

Columns for Complex Visual Object Features in the Inferotemporal Cortex: Clustering of Cells with Similar but Slightly Different Stimulus Selectivities

Keiji Tanaka

RIKEN Brain Research Institute, Wako-shi, Saitama 351-0198, Japan

Cells in the inferotemporal cortex (area TE) selectively respond to complex visual object features and those that respond to similar features cluster in a columnar region elongated vertical to the cortical surface. What are the functional roles of the column structure in the inferotemporal cortex? Selectivity of cells within a column is similar but not identical. If we emphasize the similarity among cells within a column, we can regard the columns as units for description of object features. The variety of stimulus selectivity in a column may work as a tool to disregard subtle changes in input images when the system is directed to invariant recognition. Alternatively, if we emphasize the differences in selectivity of cells within a column, the columns can be compared to differential amplifiers, each of which represents variety within a group of features. The enormous number of objects present in nature can be efficiently described by combining outputs of the multiple differential amplifiers in the inferotemporal cortex. The two modes may work in parallel, with a graded balance changing according to the behavioral context. Determining whether or not these hypotheses are valid will require further studies.

Introduction

Visual object recognition is a key function of the primate brain. It can be considerably view-invariant after exposure to multiple views of an object: we can identify the object regardless of considerable changes in input images due to changes in the viewing point and illumination condition. On the other hand, we can be highly sensitive to subtle differences in input images when we are recognizing objects at subordinate or individual levels. Thus, our visual system can either neglect or amplify differences in input images depending on the behavioral context. Columnar organization in the inferotemporal cortex may be crucial to mechanisms that satisfy these two apparently contradictory requirements.

Area TE of the inferotemporal cortex represents the final purely visual stage of the occipitotemporal pathway, which is thought to be essential for visual object recognition. The occipitotemporal pathway starts at the primary visual cortex (V1) and leads to TE after relays at V2, V4 and TEO. Although skipping projections also exist, such as those from V2 to TEO and those from V4 to the posterior part of TE, the step-by-step projections are more numerous. TE projects to various polymodal brain sites, including the perirhinal cortex, the prefrontal cortex, the amygdala and the striatum of the basal ganglia. The projections to these targets are more numerous from TE, particularly from the anterior part of TE, than from the areas at earlier stages (Ungerleider *et al.*, 1989; Yuki *et al.*, 1990; Barbas, 1992; Suzuki and Amaral, 1995). Therefore, there is a sequential cortical pathway from V1 to TE, and outputs from the pathway mainly originate from TE. In monkeys, bilateral TE ablation or complete deafferentation resulted in severe and selective deficits

in learning tasks that required visual recognition of objects (Gross, 1973; Dean, 1976; Yaginuma *et al.*, 1993).

Moderately Complex Features

An obstacle in the study of neuronal mechanisms of object vision has been the difficulty of determining the stimulus selectivity of individual cells. The variety of object features existing in the world is too great to test its entire range for a single cell while activity of the cell is being recorded. Although it is likely that the visual system scales down the variety for efficiency of representation, it remains to be determined how the brain scales down this variety. We have used an empirical reduction method that involves the use of a specially designed image-processing computer system (Fujita *et al.*, 1992; Ito *et al.*, 1994, 1995; Kobatake and Tanaka, 1994; Wang *et al.*, 1998). After spike activities from a single cell were isolated, many three-dimensional (3D) animal and plant models were first presented by hand within the animal's visual field to find the effective stimuli for the cell. Different aspects of the objects were presented in different orientations. Second, images of several most effective stimuli were taken with a video camera and displayed on a television monitor by a computer to determine the stimulus that evoked the maximal response. Finally, the image of the most effective stimulus was simplified step by step to determine which feature or combination of features contained in the image was essential for maximal activation. The minimal feature required for maximal activation was determined to be the critical feature for the cell. The magnitude of responses often increased as the complexity of an image was reduced. This may be due to the adjustment of size, orientation and shape, as well as the removal of other features, which may suppress the activation by the critical feature (Sato, 1989, 1995; Missal *et al.*, 1997; Tsunoda *et al.*, 2001).

Examples of the reduction in complexity of images for 12 TE cells are shown in Figure 1. The pictures to the left of the arrows are the original images of the most effective object stimuli and those to the right are the critical features determined after the reduction process. It should be noted that, even for the same object image, the directions of reduction and the final critical features were usually different from cell to cell. Some of the critical features were moderately complex shapes, while others were combinations of such shapes with color or texture. After determining the critical features for hundreds of cells in TE, we concluded that most cells in TE required moderately complex features for their maximal activation. The critical features for TE cells were more complex than just orientation, size, color or simple textures, which are known to be extracted and represented by cells in V1, but at the same time not sufficiently complex to represent the image of a natural object through the activity of single cells. The combined activation of multiple

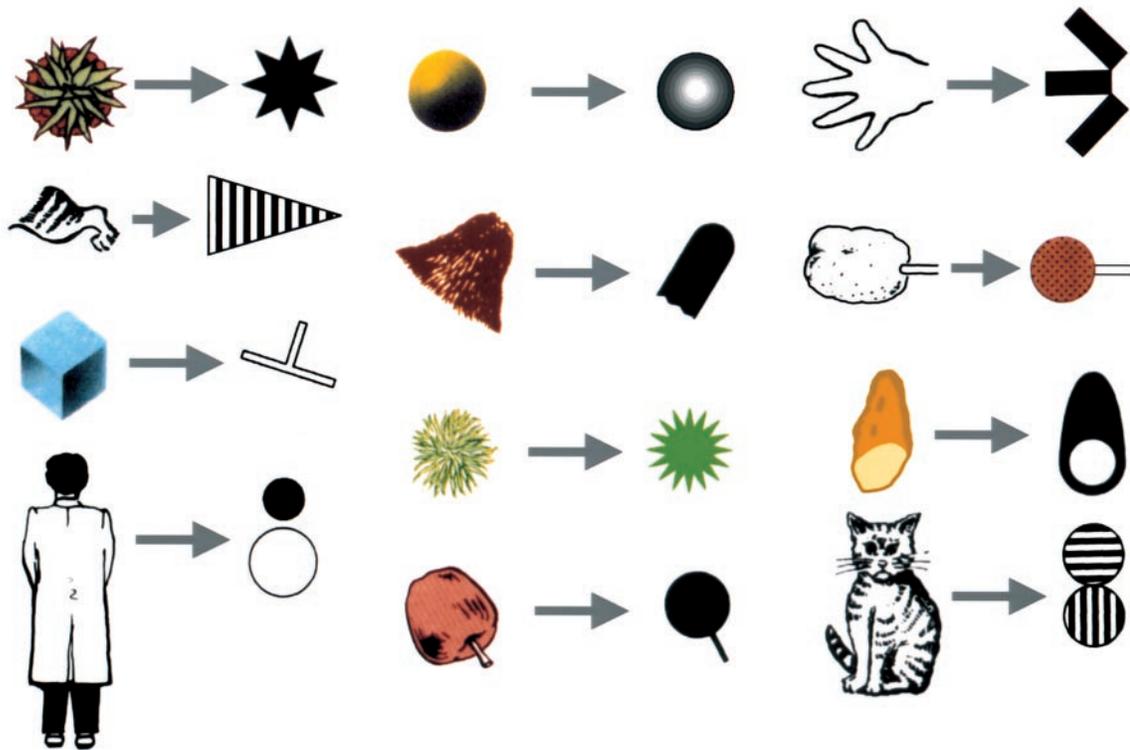


Figure 1. Examples of reductive determination of optimal features for 12 TE cells. The images to the left of the arrows represent the original images of the most effective object stimulus and those to the right of the arrows, the critical features determined by the reduction.

cells, which represent different features contained in the object image, is necessary.

