

Mechanisms of visual object recognition studied in monkeys

KEIJI TANAKA

Riken Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama, 351-0198, Japan

Received 16 May 1999; revised 16 February 2000; accepted 15 March 2000

Abstract—Cells in area TE of the inferotemporal cortex of the monkey brain selectively respond to various moderately complex object-features, and those responding to similar features cluster in a columnar region elongated vertical to the cortical surface. Although cells within a column respond to similar features, their selectivity is not identical. The data of optical imaging in TE have suggested that the borders between neighboring columns are not discrete but columns representing related features overlap one another. We have also found, by training adult monkeys for discrimination of a specific set of shapes, that such a long-term training increases the proportion of TE cells responding to the shapes used in the training even in the adult. The data suggested that TE plays important roles in discrimination of complex shapes and in visual expert learning of discriminating a certain class of objects in the adult.

1. INTRODUCTION

Area TE of the monkey inferotemporal cortex represents the final purely visual stage of the occipitotemporal pathway, which is thought to be essential for visual object recognition (Gross, 1994). The recognition process is flexible, tolerating marked changes in input images due to changes in illumination, viewing angle, and pose of the object (Ullman, 1996). We have studied the functional architecture of TE to find mechanisms underlying the flexible properties of visual object recognition in the primate.

2. MODERATELY COMPLEX FEATURES

An obstacle in the study of neuronal mechanisms of object vision has been the difficulty in determining the stimulus selectivity of individual cells. There is a great variety of object features in the natural world, and it remains to be determined how

the primate brain scales down the variety. There have been studies which 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 was linear, which is hardly expected in higher visual centers.

We have developed a systematic reduction method that involves the use of a specially designed image-processing computer system (Fujita *et al.*, 1992; Kobatake and Tanaka, 1994; Ito *et al.*, 1994, 1995; Wang *et al.*, 1998). After spike activities from a single cell were isolated, many three-dimensional animal and plant models were first presented to find the effective stimuli. Different aspects of the objects were presented in different orientations. Second, the images of the effective stimuli were taken with a video camera and displayed on a TV monitor by the computer to determine the most effective stimulus. 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 requirement for maximal activation was determined as the critical feature for the cell. There are previous studies that used similar reduction methods (Desimone *et al.*, 1984; Tanaka *et al.*, 1991), but the use of an image processing computer made our studies more efficient and enabled new findings.

Examples of the reduction are shown in Fig. 1 for 12 TE cells. The pictures to the left of the arrows are the original images of the most effective objects 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 activation. The critical features for TE cells were more complex than just the orientation, size, color or simple textures, which are known to be extracted and represented by cells in V1, but at the same time not complex enough to represent the image of a natural object through the activity of single cells. The combined activation of multiple cells, each of which represents a partial (either local or holistic) feature of the object image, is necessary.

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.*, 1992), but they have also been found in TE. Thus, there is more convergence of information onto single cells for the representation of faces than for that of non-face objects. This difference may be because discrimination of faces from other objects is not the final goal and further processing of facial images is needed to discriminate among individuals and expressions while distinguishing a non-face object from other objects is close to the final goal.

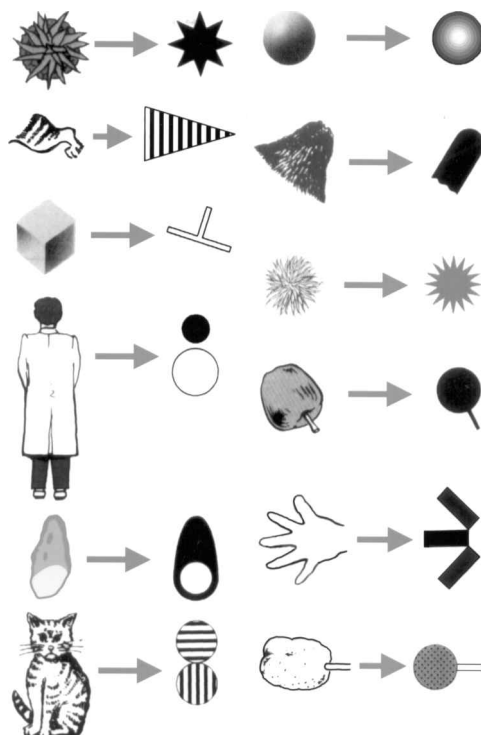


Figure 1. Examples of the reduction process to determine the feature critical for the activation of individual TE cells. The images to the left of the arrows represent the original images of the most effective object stimuli, and those to the right of the arrows, the critical features determined by the reduction.

