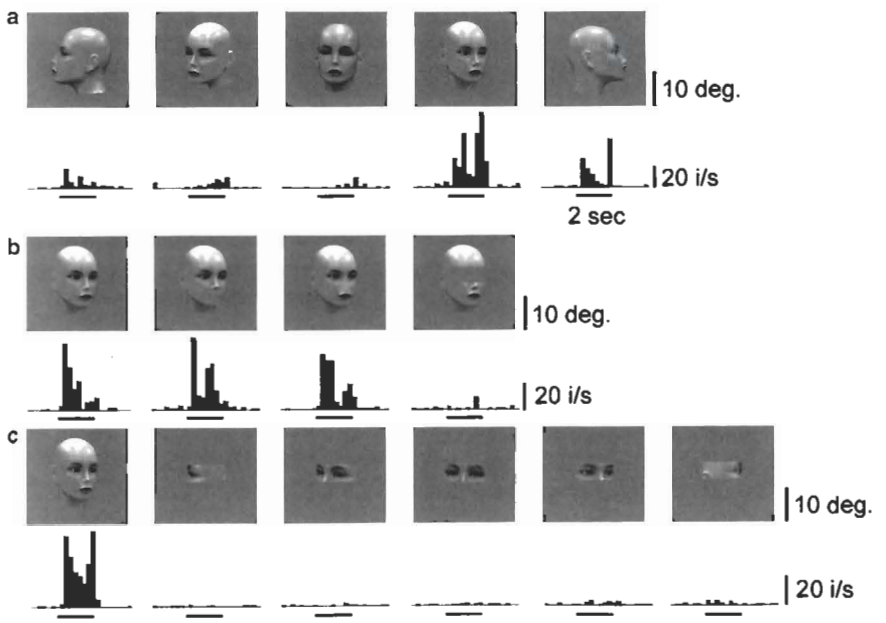

Fig. 5.11 Responses of a face neuron. The recording consists of three sessions. In the first session (**a**), the selectivity of the neuron for different views of a face is demonstrated. In the second and third sessions (**b**, **c**), selectivity of the same neuron for a face and facial parts were examined. Please note that no activation by either a face without eyes (**b**) or an eye alone (**c**) was observed. (Cited from Fukuda and Tanifuji, unpublished observation.)

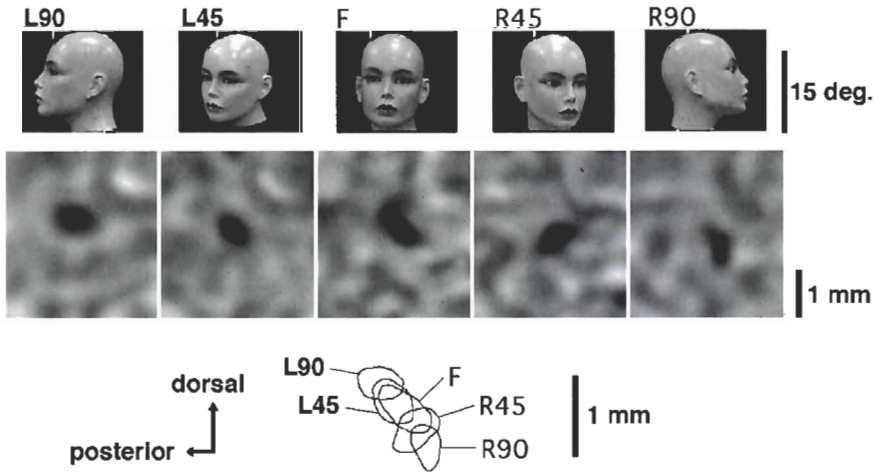

Fig. 5.12 Systematic shift in distribution of activity spots with rotation of the face. Images of the same cortical area (*middle panels*) obtained from five different views of the same mannequin face (*top panels*). The contours of the active spots are superimposed at the *bottom* (Modified from Wang et al. (1996))

