Effects of Shape-Discrimination Training on the Selectivity of Inferotemporal Cells in Adult Monkeys

EUCALY KOBATAKE,1,2 GANG WANG,1,3 AND KEIJI TANAKA1,4
1RIKEN Brain Science Institute, Wako-shi, Saitama 351-01; 2Electrotechnical Laboratory, Tsukuba-shi, Ibaraki 305; 3Department of Physiology II, Faculty of Medicine, Kagoshima University, Kagoshima-shi, Kagoshima 890; and 4Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Wako-shi, Saitama 351-01, Japan

Kobatake, Eucaly, Gang Wang, and Keiji Tanaka. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. J. Neurophysiol. 80: 324 ± 330, 1998. Through extensive training, humans can become ‘‘visual experts,’’ able to visually distinguish subtle differences among similar objects with greater ease than those who are untrained. To understand the neural mechanisms behind this acquired discrimination ability, adult monkeys were fully trained to discriminate 28 moderately complex shapes. The training effects on the stimulus selectivity of cells in area TE of the inferotemporal cortex were then examined in anesthetized preparations. Area TE represents a later stage of the ventral visual cortical pathway that is known to mediate visual object discrimination and recognition. The recordings from the trained monkeys and untrained controls showed that the proportion of TE cells responsive to some member of the 28 stimuli was significantly greater in the trained monkeys than that in the control monkeys. Cell responses recorded from the trained monkeys were not sharply tuned to single training stimuli, but rather broadly covered several training stimuli. The distances among the training stimuli in the response space spanned by responses of the recorded TE cells were significantly greater in the trained monkeys than those in the control monkeys. The subset of training stimuli to which individual cells responded differed from cell to cell with only partial overlaps, suggesting that the cells responded to features common to several stimuli. These results are consistent with a model in which visual expertise is acquired through the development of differential responses by inferotemporal cells to the images of relevant objects.

INTRODUCTION
The human ability to discriminate between similar object images appears to depend on visual circumstances. It is said that Inuits can discriminate many different kinds of snow, and mounted people can discriminate many different kinds of horses. This capacity also depends on profession. Shepherds can distinguish individual sheep, and neuroscientists can distinguish individual experimental animals. What is the neural basis of this phenomenon?

One possible mechanism is that the number of neurons responsive to the images of a relevant class of objects increases and differentiation develops among them such that individual cells respond differentially to separate members of the object class. Area TE of the inferotemporal cortex is a likely site for such changes in the monkey brain, because it is located at a later stage of the ventral visual pathway, which is essential for object discrimination and recognition, and also because cells in TE respond selectively to complex features of objects (for review see Gross 1994; Logothetis and Sheinberg 1996; Miyashita 1993; Tanaka 1996).

There are many cells specifically responsive to faces in TE and the region mediodorsally adjoining TE in the superior temporal sulcus (Bruce et al. 1981; Desimone et al. 1984; Perrett et al. 1982). It is often suggested that these cells have developed so that the monkey can distinguish individual faces and facial expressions. Indeed, it has been reported that the cells differentially respond to different faces, although their selectivity is broad (Baylis et al. 1985; Yamane et al. 1988; Young and Yamane 1992). These face-selective cells may have developed over generations or through early development. The present study was designed to determine what kind of changes occur in TE of adult monkeys that are extensively trained to discriminate among members of a shape class.

Some earlier studies relate to this issue. Sakai and Miyashita (1991, 1994) and Logothetis et al. (1995) trained adult monkeys to discriminate among fractal patterns or wireframe objects and found that many inferotemporal cells responded to the learned stimuli after training. We performed the identical recording procedures on a population of cells in both trained and untrained control monkeys, under conditions of anesthesia and separate from training, and found that the proportion of cells responsive to some of the training stimuli in the trained monkeys was greater than that in the control untrained monkeys. Portions of the present results have been previously reported in abstract form (Kobatake et al. 1992, 1993).