3. INVARIANCE OF RESPONSES

Our ability for object recognition is retained even in the event of many different kinds of translation of the objects in space. These invariances can, in part, be explained by invariant properties of single cell responses in TE. By using a set of shape stimuli composed of the individually determined critical feature and several other shape stimuli obtained by modifying the critical feature, we have observed that the selectivity for shape is preserved over large receptive fields (Ito *et al.*, 1995). However, the magnitude of the response varies within the receptive field. 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. The receptive fields of TE cells are larger than those of cells in areas in the earlier stages along the ventral visual pathway, but they are, in general, smaller than the largest receptive fields found in the dorsal visual pathway. The receptive fields of TE cells usually range from 10 degrees to 30 degrees in one-dimensional size.

The effects of changes in stimulus size, however, varied more among cells (Tanaka *et al.*, 1991; Ito *et al.*, 1995). Twenty-one percent of the TE cells tested responded to

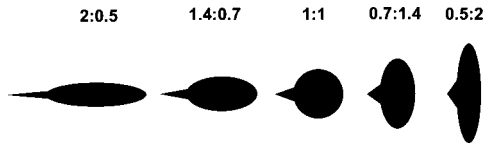


Figure 2. A series of shape stimuli obtained by changing the aspect ratio of one shape. The aspect ratios of neighboring stimuli are different by 1 octave.

a size range of >4 octaves of the critical features with $>50\%$ maximum responses, whereas 43% responded to a size range of <2 octaves. TE cells with considerable invariance for the location and size of stimuli have also been found by Lueschow *et al.* (1994) and Logothetis *et al.* (1995). The tuned cells may be present only as interneurons on the procedure to make invariant cells: those responding to various sizes of the same shape may converge to yield the size-invariant responses. Alternatively, both size-dependent and -independent processing of images may occur in TE.

A definite number of TE cells tolerated reversal of the contrast polarity of the shapes. Contrast reversal of the critical feature evoked $>50\%$ of the maximum response to the critical feature in 40% of tested cells (Ito *et al.*, 1994). Sary *et al.* (1993) found that some TE cells responded similarly to shapes defined by difference in luminosity, direction of motion of texture components and the coarseness of texture, while maintaining their selectivity for shape.

Another kind of invariance of TE cells was found for 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 the 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 (for example, those shown in Fig. 2). We observed that 51% of cells responded to an aspect ratio range of >3 octaves with $>50\%$ of the maximum responses (Esteky and Tanaka, 1998).

Responses of TE cells are more selective for the orientation of stimuli in the frontoparallel plane (Tanaka *et al.*, 1991). Rotation of the critical feature by 90 degrees decreased the response by $>50\%$ for most cells.

4. 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 have found that cells located close together in the cortex have similar stimulus selectivities (Fujita *et al.*, 1992). The critical feature of one isolated cell was determined by using the same procedure as described above, while the responses of another isolated cell, or nonisolated multiunits, were simultaneously recorded. In most cases, the second cell responded to the optimal and suboptimal stimuli of

5. OPTICAL IMAGING OF THE COLUMNAR ORGANIZATION

To further study the spatial properties of the columnar organization in TE, we used the technique of optical imaging with intrinsic signals (Wang *et al.*, 1996, 1998). The cortical surface was exposed, illuminated with red light tuned to 605 nm, and the reflected light image was taken by a CCD video camera. The reflected images for different visual stimuli were compared.

At this wavelength, the differential components of signals mostly reflect the changes of the density of deoxidized hemoglobin in the capillaries. Activated neuronal tissue takes up more oxygen from hemoglobin, so that the density of deoxidized hemoglobin in the nearby capillaries increases. Because deoxidized hemoglobin absorbs more light than oxidized hemoglobin at the specified wavelength, the region of the cortex with elevated neuronal activities becomes darker in the reflected image.

We first determined the critical features for 15 to 25 single cells in several unit-recording sessions, and then optical signals were imaged during visual stimulation with the critical features. The image during stimulation with a critical feature divided by that during stimulation with a control stimulus revealed a small dark spot covering the electrode penetration from which the critical feature was determined. The averaged diameter of the dark spot was 490 μm . Although the critical feature was determined for a single cell, a large proportion of cells in the region must be activated to produce observable metabolic change. Therefore, the localized and specific occurrence of dark spots indicates a regional clustering of cells with similar stimulus selectivity.