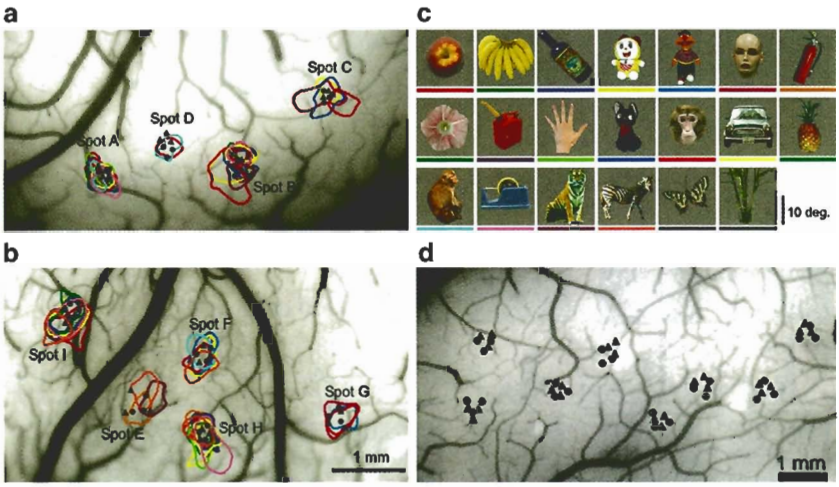

Fig. 5.13 The sites for electrophysiological recordings of single cells and MUs. Electrode penetration sites are given in filled circles and triangles. (**a, b**) The activity spots were shown by colored contours indicating area of activation by object images shown in (**c**). Contours were drawn as in Fig. 5.2. (**c**) Stimulus images for IOSI. The line type and the color are matched to the contours in **a**, and **b**. (**d**) The recording sites in the hemisphere where no IOSI was conducted beforehand. (Modified from Sato et al. (2009))

5.7 Object Representation at Different Levels: Columns and Single Cells Within a Column

Recently, several studies have quantitatively examined object selectivity of IT cells and found that IT cells were tuned to a relatively small number of object images (Tamura et al. 2005; Kreiman et al. 2006; Sato et al. 2009). One measure to quantify specificity of stimulus tuning curves is the sparseness index (Rolls and Tovee 1995). We obtained the value of the sparseness index for TE neurons for randomly chosen 80 object images as 0.19 ± 0.18 ($n = 218$). This value, 0.19, means that the neuron responds to only 19% of stimuli, if we approximate neuronal responses to stimuli in an all-or-none fashion (0 or 1). Furthermore, these studies have shown that selectivity of nearby cells was largely different from one another (Tamura et al. 2005; Kreiman et al. 2006; Sato et al. 2009). For example, five object images that evoked the strongest visual responses were different for two neurons even if these neurons were sampled from the same activity spot and spaced only for 150 μm apart (Fig. 5.13, 5.14). The correlation of object selectivity of nearby two cells within the same activity spots was as low as 0.15 in the mean value of the correlation coefficient, and the number of cell pairs that had significant correlation was only 28.5% ($p < 0.05$) (Fig. 5.15a) (Sato et al. 2009). The difference in object selectivity of nearby cells was not an artifact due to trial-by-trial variability of neuronal responses. The correlation coefficient expected from the trial-by-trial variability

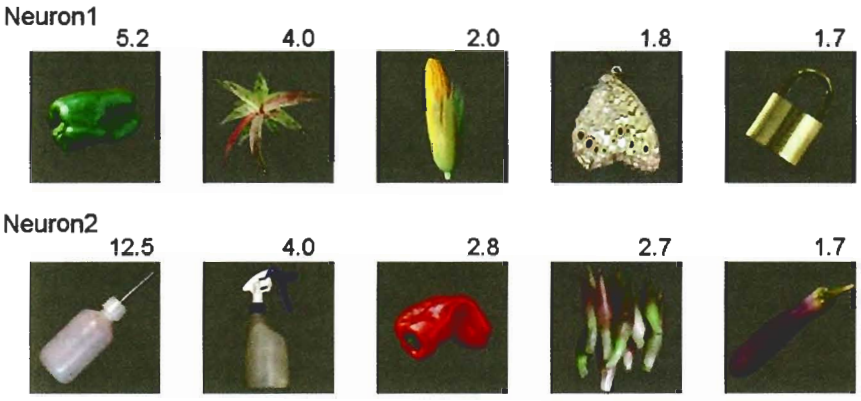

Fig. 5.14 The best five stimuli of two nearby cells spaced 150 μm apart. These two cells were isolated from spot A in Fig. 5.13. The number above each stimulus was the evoked response by the stimulus (spikes/s). The correlation coefficient of selectivity of these two cells for 100 object images was 0.07