METHODS
Training
Four adult macaque monkeys (Macaca fuscata) served as subjects. Two of these received training on a visual recognition task with the 28 shape stimuli shown in Fig. 1. One trained monkey (female) weighed 5.2 kg, and the other (male) weighed 5.5 kg at the beginning of the shape training. At that time their ages were estimated to be between 4 and 5 yr. Each trial of the task began with the presentation of a sample chosen from among the 28 stimuli on a computer display equipped with a touch screen. As soon as the monkey touched the stimulus, it disappeared from the screen. After a delay period, the sample stimulus reappeared with four distracters chosen from the same set. The monkey obtained a drop of juice as reward for touching the sample on the screen. The delay was initially set to 1 s and gradually increased to 16 s. Both the
Recording

The monkeys were prepared for repeated recordings with initial aseptic surgeries. Under anesthesia with pentobarbital sodium (35 mg/kg ip, supplemented when necessary by 10 mg/kg), a brass block for head fixation was attached to the top of the skull, two stainless steel screws for electroencephalogram recording were implanted into the skull, the zygoma was removed, and the lateral surface of the skull was exposed and covered with resin for later recording of cell activity. Before the first recording session, eye optics were measured to select appropriate contact lenses, and photographs of the fundus were taken to determine the position of the fovea.

Recordings were performed under anesthesia once a week on each monkey. The trained monkeys underwent training on the other days of the week. Recording sessions began with the induction of anesthesia with ketamine hydrochloride (10 mg/kg im). An endotracheal cannula was inserted through the tracheal opening, and a small hole was made in the resin-coated skull. Throughout the recording session, animals were immobilized with pancuronium bromide (0.08 mg/kg im, followed by 0.024 mg·kg⁻¹·h⁻¹ im), and the anesthesia was maintained by artificial ventilation with a mixture of N₂O and O₂ (70:30). The depth of anesthesia was assessed by monitoring the electrocardiogram and electroencephalogram, with isoflurane added to the gas mixture when necessary. Atropine sulfate (0.5 mg) was subcutaneously administered every 3 h to reduce salivation.

The pupils were dilated and the lenses relaxed by local application of 0.5% tropicamide-0.5% phenylephrine. The corneas were fitted with contact lenses of appropriate power with artificial pupils 3 mm diam so that stimulus images on a television display at 57 cm from the corneas would be focused on the retina. Several retinal landmarks, such as the intersection of blood vessels and the center of the optic disk, were projected onto the display with a reversible retinoscope (Sanso, Tokyo). The position of the fovea was determined indirectly from the positions of these landmarks by referring to the photographs of the fundus. The stimuli were presented to the eye contralateral to the recording site.

Extracellular unit recordings were made from the dorsolateral portion of TE (Fig. 2) with glass-coated Elgiloy electrodes (2–5 MΩ at 1 kHz). The electrodes were advanced from the lateral side through a pinhole made in the dura mater with a needle. The exposed dura mater was covered with paraffin to prevent it from drying and to reduce movements of the brain caused by pulsation and respiration. The position of penetration was determined with reference to a point marked on the resin-coated skull. The hole in the skull was filled with resin after the recording was completed. All recording procedures were conducted under aseptic conditions. Within a few hours after the last injection of muscle relaxant, spontaneous respiration resumed and became normal. The monkey was returned to its home cage after the injection of an antibiotic (Pentcilin, 40 mg/kg im; Sankyo, Tokyo). Recordings were also made from area TEO, and the border between TE and TEO was determined based on the size of the receptive fields (Kobatake and Tanaka 1994). Data from TEO cells were excluded from this paper. The experimental protocol had been approved by the Experimental Animal Committee of the RIKEN Institute. Monkeys were regu-
larly monitored by a veterinarian and cared for in accordance with the Guiding Principles for the Care and Use of Animals in the Field of Physiological Science of the Japanese Physiological Society.

Visual stimulation and procedure on individual cells

To evaluate the magnitudes of the responses to the training stimuli, we used a reference set of 75 object stimuli consisting of animal and plant imitations and laboratory junk objects [see Kobatake and Tanaka (1994) for the list of objects]. Once activity in a cell was isolated, all of the object stimuli in the set were successively hand-presented to the monkey, and the two to four most effective object stimuli were determined by listening to the evoked activity on an audiomonitor. Images of these object stimuli were then taken with a video camera and stored on a computer. The background of the stimuli was filled with a homogenous gray. Unlike our previous studies (reviewed in Tanaka 1996), in the interest of time, we did not determine which features of the stimulus images were critical for activation. Finally, the images of the object stimuli were presented on the television display in combination with the training stimuli to evaluate the relative magnitude of the responses to the training stimuli.