This image reduction method has limitations. The initial survey of effective stimuli cannot cover the entire variety of objects existing in the world, so we may miss some very effective features. In addition, the tested methods of reducing the complexity of effective object images are limited by the available time of continuous recording from a single cell and also by the imagination of the experimenter. Because of these limitations, the objectivity of this method for determining optimal features has sometimes been questioned. However, alternative methods also have limitations. For example, some studies have used mathematically perfect sets of shapes (Schwartz *et al.*, 1983; Richmond *et al.*, 1987; Gallant *et al.*, 1993, 1996). However, the generality of these sets would hold only if the system were linear, which is hardly expected in higher visual centers. Others (Pasupathy and Connor, 2001) have developed a method of presenting a large number of shapes made by combining several arcs of different curvatures. They have shown the usefulness of this method in studying the selectivity of V4 cells, but it may not be useful for TE cells, which respond to more complicated shapes than V4 cells. Yet others (Keyser *et al.*, 2001) have developed a method of analyzing responses to >1000 stimulus images in a fixation task. The stimulus images were each presented for a short time (e.g. 100 ms) without an interstimulus interval. The following stimulus presentation may inhibit the response to the previous stimulus, but, because the order of stimulus presentation is randomized and because TE cells tend to respond to a small part of the stimuli, there are no inhibitory interactions in the majority of repetitions. These two methods may be combined to explore systematically a large feature space

of complex shapes, sufficiently complex for the activation of most TE cells.

Faces and Other Extensively Learned Objects

Although the critical features for the activation of TE cells are only moderately complex in general, there are cells that respond to faces and critically require nearly all the essential features of the face. Such cells were originally found deep in the superior temporal sulcus (Bruce *et al.*, 1981; Perrett *et al.*, 1982), but they have also been found in TE (Baylis *et al.*, 1987). Thus, there is more convergence of information to single cells for representations of faces than for those of non-face objects. This difference may arise because discrimination of faces from other objects is not the final goal of face processing (since further processing of facial images is needed to discriminate among individuals and expressions), while distinguishing a non-face object from other objects may be close to the final goal of object processing.

There are suggestions that responses to whole objects will develop in TE if the subject is extensively trained in fine discrimination of similar objects. Logothetis and colleagues (Logothetis *et al.*, 1995) trained adult monkeys to recognize wire-frame objects against many other similar wire-frame objects and recorded from cells in TE of the monkeys during the same task. About 20% of cells responded to wire-frame objects more strongly than to any other tested objects. Some of the neurons responded to parts of the objects as well as to the entire images of the objects, while others did not respond to parts of the objects (Logothetis, 1998). Based on these results, it was proposed (Logothetis, 1998) that some TE cells respond to whole objects which the subjects have used to conduct fine

discriminations, while a majority of TE cells respond to features present in images of multiple different objects. However, this remains to be further studied, because the examination of selectivity described for object parts was rather preliminary (Logothetis, 1998).

Depth Structure

The critical features for the activation of TE cells included the gradient of luminosity (e.g. top middle in Fig. 1). The gradient of luminosity often provides depth structure of object surfaces with an assumption of the direction of illumination. In this sense, features represented by TE cells are not necessarily purely two-dimensional (2D); that is, they may be features that can be described in 2D space but reflect depth structures. Moreover, recent studies have found that some TE cells selectively respond to horizontal disparity in addition to the 2D shape of stimuli.

The horizontal disparity between images projected to the left and right eyes is a strong cue for perception of depth. Although it was once assumed that the selectivity for disparity is more predominant in the occipitoparietal (or dorsal visual) pathway, which is responsible for visuomotor control or spatial vision, than the occipitotemporal (or ventral visual) pathway, recent studies have shown that many cells in TE are selective to the disparity of stimuli, as well as their 2D shapes in the frontoparallel plane.

Uka *et al.* (Uka *et al.*, 2000) recorded from TE cells in monkeys performing a fixation task and examined their responses to 2D shape stimuli presented at different depths. The depth was defined relative to that of the fixation point, as in other such experiments. Cells that responded to at least one of the 11 2D shapes at zero disparity were examined for disparity selectivity. Responses of more than one-half (63%) of the cells showed statistically significant dependency on disparity. Most of the disparity-selective cells were either 'near' or 'far' neurons according to the classification of Poggio and Fischer (Poggio and Fischer, 1977). This is in contrast to the primary visual cortex and area MT, in which tuned excitatory cells constitute a large part (2/3 in V1 and 2/5 in MT) of the disparity-selective cells (Poggio and Fischer, 1977; Maunsell and Van Essen, 1983; Cumming and DeAngelis, 2001).

The stimuli used by Uka *et al.* (Uka *et al.*, 2000) were flat in the depth direction, i.e. there were no depth structures within their contours. Many objects in nature have surfaces tilted or curved in the depth direction and such depth gradient of the surface is an important feature of the object image. Janssen *et al.* (Janssen *et al.*, 1999, 2000a,b) used a stimulus set composed of stimuli having several different depth profiles in combination with several different 2D shapes. About one-half of the cells recorded from the ventral bank of the anterior part of the superior temporal sulcus exhibited selectivity for depth profile. Some of them responded to a linear gradient of depth, some to a combination of opposite linear gradients (or wedge profile) and others to a smooth concave or a convex depth curvature. They were selective for both 2D shape and depth profile. The selectivity for the depth profile was not explained by the selectivity for the depth position of a particular part of the stimulus, because the stimuli of the opposite depth profile did not activate the cells at any depth. The proportion of such cells was much lower (~10%) in the ventrolateral surface (i.e. area TE) than in the ventral bank of the superior temporal sulcus.