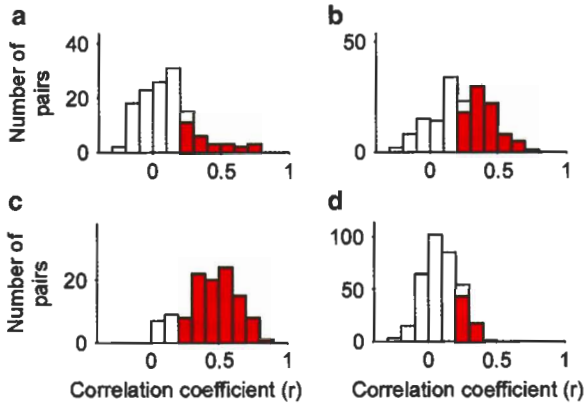

Fig. 5.15 Similarity in object selectivity between pairs quantified by correlation coefficient of evoked responses to 80 object stimuli. The pairs are made of single isolated cells (**a**), and MUs (**b**). Proportions of pairs that had significant correlation were 28.5% (**a**) and 60.0% (**b**). In (**c**), the correlation coefficient was calculated for the pair of MUs and the avgMU obtained from the same columns. Proportion of MUs that had significant correlation in object selectivity with avgMUs was 85.2%. On the other hand, in (**d**), we calculated correlation coefficient for the piars of MUs and avgMU obtained from different columns, Proportion of MUs that had significant correlation was only 16.4%. Thus, the common property represented by the avgMU was unique to the activity spot where the avgMU was calculated and was different from other spots. Horizontal axis, the value of correlation coefficient. The columns with red color represent group of pairs that has statistically significant correlation ($p<0.05$). (Modified from Sato et al. (2009))

was around 0.4 in correlation coefficient, and this value is higher than the value of similarity in object selectivity of nearby isolated cells. Thus, although evidence suggests a columnar organization in IT cortex (Fujita et al. 1992), object selectivity of nearby cells seemingly contradicts to the column hypothesis in IT cortex.

Furthermore, these findings raise a possibility that object images are specifically represented even by combinations of single cells in a local region that have different response property.

To resolve this contradiction, we analysed object responses of single cells and multi-units (MUs) recorded from activity spots (Fig. 5.13), and explained the reason for the difference in object selectivity of nearby cells that cell-specific response property obscured the common response property across the cells within a columnar region (Sato et al. 2009). In accordance with this explanation, the number of pairs of multiunits (MUs) (60%) showing significant correlation in object selectivity was higher than that of single neuron pairs (28.5%) probably because the variation caused by cell-specific responses were averaged across the cells and the common property became apparent in MU responses (Fig. 5.15a, b). Furthermore, we found that this common property (quantified as the average of MU activities recorded from the same activity spot (avg MU activity)) was different from activity spot to activity spot (Fig. 5.15c, d). Thus, although there is cell-to-cell variability in object responses, the columnar structure is maintained since these cells share a common response property. The relationship between the common property and cell-specific property is demonstrated in Fig. 5.16 by plotting object responses of individual cells against descending order of object responses of the avgMU activity. Furthermore, the PCA analysis of MUs recorded from multiple activity spots revealed that one activity spot was characterized by one or a few common properties in object responses (Fig. 5.17). As we mentioned earlier, analysis of the critical features of individual cells first revealed the columnar organization in area TE (Fujita et al. 1992). We consider that the procedure of stimulus simplification made them find the visual feature that represents the common property across the cells within a columnar region, although the authors of the paper did not intended to search for the common property across the cells. Recently, it has been shown that nearby cells in V1 also respond differently to natural images although these cells should have similar orientation preference (Yen et al. 2007). Thus, the differentiation between the response property specific at a single cell level and the response property at the columnar level may be a universal characteristic in cerebral cortices.

The critical question particularly in area TE is how to relate object representation by combinations of columns and a possible representation by combinations of single cells in a local region. Taking into account trial-by-trial variation of neuronal activity, we consider that representation at the columnar level has primary importance. Neuronal firing approximately follows Poisson distribution, and thus, trial-by-trial variability in neuronal responses is higher for the good stimuli that evoked strong mean responses than the stimuli that evoked weak mean responses. Thus, in general, it is difficult to extract information about the stimulus from activity in a single trial of a single neuron. Since we recognize an object instantaneously without trial-by-trial averaging, one plausible way to extract reliable sensory information is to take an ensemble average of responses of nearby neurons. Our results above point out that this ensemble average reflects tuning specificity of feature columns, and the specificity is generally different from single cells. Accordingly, we consider that recognition of object images presented at a