However, when we observed a larger extent of the cortical surface, we found that the presentation of a single feature activated multiple spots. In Fig. 4, the spots activated by 8 moderately complex features are indicated by different kinds of lines and superimposed, those by 4 features in the upper half and those by other 4 features in the lower half. For example, the feature 1 evoked 6 spots, and feature 2 evoked 2 spots. 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 overlapped regions, which were activated by many stimuli, likely represent columns with non-selective cells. But, others that were activated by only 2 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. 5). This experiment was also guided by a unit-recording experiment. We recorded 5 cells in one electrode penetration around the center of the imaged region, and all of them selectively responded to faces. Three of them responded maximally to the front view of the face, whereas the remaining two responded to the profile, the lateral view of the face. In the optical imaging session, 5 different views of the same doll face were presented in combination with 14 non-face features. All of the faces

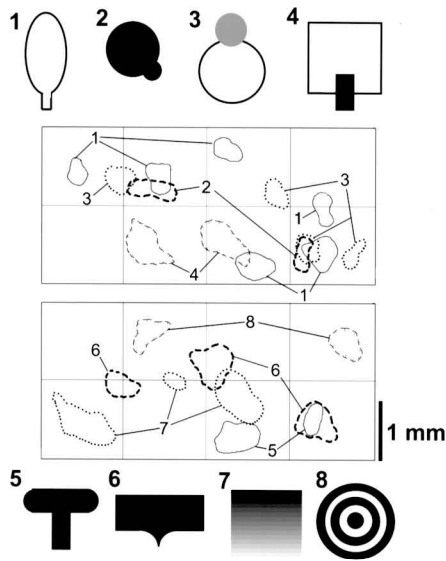


Figure 4. Activation maps evoked by the presentation of 8 moderately complex features. To obtain the map, each image was subtracted by the reference image averaged over the images obtained for all the different stimuli combined in the experiment to remove the global darkening, the activation spots were 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.

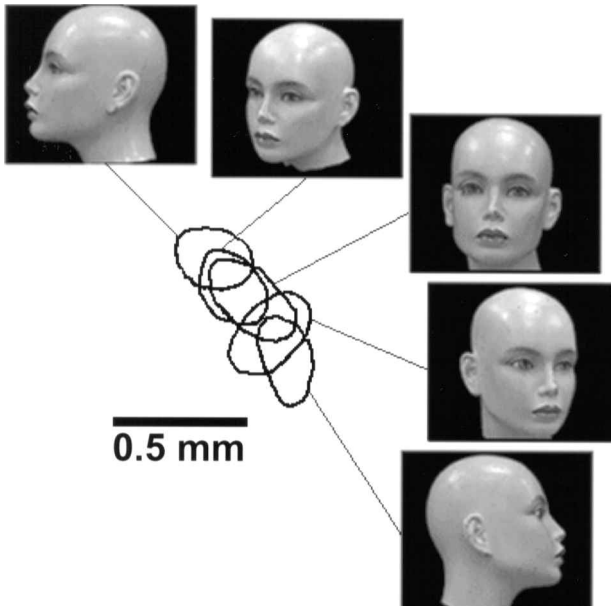


Figure 5. Systematic movement of the activation spot with rotation of the face. The images were obtained for 5 different views of the same doll face. The reference image obtained by averaging the five images has been subtracted. The contours circumscribing the pixels with t -values at $p < 0.05$, compared with the reference image, are superimposed at the bottom.

evoked activation spots around the center of the illustrated 1.3×1.3 mm region. But, 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 300 to 400 μm in diameter and the overall region was 800 μm . 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 3 other monkeys.

The data for the non-face features are more limited, but I hypothesize that there are similar structures, and propose a modified model of the columnar organization in TE in Fig. 6. The borders between neighboring columns are not necessarily distinctive. 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.

This systematic arrangement of related columns can be used for various kinds of computation necessary for object recognition. One simple possible computation is the generalization of activation by the horizontal excitatory connections to nearby columns representing related features. We may call it the selective blurring of activation. Another possible simple processing is the mutual inhibition among the nearby columns for the winner-takes-all type selection.