Because previous cytoarchitectural studies distinguished the ventral bank of the anterior part of the superior temporal sulcus (TEa and TEm) from the ventrolateral surface (TE) (Seltzer and Pandya, 1978), we have to consider the possibility that the two regions are functionally differentiated. However, H. Tanaka and

I. Fujita (personal communication) found that cells in the ventral bank were as selective for complex 2D shapes as cells in TE. Moreover, the cells in the ventral bank were much more sensitive to the direction of the disparity gradient or curvature, e.g. concave versus convex, than the quantitative values of curvature or gradient (Janssen *et al.*, 2000b). Therefore, the responses of cells in the ventral bank of the superior temporal sulcus do not represent a full reconstruction of the 3D structure of the objects. Rather, it may be the case that the representation there is still mainly 2D and the qualitative information of disparity gradient or curvature just makes the 2D representation richer. It should be also noted that the representation of 2D shapes in TE may also not be fully quantitative. The number of features represented in TE may be limited by the number of TE columns (see the section entitled 'Columnar Organization in TE') and the invariance of responses of TE cells to certain types of shape deformations makes it difficult to reconstruct the input images from responses of TE cells (see next section). Only the features useful for discrimination of objects may be selectively represented in TE.

Invariance of Responses

Our object recognition ability is retained even when objects are translated in various ways. These invariances can, in part, be explained by invariant properties of single-cell responses in TE. Using a set of shape stimuli composed of individually determined critical features and several other shape stimuli obtained by modifying the critical features, we have observed that selectivity for shape is preserved across TE receptive fields (Ito *et al.*, 1995), which usually range from 10 to 30° in a one-dimensional size. However, the maximum response is usually obtained around the geometrical center of the receptive field and the magnitude of response decreases toward the edges of the receptive field (Ito *et al.*, 1995; Op de Beeck and Vogels, 2000). Moreover, the receptive-field centers of TE cells are scattered around the fovea (Kobatake and Tanaka, 1994; Op de Beeck and Vogels, 2000). Therefore, responses of TE cells carry information about the position of stimuli as well as detailed information about their shape, color and texture.

The effects of changes in stimulus size varied among cells (Tanaka *et al.*, 1991; Ito *et al.*, 1995). Twenty-one percent of the TE cells tested responded to a size range of more than four octaves of the critical features with >50% maximum responses, whereas 43% responded to a size range of less than two octaves. TE cells with considerable invariance for the location and size of stimuli have also been found by Lueschow *et al.* and Logothetis *et al.* (Lueschow *et al.*, 1994; Logothetis *et al.*, 1995). The tuned cells may only function in the process of making invariant responses: those responding to various sizes of the same shape converge to a target cell to yield the size-invariant responses with sharp shape selectivity. Alternatively, both size-dependent and -independent processing of images may occur in TE.

A number of TE cells tolerated reversal of the contrast polarity of the shapes. Contrast reversal of the critical feature evoked >50% of the maximum responses in 40% of tested cells (Ito *et al.*, 1994). Other workers (Sary *et al.*, 1993) found that some TE cells responded similarly to shapes defined by differences in luminosity, direction of motion of texture components and the coarseness of texture, while maintaining their selectivity for shape. Tanaka *et al.* (Tanaka *et al.*, 2001) found that about a quarter of TE cells responded similarly to shapes defined by difference in horizontal disparity of texture components, to those defined by differences in size of texture components and to those defined by differences in luminosity.

Another kind of invariance of TE cells was found with regards

to the aspect ratio of shapes. The aspect ratio is the ratio of the size along one axis of the stimulus to that along the orthogonal axis. When an object rotates in depth, the features contained in the image change their shapes. Unless occlusion occurs, changes occur in the aspect ratio. For individual TE cells, we first determined the critical feature using the reduction method and then tested the effects of changes in the aspect ratio of the critical feature. We observed that 51% of cells responded to an aspect ratio range of more than three octaves with >50% of the maximum responses (Hossein and Tanaka, 1998).

In Figure 1 and our previous studies, we drew the features determined to be critical for the activation of individual TE cells as 2D images. However, this was for the sake of description and it does not necessarily mean that the cells were tuned to 2D images. Selectivity can only be defined in terms of a list of tested stimulus deformations and their associated response reductions. The above-described invariances of TE cells suggest that they are actually more sensitive to certain types of deformations than others. The types of deformations that often occur when an object moves around appear to be more tolerated. A related discussion has been presented elsewhere (Vogels *et al.*, 2001).

Columnar Organization in TE

We examined the spatial distribution of the cells responding to various critical features in TE. By recording two TE cells simultaneously with a single electrode, we found that cells located close together in the cortex had similar stimulus selectivities (Fujita *et al.*, 1992). The critical feature of one isolated cell was determined using the same procedure as described above, while the responses of another isolated cell, or non-isolated multiunits, were simultaneously recorded. In most cases, the second cell responded to the optimal and suboptimal stimuli of the first cell. The selectivities of the two cells differed slightly, however, in that the maximal response was evoked by slightly different stimuli, or the mode of the decrease in response was different when the stimulus was changed from the optimal stimulus.

To determine the spatial extent of the clustering of cells with similar selectivities, we examined the responses of cells recorded successively along long penetrations vertical or oblique to the cortical surface (Fujita *et al.*, 1992). The critical feature for a cell located at the middle of the penetration was first determined. A set of stimuli was then constructed, including the critical feature for the first cell, its rotated versions and ineffective control stimuli; cells recorded at different positions along the penetration were then tested with the fixed set of stimuli. As in the example shown in Figure 2, cells recorded along the vertical penetrations commonly responded to the critical feature for the first cell or some related stimuli. The commonly responsive cells spanned nearly the entire cortical thickness from layers 2 to 6. In the case of penetrations that were made oblique to the cortical surface, however, the cells that were commonly responsive to the critical feature of the first cell or related stimuli were limited to within a short span around the first cell. The horizontal extent of the span was, on average, 400 μm . Cells outside the span did not respond to any of the stimuli included in the set, or responded to some stimuli that were not effective in activating the first cell and were included in the set as ineffective control stimuli. Based on these results, we proposed that TE is composed of columnar modules, in each of which cells respond to similar features (Fig. 3).

It should be noted that precise determination of the optimal features is essential to observe the similarity of stimulus selectivities between neighboring cells clustered in a columnar region. Several studies, which used a fixed set of arbitrarily