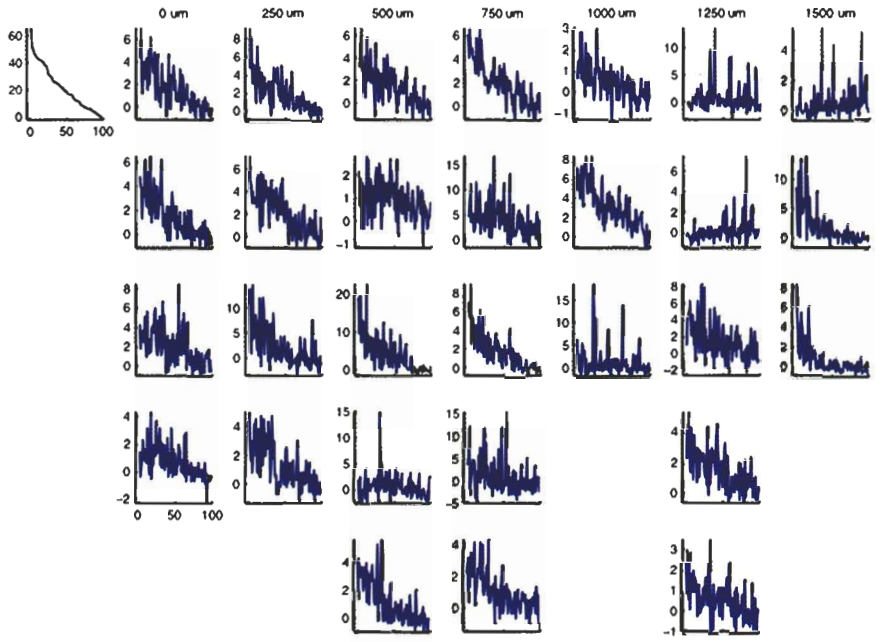

Fig. 5.16 The common response property revealed in tuning curves of the individual cells. Evoked responses of each cells were plotted against object stimuli rank ordered by the avgMU responses. The graph at the *left upper* corner represents the tuning curve of averaged MUs and the rests represent tuning curves of single cells at different depths. Depth of cells in each column is indicated at the *top*. Horizontal axes are rank-ordered according to the magnitude of evoked responses of averaged MUs to the 100 object stimuli in descending order. Vertical axes represent mean firing rate (spikes/s). As you see in single cell tuning curves, though the object that elicited the strongest responses was different from cell to cell, tuning curves of these cells have general tendency that rightward objects evoked weaker responses than leftward objects. (Modified from Sato et al. (2009))

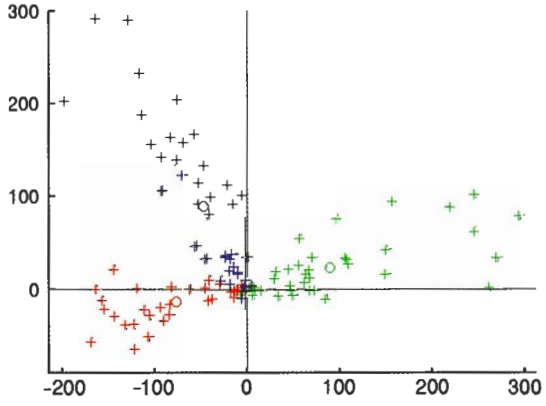

Fig. 5.17 One response property for each spots is approximated in comparison with difference in response property across activity spots. Object Responses of MUs of three spots (crosses; different color indicates different spots) are plotted in a multiple dimensional space made of 80 object images, and projection onto the 2-dimensional plane that includes responses of avgMU in three activity spots (open circles) is shown. (Modified from Sato et al. (2009))

glance is achieved based on representation of object images not at the level of single cells but at the columnar level. It is still an open question in what occasion representation of object images at the level of single cells is useful (but see Summery and Discussion).

Finally, unlike the columnar organizations in V1, we propose that the columnar organization may not cover entire cortex uniformly (Sato et al. 2009). In one hemisphere, we analyzed similarity in object selectivity of cells not from activity spots identified by IOSI but from arbitrary assigned local regions that have about the size of activity spots (Fig. 5.13d). If the columnar organization uniformly covers the cortex, the results would be the same as those from activity spots. However, we found that correlation in object selectivity among MU pairs greatly reduced. This observation suggests that in part of the cortex common property across the cells are not obvious than that in the activity spots.