The stimuli were intermixed and presented 10 times in cyclic order in this quantitative test. They were presented for 1 s, with 2-s blank intervals between trials. During the presentation they moved along a circular path (0.29° in radius and 0.96 s/cycle, without change in orientation), to avoid sensory adaptation in the paralyzed state. The magnitude of responses was determined as the mean firing rate during the stimulus presentation minus the spontaneous firing rate during the 1-s period immediately before the stimulus presentation. Considering the response latency, we shifted the time window for the response within a range of 50–250 ms from the precise onset of stimulus presentation so as to maximize the mean firing rate. The statistical significance of the responses was determined by comparing the mean firing rates within the window for 10 individual responses, with 10 spontaneous firing rates immediately preceding each stimulus presentation, using the Kolmogorov-Smirnov (K-S) test. The magnitude of responses, after subtraction of the spontaneous firing rate, was normalized with respect to the cell’s maximal responses (i.e., the larger of the strongest response to the object stimulus and the strongest response to the training stimuli). In comparing the trained and control monkeys, we used normalized responses rather than absolute magnitudes because the absolute magnitude considerably varies from cell to cell and changes according to conditions of the preparation such as the depth of anesthesia and partial damage by the electrode to the recorded cells.

A proper comparison of normalized values, however, depends on the selection of reference object stimuli. One may wonder whether the routine use of a fixed set of reference stimuli in the quantitative test is more objective than our method based on the on-line, not fully quantitative selection of the two to four most effective object stimuli from a large set of object stimuli. We thought that the routine use of a moderately sized set (e.g., 20) of stimuli would not match our purposes here. The cells in TE are individually tuned to different complex features, and these preferred features, as a whole, define a huge feature space (Desimone et al. 1984; Fujita et al. 1992; Ito et al. 1994, 1995; Kobatake and Tanaka 1994; Steinberg and Logothetis 1997; Tanaka et al. 1991). A fixed set of stimuli of medium size (e.g., 20) might fail to hit the effective stimulus range of many recorded cells. The reference stimuli will work properly if most cells are activated by at least some of them. The whole set of object images contains a larger set of partial features; therefore the chances are higher that some view of an object will contain the features effective for the activation of a recorded cell. We also considered the possibility that the experimenters inadvertently searched for effective reference stimuli more extensively in cells recorded in the control monkeys. This possibility was dismissed by inspecting the distribution of the magnitude of responses to the selected reference object stimuli between the two groups of cells recorded from the trained and control monkeys (Fig. 4).

To avoid experimenter bias in sampling cells, we followed two rules during recording. First, we made sure that recording positions be as evenly distributed as possible (Fig. 2), and that the distance between any two penetrations be at least 1.5 mm. Second, within penetrations we sampled at intervals >200 μm. Several cells were sampled at shorter intervals to examine the clustering of cells, but their data were excluded from the present analysis.

RESULTS

The data set used in this paper consisted of 131 cells recorded from the 2 trained monkeys and 130 cells recorded from the 3 untrained control monkeys. These cells were selected according to the criteria that 1) at least one stimulus (either a training stimulus or a reference object stimulus) evoke statistically significant responses (P < 0.05); 2) the cells be located in TE and 3) they be removed by >200 μm from the last studied cell along the penetration.

Training effects were first examined by comparing the distribution of the strongest responses of individual cells to the training stimuli, between the trained and control monkeys. The magnitude of the responses was normalized with respect to the individual cell’s maximal response (i.e., 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). A value of 1 means that the cell responded more strongly to some of the training stimuli than to the most effective object stimulus, as in the case of the cell whose responses are shown in Fig. 1. A value of 0 means that the cell did not respond to any of the training stimuli. The distribution in the trained monkeys is significantly shifted toward 1 as compared with that in the untrained control monkeys (Fig. 3, left; P < 0.001 with K-S test). Especially, cells with a value of 1 accounted for 25% of the cells recorded in the trained monkeys, but only 5% of those recorded in the control monkeys. The mean value from the trained monkeys is 0.53, whereas that from the control monkeys is 0.34. The proportion of cells with a value larger than 0.5 was 47% in the trained monkeys, but only 22% in the controls.

The distribution in each trained monkey also differed significantly from that in each control monkey in all pairs (Fig. 3, right; P < 0.05 for all pairs with Mann-Whitney U test, P < 0.1 for “trained 2” vs. “control 3” and P < 0.05 for other pairs with K-S test). Moreover, the average values in individual trained monkeys (0.59 and 0.50) are significantly greater than those in individual control monkeys (0.31, 0.35, and 0.36) with t-test (P < 0.01). Finally, the responses recorded after the shape training from the left TE (trained 2) were significantly (P < 0.01 with K-S test) greater than the responses recorded before the shape training from the right TE of the same monkey (control 2).