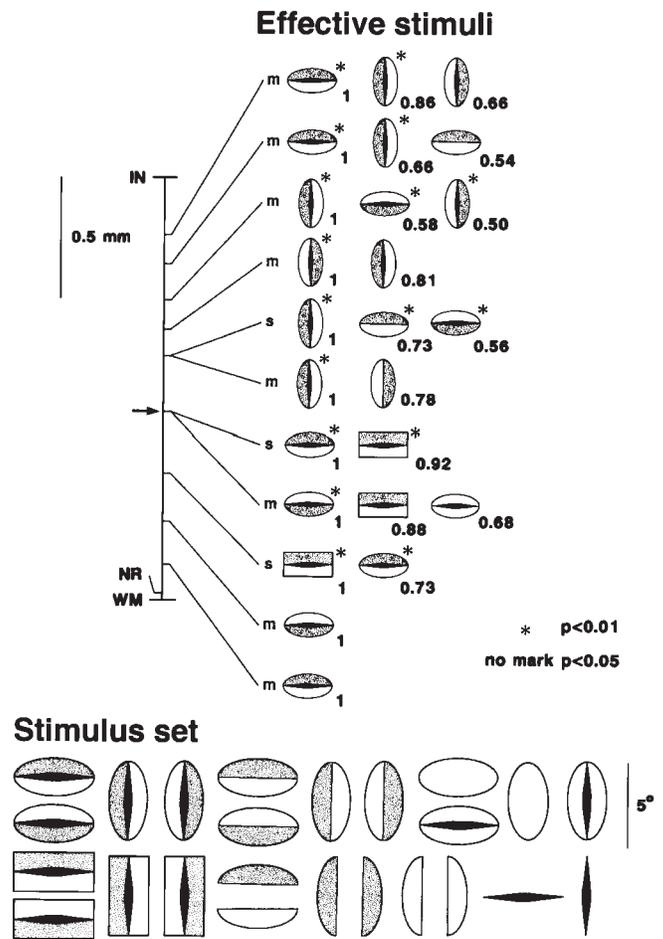


Figure 2. Responses of cells recorded along a vertical penetration in TE. The responsiveness of the cells was tested with the set of stimuli shown at the bottom, which were constructed with reference to the critical feature of the first cell indicated by the arrow. Effective stimuli are listed separately for individual recording sites, in the order of effectiveness. 'm' indicates recording from multiunits and 's' from a single unit. Other abbreviations: IN, initial neuron; NR, multiunits with no responses; WM, white matter. Cited from Fujita *et al.* (Fujita *et al.*, 1992).

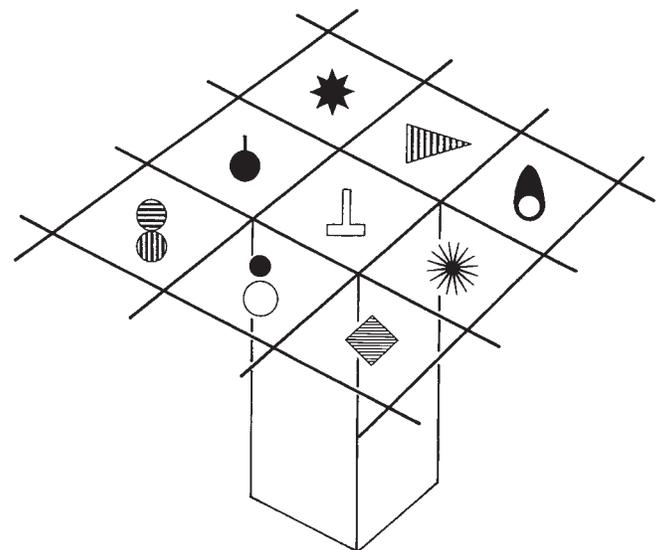


Figure 3. Schematic drawing of the columnar organization in TE.

selected object images, failed to find the similarity. The optimal features for the activation of TE cells are complex and defined by many dimensions. The preference of cells within a column is similar in some dimensions, but different in other dimensions. For example, cells in a column respond to star-like shapes, or shapes with multiple protrusions. They are similar in that they respond to star-like shapes, but they may differ in the preferred number of protrusions or the amplitude of the protrusions. Therefore, if only a fixed set of object images is used, cells within the column may respond to different objects, because star-like shapes with different numbers of protrusions appear in different objects. The same is true for the primary visual cortex. Cells within an orientation column share the preferred orientation, while they differ in the preferred width and length of stimuli, binocular disparity and the sign of contrast. If a set of stimuli that vary not only in orientation but also in all other parameters is used, cells within an orientation column will not show clear similarity in selectivity.

Spatial Arrangement of Columns

To study further the spatial properties of the columnar organization in TE, we used optical imaging with intrinsic signals (Wang *et al.*, 1996, 1998). In optical imaging with intrinsic signals, the region of the cortex with elevated neuronal activities appears darker than other regions in the reflected image. We first recorded the responses of single cells with a microelectrode to determine the critical feature and then conducted optical imaging. In the experiment shown in Figure 4, the critical feature determined for a cell recorded at the cortical site indicated by a cross was the combination of white and black horizontal bars. The PST histograms on the left represent the responses of the cell. The combination evoked a strong response in the cell, but a white bar alone or a black bar alone did not

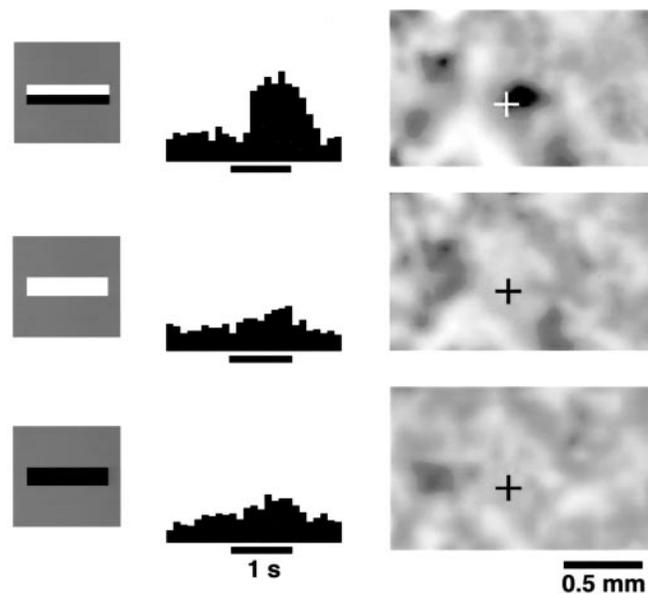


Figure 4. Correspondence of optical signals with neuronal activity. The histograms on the left show the responses of a cell recorded at the site indicated by the crosses in the optical images. The cell selectively responded to the combination of a white bar and black bar. The white bar alone or a black bar alone evoked much smaller responses. Correspondingly, the optical image showed a black spot covering the recording site during the time the monkey was viewing the combination shape, whereas there were no dark spots for the control stimuli. Cited from Wang *et al.* (Wang *et al.*, 1996).

activate the cell. The images on the right were taken from the same 1×1.5 mm cortical region. A dark spot appeared around the penetration site when the monkey saw the combination of the two bars, whereas there were no dark spots around the site when the monkey saw the simpler features. Similar results were obtained in 11 out of 13 cases. Tsunoda *et al.* (Tsunoda *et al.*, 2001) further confirmed the correlation of optical signals with neuronal responses in TE. Although the critical feature was determined for a single cell, a large proportion of cells in the region must be activated to produce an observable metabolic change. Therefore, the localized and specific occurrence of dark spots indicates a regional clustering of cells with similar stimulus selectivities. In the small number of cases in which the correlation was not found – for example, 2 out of 13 cases in Wang *et al.* (Wang *et al.*, 1998) – the cell for which the critical feature was determined might be located at an eccentric position in the range of selectivity variety within the column and the stimulus may have activated only a small proportion of cells in the column.