The continuous mapping of different views of faces cannot be generalized to non-face objects. Because the critical features of TE cells are only moderately complex except for faces, the image of a non-face object has to be represented by a

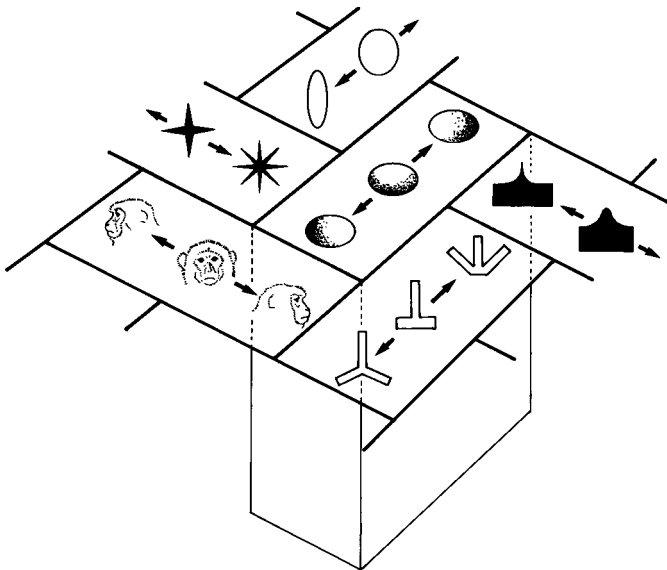


Figure 6. Revised schematic diagram of the columnar organization in TE.

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 in TE.

6. CHANGEABILITY OF THE SELECTIVITY IN THE ADULT

The selectivity of inferotemporal cells can be changed in adult animals by a long-term training. We found this by training two adult monkeys for the recognition of the 28 moderately complex shapes shown in Fig. 7 (left) and recording from inferotemporal cells after the training (Kobatake *et al.*, 1998). The training paradigm was a kind of delayed matching to sample (Fig. 7, right). One stimulus that was randomly selected from the set of 28 stimuli appeared on a television display as a sample, the monkey touched it, and the sample disappeared. After a delay period, the sample appeared again on the display but this time together with 4 other stimuli that were randomly selected from the stimulus set. The monkey selected the sample and touched it to get a drop of juice as a reward. After an intertrial interval, the trial was repeated with a different sample. The monkey performed the task on a stand-alone apparatus placed in front of the home cage. The monkey came to the apparatus and practiced the task whenever it wanted. The training was started with a 1 s delay, and the delay was increased gradually to 16 s. At the end of the training, the monkey performed 500 successful trials with >80% performance.

After the training was completed, we prepared the monkeys for repeated recordings, and conducted recordings from TE under anesthesia, once a week for 3 to 4

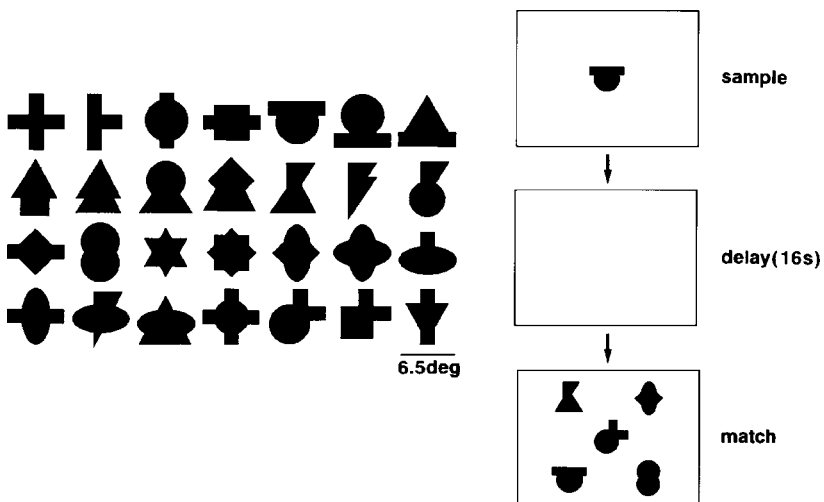


Figure 7. The 28 shapes used for the training (left) and the paradigm (right).

months. The training was continued on the days when the recordings were not conducted. We determined for individual cells the most effective stimulus from the set of animal and plant models, and the response to this best object stimulus was compared with the responses of the same cell to the training stimuli. In this experiment, we did not conduct the reduction process, but just took the images of several most effective object stimuli with a video camera and presented them under computer control in combination with the training stimuli.