5.8 Summary and Discussion

In this chapter, we have described object representation in IT cortex revealed by OISI: object images were represented by combinations of active and inactive feature columns, and these feature columns carry information about both local visual features as well as global features such as configuration of object parts. Though we showed representation of these two types of features in separate experiments, we do not intend to emphasize dichotomy between local and global features. It could be the case that some columns represent a visual feature that carries partly global and partly local information although such columns have not been explored yet. The visual features shown here, that is, protrusions, curvature, and rectangular shape for local features and the vertical alignment of parts and facial configuration for global features, are representatives that are easily described with a simple concept. As mentioned in an earlier part of the chapter, plasticity of IT cells make them adjust their response property to the new environment that monkeys react to. Thus, IT cells could tune to the visual features that are useful in certain behavior regardless of whether the represented features are local, global, or include both. Thus, in general, we may not easily describe visual features represented in IT cortex. For example, it would be difficult to characterize the visual features useful for Japanese to recognize Kanji characters at a glance and for Arabic to recognize Arabic characters at a glance.

Because of trial-by-trial variability in neuronal responses, we have discussed that object representation at the level of columns may have primary importance. Then, why do individual cells have a relatively high object tuning property that is different from cell to cell? One thought would be that variability in object responses among cells is necessary for shaping a response property essential for object representation at the columnar level. Alternatively, however, the object representation at the level of single cells may play an essential role in object recognition under certain circumstances. For example, when we inspect an object image, we gaze at the same portion of the image multiple times (Yarbus 1967).

Repeated fixations may play a role in making trial-by-trial averaging of responses at the level of single cells. Then, object recognition may have two stages: the first stage is to capture an object image at the level of feature columns by using ensemble averaging, and for the following careful inspection of the object image, we use representation at the level of single cells with trial-by-trial averaging through repeated fixations. At present, there is no evidence supporting this two stage model. At least, we need to investigate fundamental problems related to solving trial-by-trial variation in neuronal responses. In the case of ensemble averaging, it is essential whether the trail-by-trail variations are uncorrelated across neurons or not. On the other hand, in the case of solving trial-by-trial variations with repeated fixations, a certain kind of short term memory systems is required for holding sensory responses obtained in different time.

As described above, columnar organizations have a potential role in instantaneously and reliably detecting visual features at the columnar level. Thus, to maximize efficiency of plastic changes in IT cortex, plasticity may not only shape response property of individual cells depending on experience but also it may make clusters of these cells and generate new feature columns that are adjusted to the new environment. As described above, we suggest that columnar organization for visual features do not cover the entire region of IT, the rest of the region would be reserved area for generating new feature columns where neurons represent a common visual feature that is useful in certain behavior.

Since the results shown here were obtained from anesthetized monkeys, investigations of dynamic properties of object representation with behaving monkeys are required to relate object representation to recognition. For example, we have examined representation of a single object presented in the visual field. This experimental condition is very much far from natural conditions where many objects are presented simultaneously. Thus, one unsolved problem is how to deal with multiple objects in the visual field. One possible mechanism is proposed by Reynolds et al. (1999). They found that, regardless of whether the visual stimulus was presented in isolation or together with the other stimulus, visual responses of single V4 cell were the same when monkey pays attention to the stimulus. If this is the case, when multiple objects were presented to a monkey, spatial patterns of activity would change from one pattern to another depending on which object the monkey pays attention to. To address this kind of questions, investigations on representation by combinations of feature columns in behaving animals are required. So far in the literature, most of the investigations with behaving monkeys are based on physiological recordings from a single neuron at a time. Since the results presented here revealed that object representation is made of combinations of feature columns, we think that recordings from population activity particularly at the columnar level is essential. However, because of the slow time course of intrinsic signals (Fig. 5.1), OISI may not be the appropriate technique to address object recognition with behaving monkeys in which representation of object images supposed to be updated in a range of a few hundred milliseconds. Novel techniques with high temporal resolution as well as spatial resolution are required. Recently developed densely arranged multiple electrode arrays would be one of the such techniques (deCharms et al. 1999; Miyakawa et al. 2007).