The possibility that the difference obtained was due to experimenter bias in the selection of effective object stimuli was dismissed based on the following two facts. First, the absolute magnitudes of the individual cells’ largest responses to the reference object stimuli did not show significant differ-
The responses recorded from the trained monkeys (filled bars) were significantly larger than those recorded from the control monkeys (open bars) up to 14th maximal responses \((P < 0.01 \text{ (K-S test) }\) for 1st to 8th responses and \(P < 0.05 \text{ (K-S test) }\) for 9th to 14th responses).

We next asked whether the changes in responses of TE cells could underlie the learned discrimination. Because cells were generally responsive to multiple members of the training stimuli, the discrimination should have been dependent on population responses. The potential of population coding can be evaluated by examining distances among the training stimuli in the space spanned by the responses of recorded TE cells (Gochin et al. 1994; Young and Yamane 1992).

We let responses of one cell represent one dimension of the space. The number of dimension was thus equal to the number of cells, and one training stimulus was represented by one point in this space. The distance between two stimuli in this space was calculated by

1. taking a difference between responses of one particular cell to the two stimuli,
2. multiplying the difference by itself,
3. summing the square value across cells, and
4. taking a square root of the sum.

To compare the distances between the trained and control monkeys, from which different numbers of cells were recorded, the distances were normalized by the square root of the cell numbers (\(\sqrt{131} / 2\) in the trained monkeys and \(\sqrt{130} / 2\) in the control monkeys). The distances in the trained monkeys were significantly larger than those in the control monkeys (Fig. 4, \(P < 0.05\) with both K-S test and Mann-Whitney U test). The means were 18.1 spikes/s for the cells recorded from the trained monkeys versus 16.8 spikes/s for controls. Second, the absolute magnitudes of the individual cells’ largest responses to the training stimuli in the trained monkeys were significantly greater than those in the controls (Fig. 5, \(P < 0.001\) with K-S test). These findings illustrated through Figs. 3–5 indicate that the responsiveness of TE cells to the training stimuli increased as a result of training.

Cells’ responsiveness was not sharply tuned to particular training stimuli. In the cell illustrated in Fig. 1, for example, five training stimuli evoked responses higher than 50% of the cell’s maximal response. On the average, among the 28 cells that were recorded from the trained monkeys and that maximally responded to some of the training stimuli, three training stimuli evoked responses >50% of individual cells’ maximal responses, and eight training stimuli evoked responses in excess of 25%. The broad tuning may suggest that the discrimination depended on activity of cell population.

The effect of training was also found in the frequency of these moderately strong responses to suboptimal training stimuli. The responses of the same rank order in individual cells were compared between the trained and control monkeys. Figure 6 shows the comparisons for the maximal (1st), 2nd, 4th, 10th, 16th, 22nd maximal, and the smallest (28th) responses. The responses recorded from the trained monkeys (filled bars) were significantly larger than those recorded from the control monkeys (open bars) up to 14th maximal responses \((P < 0.01 \text{ (K-S test) }\) for 1st to 8th responses and \(P < 0.05 \text{ (K-S test) }\) for 9th to 14th responses).

**Fig. 3.** Distribution of the normalized magnitude of the individual cells’ strongest responses to the training stimuli. The overall distributions for 131 cells recorded from the 2 trained monkeys and for 130 cells recorded from the 3 control monkeys are shown at left, whereas the distributions in individual monkeys are shown at right. The magnitude of the response, after subtracting the spontaneous firing rate, 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).

**Fig. 4.** Distribution of the absolute magnitudes of the responses to the stimuli. The responses of the same rank order in individual most effective reference-object-stimuli for the 131 cells recorded from the 2 trained monkeys (top) and for the 130 cells recorded from the 3 control monkeys (bottom).
We trained adult monkeys to discriminate among a class of shape stimuli and found that the proportion of inferotemporal cells responsive to some of the training stimuli in the trained monkeys was greater than that in the control untrained monkeys. Because consistent results were obtained in two trained monkeys and three control monkeys, we take the results as evidence that the proportion of such cells increased in the inferotemporal cortex through the training. Individual cells responded to different subsets of the training stimuli. This change in responsiveness fulfilled the requirements of the task: the development of diverging responsiveness increased the distances among the training stimuli representations in the feature space spanned by responses of inferotemporal cell population, which in turn made the discrimination easier. Single cells responded to multiple members of the training stimuli, which suggests that the discrimination was based on the activity of cell population. The increase in distances would contribute to either the passive or active mechanism, which has previously been pro-

![Figure 5](image1.png)

**FIG. 5.** Distributions of the absolute magnitudes of the strongest responses to the training stimuli for the 131 cells recorded from the 2 trained monkeys (top) and for the 130 cells recorded from the 3 control monkeys (middle), and the difference between them (bottom). Note that the ordinate scale on the bottom is twice that in the top and middle.