The cell illustrated in Fig. 8 responded maximally to the sight of a watermelon among the object stimuli. But, the cell responded more strongly to the cross shape,

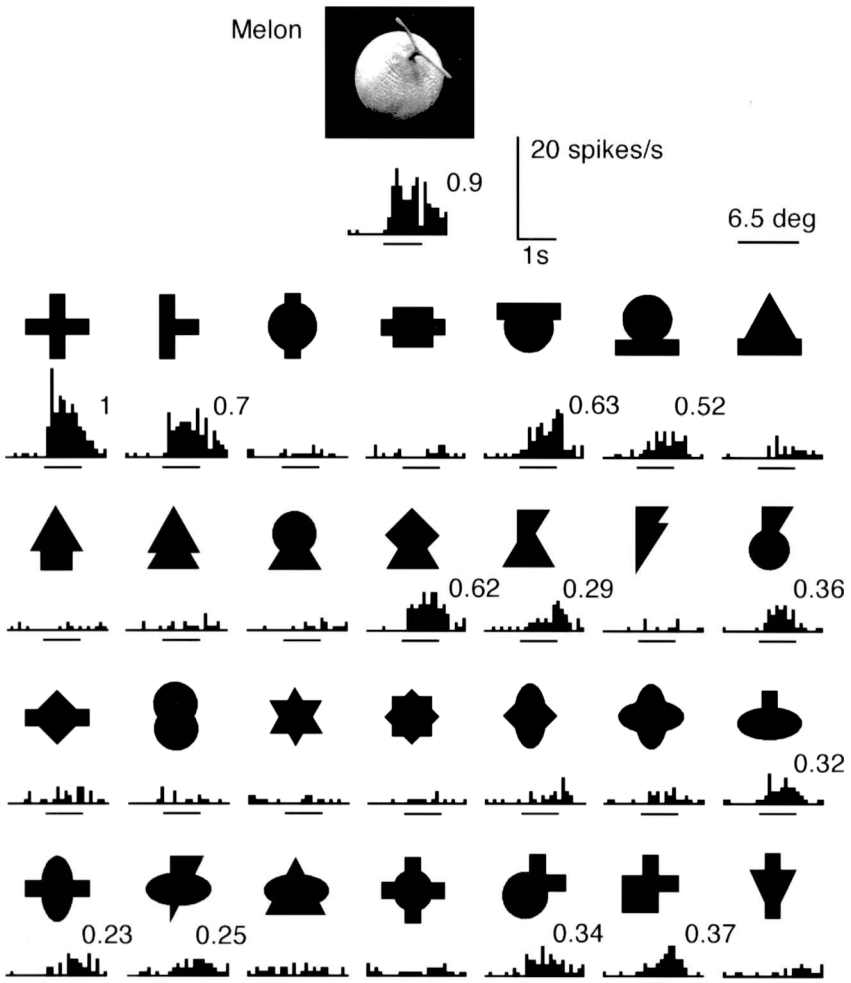


Figure 8. Responses of one TE cell to the image of the most effective stimulus of the object stimulus set (top) and its responses to the 28 shape stimuli used for the training. This cell was recorded from a monkey that had been trained with the 28 stimuli. Statistically significant responses ($p < 0.05$) are labeled with their relative response magnitudes. Cited from Kobatake *et al.* (1998).

which was one of the training stimuli. There were also responses to several other training stimuli. Many cells recorded from the trained monkeys responded more strongly to some of the training stimuli than to the best object stimulus, like in this example.

To quantitatively compare the results obtained from the 2 trained monkeys with those obtained from 3 untrained control monkeys, we calculated the ratio of the maximal response to the training stimuli to the cell's overall maximal response and compared its distribution between the two groups (Fig. 9). The x -axis in the figure is the ratio, and the y -axis the proportion of cells. The top histogram shows the distribution among 131 cells recorded in the trained monkeys, and the bottom histogram shows the distribution for 130 cells recorded from the 3 control monkeys. One on the x -axis signifies that the cell was maximally activated by some of the training stimuli, and 0 signifies that the cell was not activated at all by any of the training stimuli. Twenty-five percent of the cells recorded from the trained monkeys responded maximally to some of the training stimuli, but only 5% of the cells in the control monkeys responded maximally to some of the stimuli. These results indicate that the number of cells maximally responsive to training stimuli increased during the period of the discrimination training. Sakai and Miyashita (1991, 1994) and Logothetis *et al.* (1995) trained adult monkeys to discriminate among fractal patterns or wire-frame objects and found that many TE cells responded to the learned stimuli after the training. A unique contribution of