However, when we observed a larger area of the cortical surface, we found that the presentation of a single feature activated multiple spots. In Figure 5, the spots activated by eight moderately complex features are indicated by different kinds of lines and superimposed, i.e. spots activated by one set of four features are shown in the upper half and those by another set of four features in the lower half. For example, feature 1 evoked six spots and feature 2 evoked two spots. This example demon-

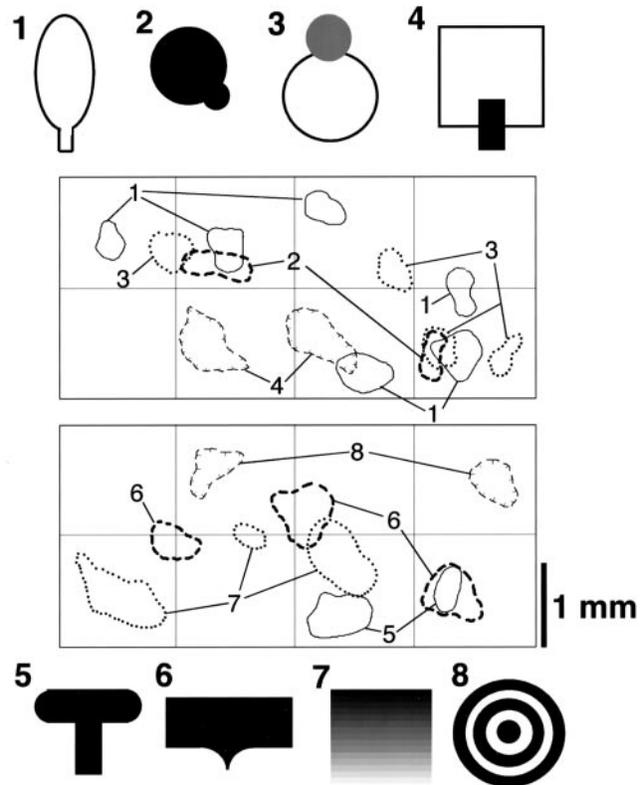


Figure 5. Map of activation evoked by the presentation of eight moderately complex features. To obtain the map, each image was subtracted by the reference image averaged over the images obtained for all the different stimuli combined in the experiment to remove the global darkening. The activation spots were delineated at $1/e$ of the maximum intensity in individual images and the contours of spots in the images for different stimuli are indicated by different line types.

strates that a single feature is processed in multiple columns in TE.

Another interesting observation here is the partial overlaps between the activation spots evoked by different features. Some of the overlapping regions, which were activated by many stimuli, likely represent columns of non-selective cells. However, others that were activated by only two of the stimuli may represent specific overlaps. For many of these overlaps, we can find similarity between the two features, although the judgment of similarity is only subjective.

The partial overlapping of columns responding to different but related features was most clearly observed for faces presented in different views (Fig. 6). This experiment was also guided by a unit-recording experiment. We recorded five cells in one electrode penetration around the center of the imaged region; all of these cells selectively responded to faces. Three of them responded maximally to the front view of the face, whereas the remaining two responded to the profile, i.e. the lateral view of the face. In an optical imaging session, five different views of the face of the same doll were presented in combination with 14 non-face features. All of the faces evoked activation spots around the center of the illustrated 3×3 mm region. However, their center positions were slightly different. The contours of the dark spots are superimposed at the bottom. The activation spot moved in one direction as the face was rotated from the left profile to the right profile through the front view of the face. Individual spots were 0.4–0.8 mm in diameter and the overall region was 1.5 mm. These regions were not activated by the 14 non-face features. Similar results, namely selective activation by faces and systematic shift of the activation spot with the rotation of the face, were obtained for three other monkeys. In these three monkeys, optical imaging was not guided by unit recording. The recording chamber, with an inner diameter of 18

mm, was placed in the same part of TE and the face-selective activation was found at approximately the same location (approximately the posterior third of TE on the lateral surface, close to the lip of the superior temporal sulcus).

The effects of rotating the face around a different axis (the chin-up and -down) and of changing the facial expression were also determined in some of the experiments, but neither of these caused a shift in the activation spot. Only two faces were tested: a human face and a doll's face. The two faces activated regions that mostly overlapped. There are two possible interpretations of this result. One is that the variations other than those with horizontal rotation are represented at different sites not covered by the recording chamber in the experiments. Alternatively, it is possible that only the variations along the horizontal rotation are explicitly mapped along the cortical surface and other variations are imbedded in overlapping cell populations.

The data for the non-face features are fewer, but I hypothesize that there are similar structures for representing non-face objects and I propose a modified model of the columnar organization of neurons in TE as shown in Figure 7. The borders between neighboring columns are not necessarily distinct. Instead, multiple columns that represent different but related features partially overlap with one another and as a whole compose a larger-scale unit. At least in some cases, some parameter of the features is continuously mapped along the cortical surface.

The systematic arrangement of related columns could be used for various computations necessary for object recognition. For example, object generalization might be mediated by horizontal excitatory connections between nearby columns representing related features, to achieve a selective blurring of activation. In addition, object discrimination might be achieved through mutual inhibition among nearby columns for winner-take-all selection. The continuous mapping of different views of faces