References

Baylis GC, Rolls ET, Leonard CM (1985) Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. Brain Res 342(1):91–102

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

Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46(2):369–384

Das A, Gilbert CD (1995) Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. Nature 375:780–784

deCharms RC, Blake DT, Merzenich MM (1999) A multielectrode implant device for the cerebral cortex. J Neurosci Methods 93:27–35

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4(8):2051–2062

Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. Nature 360(6402):343–346

Fukuda M, Rajagopalan UM, Homma R, Matsumoto M, Nishizaki M, Tanifuji M (2005) Localization of activity-dependent changes in blood volume to submilimeter-scale functional domains in cat visual cortex. Cereb Cortex 15(6):823–833

Grinvald A, Shoham D, Shmuel A, Glaser D, Vanzetta I, Shtoyerman E, Slovin H, Wijnbergen C, Hildesheim R, Arieli A (1999) In-vivo optical imaging of cortical architecture and dynamics. In: Windhorst U and Johansson H (eds) Modern techniques in neuroscience research. Springer, Berlin Heidelberg New York pp. 893–970

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

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

Homma R, Tanifuji M (2003) Comparison of functional MAPs in macaque area TE revealed by *in vivo* optical imaging with voltage-sensitive dye and intrinsic signal imaging. Abstr. viewer/ Itinerary Planner. Society for Neuroscience, Washington, DC. 818.21

Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol 71(3):856–867

Kobatake E, Wang G, Tanaka K (1998) Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. J Neurophysiol 80(1):324–330

Kreiman G, Hung CP, Kraskov A, Quiroga RQ, Poggio T, DiCarlo JJ (2006) Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. Neuron 49:433–445

MacVicar BA, Hockman D (1991) Imaging of synaptically evoked intrinsic optical signals in hippocampal slices. J Neurosci 11:1458–1469

Miyakawa N, Vidal-Naquet M, Blake D, Merzenich M, Tanifuji M (2007). Activities from combination of columns in macaque area TE can encode object identity across viewing angles, Abstr. viewer/Itinerary Planner. Society for Neuroscience, San Diego. Online, 554.11.

Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res 47(3):329–342

Perrett DI, Oram MW, Harries MH, Bevan R, Hietanen JK, Benson PJ, Thomas S (1991) Viewer-centred and object-centred coding of heads in the macaque temporal cortex. Exp Brain Res 86(1):159–173

Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1984) Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. Hum Neurobiol 3(4):197–208

Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci 19:1736–1753

Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol 73:713–726

Sato T, Uchida G, Tanifuji M (2009) Cortical columnar organization is reconsidered in intorial tempral cortex. Coreb Corlex 19:1870–1880

Tamura H, Kaneko H, Fujita I (2005) Quantitative analysis of functional clustering of neurons in the macaque inferior temporal cortex. Neurosci Res 52:311–322

Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol 66(1):170–189

Tsunoda K, Oguchi Y, Hanazono G, Tanifuji M (2004) Mapping cone- and rod-induced retinal responsiveness in macaque retina by optical imaging. Invest Ophthalmol Vis Sci 45(10):3820–3826

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M (2001) Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat Neurosci 4(8):832–838

Vanzetta I, Grinvald A (1999) Increased cortical oxidative metabolism due to sensory stimulation: implications for functional brain imaging. Science 286(5444):1555–1558

Vanzetta I, Solvin H, Omer DB, Grinvald A (2004) Columnar resolution of blood volume and oximetry functional maps in the behaving monkey; implications for fMRI. Neuron 42(5):843–854

Wang G, Tanaka K, Tanifuji M (1996) Optical imaging of functional organization in the monkey inferotemporal cortex. Science 272(5268):1665–1668

Wang G, Tanifuji M, Tanaka K (1998) Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. Neurosci Res 32(1):33–46

Yarbus AL (1967) Eye movements and vision. Plenum Press, New York

Yamane S, Kaji S, Kawano K (1988) What facial features activate face neurons in the inferotemporal cortex of the monkey? Exp Brain Res 73(1):209–214

Yamane Y, Tsunoda K, Matsumoto M, Phillips AN, Tanifuji M (2006) Representation of the spatial relationship among object parts by neruons in macaque inferotemporal cortex. J Neurophysiol 96:3147–3156

Yen S-C, Baker J, Gray CM (2007) Heterogeneity in the Responses of adjacent neurons to natural stimuli in cat striate cortex. J Neurophysiol 97:1326–1341

Young MP, Yamane S (1992) Sparse population coding of faces in the inferotemporal cortex. Science 256(5061):1327–1331