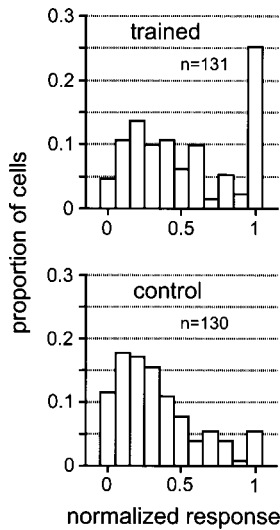


Figure 9. Distribution of the normalized magnitude of the strongest responses of the individual cells to the training stimuli. Those from 131 TE cells recorded from the 2 trained monkeys are shown at the top, and those from 130 TE cells recorded from the 3 control monkeys are shown at bottom. The magnitude of the response was normalized with respect to the maximal response of the cell (the larger of the strongest response of the cell to the reference object stimuli and the strongest response of the cell to the training stimuli). Cited from Kobatake *et al.* (1998).

our study is the demonstration that training increases the proportion of TE cells that respond to the learned stimuli as measured against untrained controls.

The responses to the training stimuli were not sharply tuned to a particular stimulus but rather distributed for several different training stimuli. It was not the case that detectors of particular training stimuli appeared. This character of the effects of training was also found in responses of the cells to the 8 stimuli shown in Fig. 10A. These stimuli were not used during the training, but were presented for the single-cell recordings under anesthesia. They were composed of the same primitives as those of the training stimuli, so they as a whole covered the feature space occupied by the set of training stimuli. These 8 stimuli are referred to as hidden stimuli. Cells recorded from the 2 trained monkeys responded to these hidden stimuli very well. The histogram at the middle of Fig. 10B shows the distribution of the normalized magnitude of individual responses to the hidden stimuli of TE cells recorded from the trained monkeys. It was nearly the same as the distribution of the normalized magnitude of the same cells to the training stimuli shown at the top of Fig. 10B. The difference between the two distributions is plotted at the bottom, and was not significant. These results suggest that the changes induced in TE as a result of the training did not consist of individual cells becoming tuned to particular stimuli, but could be attributed to the feature space in which the training stimuli were distributed becoming more densely covered by TE cells.

We investigated the properties of the distributed coding of the training stimuli, by examining the correlation between the responses of two cells. The cell pairs were not necessarily recorded simultaneously. They were recorded at different times, and in different penetrations. The correlations in 2 pairs are illustrated in the upper part of Fig. 11. The x -axis is the normalized response of cell #1, the y -axis the normalized response of cell #2, and there are 28 dots corresponding to the 28 training stimuli. In many pairs, the two cells responded to different sets of stimuli. These pairs are not of interest for the purpose of analysis here. There were pairs, whose responses overlapped partially, namely, the same stimuli evoked strong responses in both cells (as the two pairs shown in the upper part of Fig. 11). In these pairs, a group of stimuli evoked strong responses in both cells, but there were always a second group of stimuli that evoked strong responses only in cell #1 and a third group that evoked strong responses only in cell #2. These results are consistent with the speculation that the responses of the cells were tuned to features contained in the stimuli, and not to the whole shape of the stimuli. Cell #1 responded to feature #1, while cell #2 responded to feature #2. Both cells responded to the first group of stimuli because the stimuli had both feature #1 and feature #2. Only one cell responded to the second and third group of stimuli because they had only one of the features. The correlation coefficient between the responses of the two cells tended to be very small and distributed around 0, as shown at the bottom of Fig. 11. Only a few pairs showed correlation coefficients larger than 0.5.