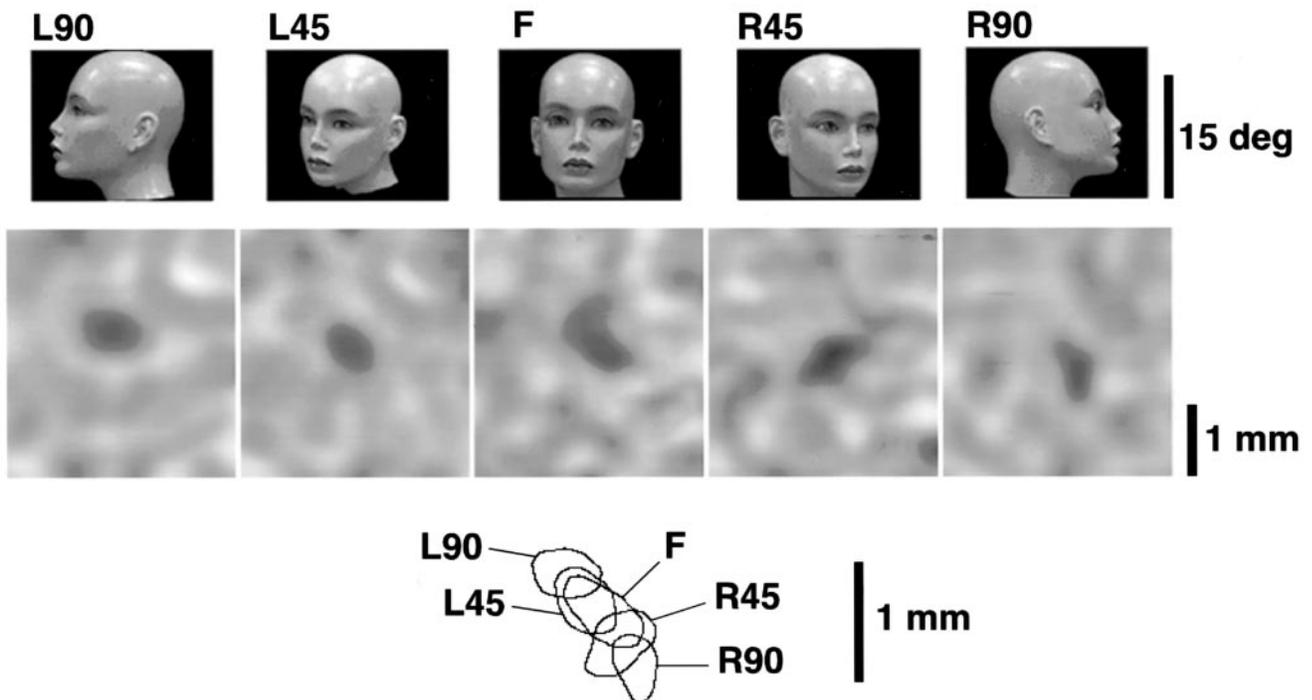


Figure 6. Systematic movement of the activation spot with rotation of the face. The images were obtained for five different views of the face of the same doll shown on top. The reference image obtained by averaging the five images was subtracted. The contours circumscribing the pixels with t -values at $P < 0.05$, compared with the reference image, are superimposed at the bottom. Cited from Wang *et al.* (Wang *et al.*, 1996).

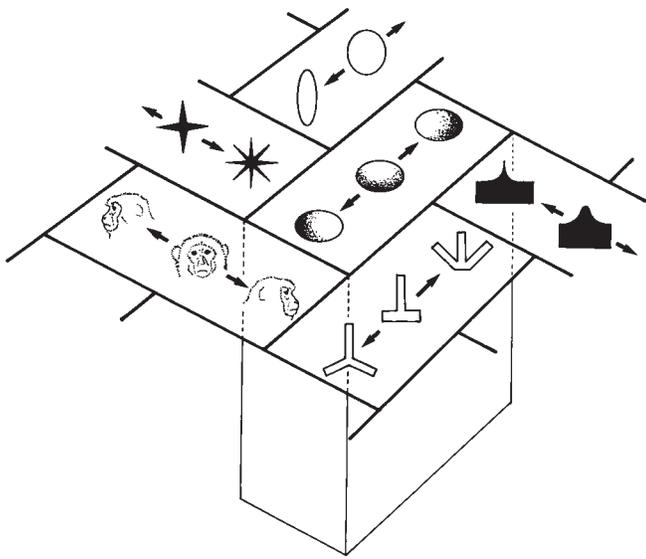


Figure 7. Revised schematic diagram of the columnar organization in TE.

cannot be generalized to non-face objects. Because the critical features for TE cells are only moderately complex except for faces, the image of a non-face object has to be represented by a combination of activations at multiple cortical sites. Rotation of a non-face object causes shifts of activation at multiple cortical sites, each of which corresponds to the partial change of a feature. The parameters along which the activation moves in non-face columns should be examined further to uncover the functional architecture of TE.

Representation of Features and of Objects

Since most inferotemporal cells represent features of object images but not the whole object images, the representation of the image of an object requires a combination of multiple cells representing different features contained in the image of the object. This process of combination presents unique scientific problems. Objects often appear in a clutter. A part of features belonging to one object may be mistakenly combined with a part of features belonging to another object. This erroneous combination causes a false perception of an object that is not visually present. How does the brain avoid such an erroneous combination?

Previously, the synchronization of spiking activity between cells was proposed as the mechanism for binding the features belonging to one object. Some experiments found a correspondence between cortical spike synchronization and perception of object borders (Singer, 1999), while others did not (Lamme and Spekreijse, 1998). Another possible means of avoiding erroneous feature combination is to have features partially overlapping with one another (Mel and Fiser, 2000). Suppose we are to represent four-letter strings. There will be an erroneous combination if we use only representation units coding single letters (e.g. ABCD is not discriminated from BADC, CDAB and so on, if units code A, B and C), while there will be no erroneous combinations if we use units specifying two consecutive letters and those specifying letters at the top and end of three consecutive letters (e.g. ABCD is the only four-letter string that contains AB, CD and A_C). The spatial relation between the units does not need to be represented.

Tsunoda *et al.* (Tsunoda *et al.*, 2001) compared activation of

the inferotemporal cortex by object images and activation by features included in the object images using a combination of optical imaging and single-cell recordings. The image of an object usually activated several spots within the imaged region (6×8 mm) and a feature contained in the object image activated a subset of the spots, as in the case shown in Figure 8A. This result was consistent with the idea that different spots were activated by different features contained in the object image. However, activation by a feature often included new spots that had not been activated by the whole-object image, as illustrated in Figure 8B. Single-cell recordings revealed that cells within such spots were activated by one feature while inhibited by another feature included in the object image. Previous single-cell recording studies had also shown that the response of inferotemporal cells to the optimal stimulus was suppressed by the simultaneous presentation of a second stimulus (Sato, 1989, 1995; Missal *et al.*, 1997, 1999). These results indicate that the stimulus selectivity of inferotemporal columns should be described by both the simplest feature for maximum activation and the features that suppress activation. Even with the same optimal feature for excitation, the range of features that suppresses excitation can vary from column to column and probably also from cell to cell. This complexity of the overall stimulus selectivity of inferotemporal columns and cells may help to reduce the chance of erroneous detection of non-existing objects.

Another study (Yamane *et al.*, 2001) also used a combination of the optical imaging and single-cell recordings, and found that some of the columns activated by an object image were activated, not by local features, but by a global feature of the object image. These columns were more sensitive to the global arrangement of object parts than to the properties of the parts. For example, one column responded to two vertically aligned black parts, regardless of the shape of either part. These columns representing global features could also help to reduce the possibility of erroneous detection of non-existing objects.

Intrinsic Horizontal Connections within TE

Intrinsic horizontal connections span up to 8 mm in TE. The projection terminals are more or less continuously distributed within 1 mm of the cells of origin, whereas they are clustered in patches in more distant regions (Fujita and Fujita, 1996; Tanigawa *et al.*, 1998). The cells of origin of these horizontal connections contain inhibitory neurons within 1 mm, but they are exclusively composed of excitatory cells (mostly pyramidal cells) for longer connections (Tanigawa *et al.*, 1998). Iontophoretic injection of bicuculline methiodide, an antagonist of the inhibitory synaptic transmitter γ -aminobutyric acid (GABA), reduced the stimulus selectivity of TE cells; in particular, the stimuli optimal for nearby cells turned out to evoke excitatory responses during the blockage of inhibition (Wang *et al.*, 2000). Inhibitory components of horizontal connections contribute to the formation of stimulus selectivity. The functional roles of the excitatory components are not known. It is possible that they connect columns responding to similar features, as is the case in the primary visual cortex (Gilbert and Wiesel, 1989). The combination of optical imaging and anatomical tracing methods (Tanifuji *et al.*, 2001) will provide insights into this issue.