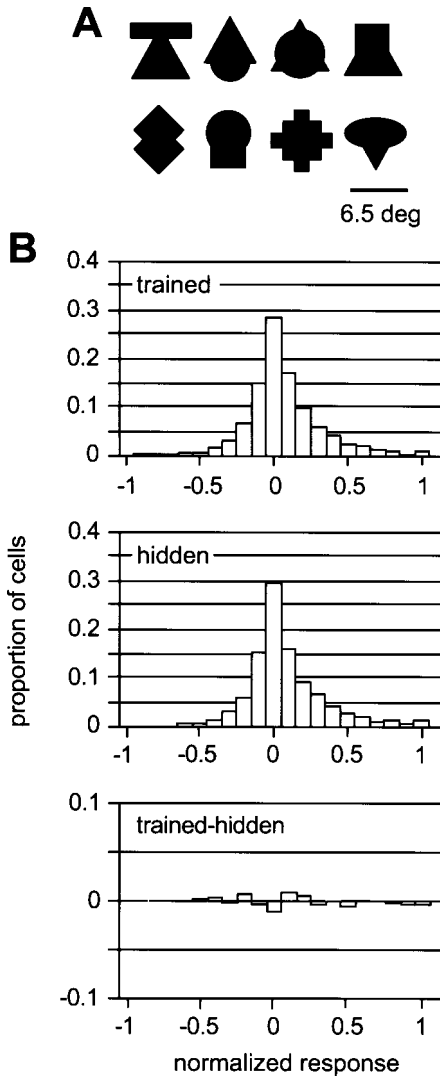


Figure 10. The spread of training effects. A: Eight stimuli not used for training but presented during the recordings under anesthesia. They are referred to as ‘hidden stimuli’. B: Distribution of the response magnitudes to the individual training stimuli of the cells recorded in the trained monkeys (top), distribution of the magnitudes of responses of the same cells to the individual hidden stimuli (middle) and the difference between the two distributions (bottom).

To effectively discriminate shapes, the unit of neural circuitry may either code for the entire image of an exemplar, or features or aspects common to some of the exemplars. The former is more straightforward when the mechanism of the change is considered, whereas the latter makes it easier to generalize the training effect to novel but similar shapes. The task in which the monkeys were trained used a fixed set of 28 shapes and thus did not require generalization. Nevertheless,

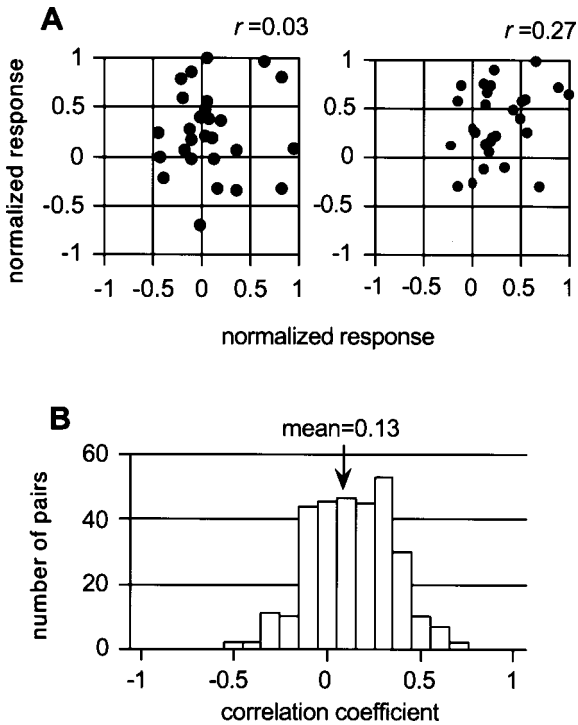


Figure 11. Independence of responsivity to different training stimuli among cells recorded in trained monkeys. A: Two examples of the scatter diagrams showing the correlation between the responses of two cells to the 28 training stimuli. The x -value of an individual dot represents the magnitude of the response elicited by one training stimulus in one cell of the pair, and the y -value of the dot represents the magnitude of the response elicited by the same stimulus in the other cell of the pair. There are 28 dots corresponding to the 28 training stimuli. The values of r represent Pearson's correlation coefficient for the distribution. B: The distribution of Pearson's correlation coefficients among 309 cell pairs in which at least one training stimulus evoked responses exceeding 50% of the maximal response of either cell. Cited from Kobatake *et al.* (1998).

the ability to generalize must certainly hold selective advantage in nature and may therefore constitute a cortical operating principle that is always in effect. The results are more consistent with the hypothesis that inferotemporal cells develop responses to the learned stimuli by coding partial features, or aspects, common to multiple exemplars.

7. FUNCTIONS OF THE TE COLUMNS

The finding of columnar organization in TE has prompted investigation of the mechanism of object recognition on the bases of the anatomical structure and physiological properties of neurons and networks.