Functions of TE Columns

Representation by multiple cells in a columnar module, in which the precise selectivity varies from cell to cell while selectivities for most effective stimuli largely overlap, can satisfy two apparently conflicting requirements in visual recognition: disregard-

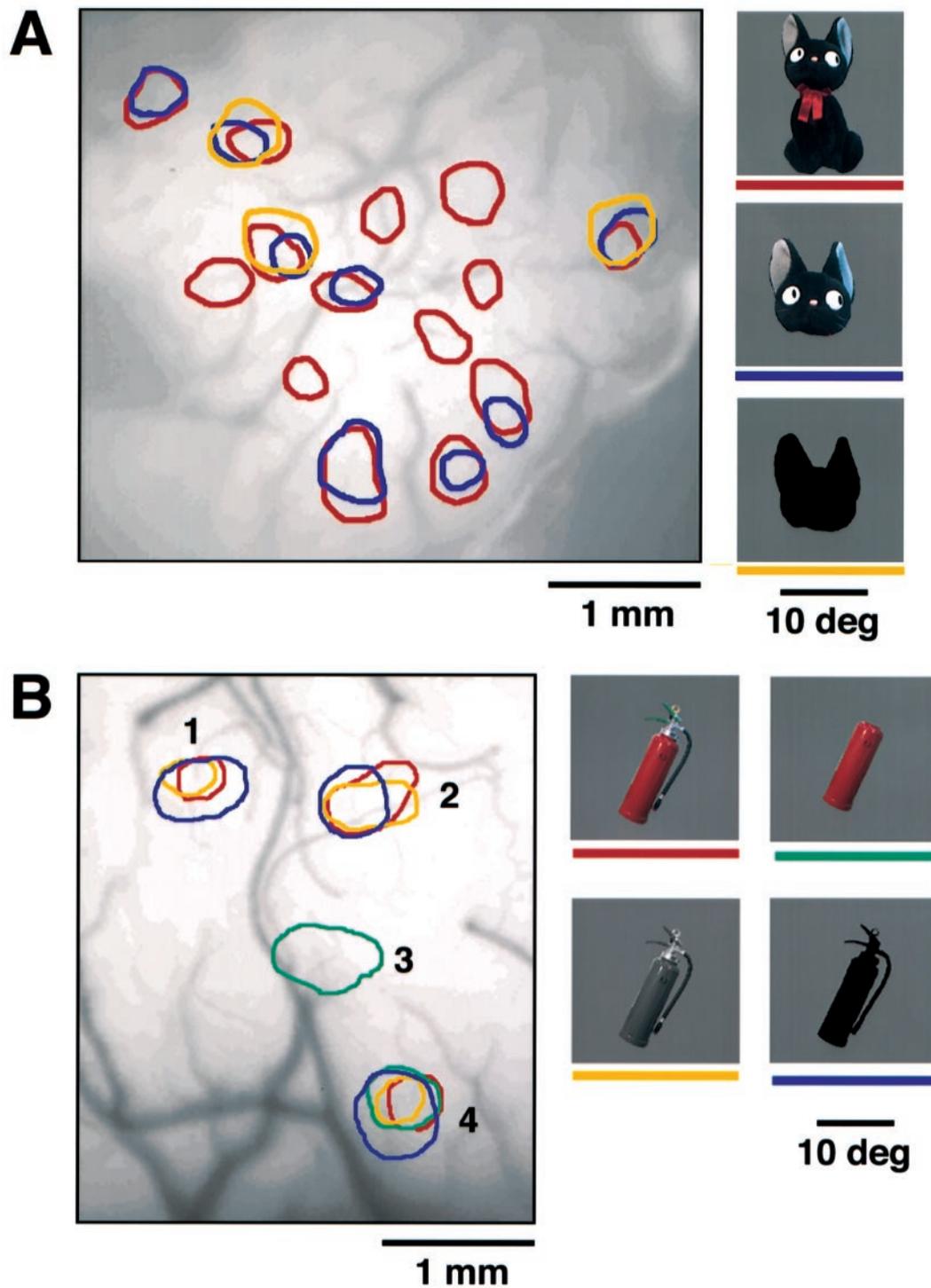


Figure 8. Activations by object images and partial features. The outline of the activation spots was determined at one-half of the maximal signal in an individual spot. Only the spots including pixels with significant activation during the stimulation compared with the blank period are shown. (A) The head activated a partial set of the spots activated by the cat and the silhouette of the head activated a partial set of the spots activated by the complete head. (B) The color and gray image and silhouette of a fire extinguisher activated three spots. The body of the fire extinguisher activated a new spot, as well as two of the three spots activated by the fire extinguisher. Single-cell recordings conducted after the optical imaging showed that Spot 1 was activated by protrusions, Spot 2 by a curved line and Spots 3 and 4 by a rectangular shape. Cells in Spot 3 were inhibited by a curved line, but those in Spot 4 were not. Cited from Tsunoda *et al.* (Tsunoda *et al.*, 2001).

ing subtle changes in input images under different viewing conditions; and achieving a preciseness of representation in discrimination of objects in subordinate or individual levels.

Clusters of cells having overlapping and slightly differing

selectivities may work together to confer object recognition abilities that are invariant to viewing conditions. Although single cells in TE tolerate some changes in size, contrast polarity and aspect ratio, these invariant properties at the single-cell level are

not sufficient to explain the entire range of flexibility of object recognition. In particular, the responses of TE cells are generally selective for the orientation of the shape in the frontoparallel plane. Cells preferring different orientations and other parameters of the same 3D shape may be combined in a column to provide invariant outputs. Whether signals from these selective cells converge to a group of single cells that show invariant responses is a matter for further investigation. One possibility is that outputs of cells preferring different orientations, sizes, aspect ratios and contrast polarities of the same shape overlap in the target structure, thereby evoking the same effects. An anatomical study with an injection of anterograde tracer into a focal site in TE suggested that projections from TE to the ventrocaudal striatum of the basal ganglia exhibit this property (Cheng *et al.*, 1997). Another possibility is that activation of cells is transmitted to other cells within a column and to nearby columns that represent related features through horizontal excitatory connections, in the presence of top-down signals from other brain sites, the prefrontal cortex for example. Multiple, but a limited number of, ways of activation transmission are hardwired in the network within a column and the arrangement of columns at nearby positions, and the top-down signals select one from them according to the behavioral context. Clusters of cells having overlapping and slightly differing selectivities may also serve to extract common features, but disregard differences between individual members in a category of objects when the system is directed to categorical object recognition.

Representation by multiple cells with overlapping selectivities can be more precise than a mere summation of representations by individual cells. A subtle change in a particular feature, which does not markedly change the activity of individual cells, can be coded by the differences in the activities of cells with overlapping and slightly different selectivities. Projections from the ventroanterior part of TE to the perirhinal cortex extensively diverge (Saleem and Tanaka, 1996). Projection terminals from a single site of ventroanterior TE cover ~50% of the perirhinal cortex. This divergence in projections may distribute the subtle differences over a larger area of the perirhinal cortex, so that objects recognized at individual levels can be distinctively associated with other kinds of information. The subtle differences can also be emphasized by mutual inhibition between cells or nearby columns for winner-take-all-type selection. The inhibition may also be under the top-down control.

Notes

Address correspondence to Keiji Tanaka, RIKEN Brain Research Institute, Wako-shi, Saitama 351-0198, Japan. Email: keiji@postman.riken.go.jp.

References

Barbas H (1992) Architecture and cortical connections of the prefrontal cortex in the rhesus monkey. In: *Advances in neurology* (Chauvel P, *et al.*, eds), vol. 57, pp. 91–115. New York: Raven Press.

Baylis GC, Rolls ET, Leonard CM (1987) Functional subdivisions of the temporal lobe neocortex. *J Neurosci* 7:330–342.

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

Cheng K, Saleem KS, Tanaka K (1997) Organization of corticostriatal and corticoamygdalar projections arising from the anterior inferotemporal area TE of the macaque monkey: a *Phaseolus vulgaris* leucoagglutinin study. *J Neurosci* 17:7902–7925.

Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. *Annu Rev Neurosci* 24:203–238.

Dean P (1976) Effects of inferotemporal lesions on the behavior of monkeys. *Psychol Bull* 83:41–71.

Fujita I, Fujita T (1996) Intrinsic connections in the macaque inferior temporal cortex. *J Comp Neurol* 368:467–486.

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

Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259:100–103.

Gallant JL, Connor CE, Rakshit S, Lewis JW, Van Essen DC (1996) Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76:2718–2739.

Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J Neurosci* 9:2432–2442.

Gross CG (1973) Visual functions of inferotemporal cortex. In: *Handbook of sensory physiology* (Jung R, ed.), vol. 7, part 3B, pp. 451–482. Berlin: Springer Verlag.

Hossein E, Tanaka K (1998) Effects of changes in aspect ratio of stimulus shape on responses of cells in the monkey inferotemporal cortex. *Soc Neurosci Abstr* 24:899.

Ito M, Fujita I, Tamura H, Tanaka K (1994) Processing of contrast polarity of visual images in inferotemporal cortex of the macaque monkey. *Cereb Cortex* 5:499–508.

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

Janssen P, Vogels R, Orban GA (1999) Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proc Natl Acad Sci USA* 96:8217–8222.

Janssen P, Vogels R, Orban GA (2000a) Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* 288:2054–2056.

Janssen P, Vogels R, Orban GA (2000b) Three-dimensional shape coding in inferior temporal cortex. *Neuron* 27:385–397.

Keysers C, Xiao D-K, Foldiak P, Perrett DI (2001) The speed of sight. *J Cognit Neurosci* 13:90–101.

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

Lamme VA, Spekreijse H (1998) Neuronal synchrony does not represent texture segregation. *Nature* 396:362–366.

Logothetis NK (1998) Object vision and visual awareness. *Curr Opin Neurobiol* 8:536–544.

Logothetis NK, Pauls J, Poggio T (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5:552–563.

Lueschow A, Miller EK, Desimone R (1994) Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex* 5:523–531.

Maunsell JHR, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol* 49:1148–1167.

Mel BW, Fiser J (2000) Minimizing binding errors using learned conjunctive features. *Neural Comput* 12:731–762.

Missal M, Vogels R, Orban GA (1997) Responses of macaque inferior temporal neurons to overlapping shapes. *Cereb Cortex* 7:758–767.

Missal M, Vogels R, Chao-yi L, Orban GA (1999) Shape interactions in macaque inferior temporal neurons. *J Neurophysiol* 82:131–142.

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

Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86:2505–2519.

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

Poggio GF, Fischer B (1977) Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurophysiol* 40:1392–1405.

Richmond BJ, Optican LM, Podell M, Spitzer H (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J Neurophysiol* 57:132–146.

Saleem KS, Tanaka K (1996) Divergent projections from the anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey. *J Neurosci* 16:4757–4775.

Sato T (1989) Interactions of visual stimuli in the receptive fields of

- inferior temporal neurons in awake macaques. *Exp Brain Res* 77:23–30.
- Sato T (1995) Interactions between two different visual stimuli in the receptive fields of inferior temporal neurons in macaques during matching behaviors. *Exp Brain Res* 105:209–219.
- Sary G, Vogels R, Orban GA (1993) Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* 260:995–997.
- Schwartz EL, Desimone R, Albright TD, Gross CG (1983) Shape recognition and inferior temporal neurons. *Proc Natl Acad Sci USA* 80:5776–5778.
- Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res* 149:1–24.
- Singer W (1999) Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24:49–65.
- Suzuki WA, Amaral DG (1995) Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J Comp Neurol* 349:1–36.
- Tanaka H, Uka T, Yoshiyama K, Kato M, Fujita I (2001) Processing of shape defined by disparity in monkey inferior temporal cortex. *J Neurophysiol* 85:735–744.
- Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189.
- Tanifuji M, Li H, Yamane Y, Rockland KS (2001) Horizontal intrinsic connections as the anatomical basis for the functional columns in the macaque inferior temporal cortex. *Soc Neurosci Abstr* 27:1633.
- Tanigawa H, Fujita I, Kato M, Ojima H (1998) Distribution, morphology, and γ -aminobutyric acid immunoreactivity of horizontally projecting neurons in the macaque inferior temporal cortex. *J Comp Neurol* 401:129–143.
- 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* 4:832–838.
- Uka T, Tanaka H, Yoshiyama K, Kato M, Fujita I (2000) Disparity selectivity of neurons in monkey inferior temporal cortex. *J Neurophysiol* 84:120–132.
- Ungerleider LG, Gaffan D, Pelak VS (1989) Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Exp Brain Res* 76:473–484.
- Vogels R, Biederman I, Bar M, Lorincz A (2001) Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *J Cogn Neurosci* 13:444–453.
- Wang G, Tanaka K, Tanifuji M (1996) Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272:1665–1668.
- Wang G, Tanifuji M, Tanaka K (1998) Functional architecture in monkey inferotemporal cortex revealed by *in vivo* optical imaging. *Neurosci Res* 32:33–46.
- Wang Y, Fujita I, Murayama Y (2000) Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nat Neurosci* 3:807–813.
- Yamane Y, Tsunoda K, Matsumoto M, Phillips A, Tanifuji M (2001) Decomposition of object images by feature columns in macaque inferotemporal cortex. *Soc Neurosci Abstr* 27:1050.
- Yaginuma S, Osawa Y, Yamaguchi K, Iwai D (1993) Differential functions of central and peripheral visual field representations in monkey prestriate cortex. In: *Brain mechanisms of perception and memory: from neuron to behavior* (Ono T, *et al.*, eds), pp. 1–33. New York: Oxford University Press.
- Yukie M, Takeuchi H, Hasegawa Y, Iwai E (1990) Differential connectivity of inferotemporal area TE with the amygdala and the hippocampus in the monkey. In: *Vision, memory and the temporal lobe* (Iwai E, Mishkin M, eds), pp. 129–135. New York: Elsevier.