Representation by multiple cells in a columnar module, in which the selectivity varies from cell to cell while effective stimuli largely overlap, can satisfy two apparently conflicting requirements in visual recognition: one is the ability to disregard subtle changes in input images, and the other, the preciseness of representation. While the image of an object projected onto the retina changes due to changes in illumination, viewing angle and pose of the object, the global organization of outputs from TE must change little. The clustering of cells having overlapping and slightly differing selectivities works as a buffer to absorb the changes. Although the responses of 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 whole range of flexibility of our 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 three-dimensional shape may be packed 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 discussion and investigation. A possibility is that outputs of the cells preferring different orientations, sizes, aspect ratios and contrast polarities of the same shape overlap in the target structure to evoke the same effects. Anatomical study with an injection of anterograde tracer into a focal site in TE suggested that the projections from TE to the ventrocaudal striatum of the basal ganglia have this property (Cheng *et al.*, 1998).

The 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. The projections from the ventroanterior part of TE to the perirhinal cortex have extensive divergence (Saleem and Tanaka, 1996). The projection terminals from a single site of ventroanterior TE covers about 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 function of the columnar organization in TE may go beyond the discrimination of input images. The results of the optical imaging experiments suggested that there is a continuous mapping of features within cortical units about 1 mm in length across the cortical surface. The continuous mapping may be the structural basis to conduct computations involving the mapped features based on the local neuronal connections between the cells representing the related but different features. The computations may serve to translate the image of an object for three-dimensional rotations, and produce the image under different illumination conditions or different articulation poses.

Thus, the columnar organization of TE may enable overlapping and continuous representation of object features, upon which various kinds of calculations can be performed.

REFERENCES

- Bruce, C., Desimone, R. and Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque, *J. Neurophysiol.* **46**, 369–384.
- Cheng, K., Saleem, K. S. and 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.* **15**, 7902–7925.
- Desimone, R., Albright, T. D., Gross, C. G. and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque, *J. Neurosci.* **4**, 2051–2062.
- Esteky, H. and 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.
- Fujita, I., Tanaka, K., Ito, M. and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex, *Nature* **360**, 343–346.
- Gallant, J. L., Braun, J. and Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex, *Science* **259**, 100–103.
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W. and Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and cartesian gratings in area V4 of the macaque monkey, *J. Neurophysiol.* **76**, 2718–2739.
- Gross, C. G. (1994). How inferior temporal cortex became a visual area, *Cerebral Cortex* **5**, 455–469.
- Ito, M., Fujita, I., Tamura, H. and Tanaka, K. (1994). Processing of contrast polarity of visual images in inferotemporal cortex of the macaque monkey, *Cerebral Cortex* **5**, 499–508.
- Ito, M., Tamura, H., Fujita, I. and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex, *J. Neurophysiol.* **73**, 218–226.
- Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex, *J. Neurophysiol.* **71**, 856–867.
- Kobatake, E. and Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys, *J. Neurophysiol.* **80**, 324–330.
- Logothetis, N. K., Pauls, J. and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys, *Curr. Biol.* **5**, 552–563.
- Lueschow, A., Miller, E. K. and Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition, *Cerebral Cortex* **5**, 523–531.
- Perrett, D. I., Rolls, E. T. and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex, *Exp. Brain Res.* **47**, 329–342.
- Richmond, B. J., Optican, L. M., Podell, M. and 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.
- Sakai, K. and Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates, *Nature* **354**, 152–155.
- Sakai, K. and Miyashita, Y. (1994). Neuronal tuning to learned complex forms in vision, *NeuroReport* **5**, 829–832.
- Saleem, K. S. and 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.
- Sary, G., Vogels, R. and Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons, *Science* **260**, 995–997.

- Schwartz, E. L., Desimone, R., Albright, T. D. and Gross, C. G. (1983). Shape recognition and inferior temporal neurons, *Proc. Natl. Acad. Sci. USA* **80**, 5776–5778.
- Tanaka, K., Saito, H., Fukada, Y. and Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey, *J. Neurophysiol.* **66**, 170–189.
- Ullman, S. (1996). *High-level Vision: Object Recognition and Visual Cognition*. The MIT Press, Bradford Book, Cambridge, Massachusetts and London, England.
- Wang, G., Tanaka, K. and Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex, *Science* **272**, 1665–1668.
- Wang, G., Tanifuji, M. and Tanaka, K. (1998). Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging, *Neurosci. Res.* **32**, 33–46.