Thus likely underlay the learned discrimination in the trained monkeys.

Many of the cells recorded from the trained monkeys were responsive to multiple members of the training stimuli, and different cells responded to different subsets of the training stimuli. To infer the bases of their selectivity, we examined the correlation in responsiveness among pairs of cells. Pairs were made among the 62 cells recorded in the trained monkeys that gave responses >50% of the maximal to at least one training stimulus. Cells were paired only within each trained monkey, because of the possibility that different monkeys used different strategies in discriminating the training stimuli. Out of the resulting 1,958 pairs, 309 showed response overlaps, i.e., at least one training stimulus commonly evoked responses >50% of the maximal in both cells. However, as can be seen in the 2 examples shown in Fig. 8A, there was no systematic correlation in the overall responsiveness of these 2 cells to the 28 training stimuli. Figure 8B shows the distribution of Pearson’s correlation coefficients for the 309 pairs. The mean value is 0.13. These results show that responses to the training stimuli were determined by largely independent criteria in different cells. However, it is not likely that the cells were responding solely on the basis of the components (bar, circular disk, ellipse, triangle, square) from which we composed the training stimuli, because the response pattern of none of the cells could be explained by either the presence or absence of a particular component.

**DISCUSSION**

![Figure 6](image2.png)

**FIG. 6.** Comparison of the distribution of response magnitudes between responses of the same rank-order in individual cells recorded from the trained (filled bars) and control (open bars) monkeys.
The fact that the proportion of cells responding to the training stimuli was moderate while the tuning of their responses among the training stimuli was rather broad is consistent with the idea that the training stimuli are represented as a class of shapes rather than being scattered over the whole shape space. Thus the learned shapes recruited a limited but definite portion of the representation space in inferotemporal cortex. The remaining space would be available for the analysis of other shapes and features unrelated to the training stimuli.

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 (Hinton et al. 1988). The task in which the monkeys were trained used a fixed set of 28 shapes and thus did not require generalization. Nevertheless, 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 present 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. (Note that in referring to “partial features” we do not exclude holistic features.)

We cannot determine the precise time course of the changes because the recordings were performed >3 mo after posed to solve the delayed matching to sample task (Miller and Desimone 1994; Miller et al. 1991).

Sakai and Miyashita (1991, 1994) trained adult monkeys to discriminate among many fractal patterns and found that many inferotemporal cells responded to the learned patterns. Logothetis et al. (1995) trained adult monkeys to discriminate many wire-frame objects from each other and found many inferotemporal cells responding to the images of the learned objects. The present results are consistent with these previous findings; quantitatively, the finding that 25% of inferotemporal cells responded maximally to the learned stimuli agrees well with the result of Logothetis et al. (1995) that 28.5% of inferotemporal cells responded to some of the learned object stimuli more strongly than the control stimuli. A unique contribution of the present study is the demonstration that training increases the proportion of inferotemporal cells that respond to particular stimuli as measured against untrained controls. Because the present results were obtained in anesthetized preparations, and cells in the perirhinal cortex, which is one step downstream from TE, scarcely respond to visual stimuli in anesthetized preparation (H. Tamura and K. Tanaka, unpublished observation), the changes in selectivity were likely due to changes in the neuronal network up to TE. Vogels and Orban (1994) trained adult monkeys to discriminate between gratings of a limited range of orientations but did not find an increase in inferotemporal cells responsive to the range of orientations used in training despite an improvement in the monkeys’ discrimination performance for the trained range of orientations. As Vogels and Orban (1994) argued, it is likely that the change of responsiveness in the inferotemporal cortex occurs only in the domain of features more complex than the orientation of gratings.
the monkeys mastered the task. However, the changes observed here are likely to have occurred over a period of at least 1 mo. We base this assertion on the fact that, in one trained monkey, the stimulus set shown in Fig. 1 was introduced after the basics of the task had been mastered with simpler, colored stimuli, yet even at the shortest delay, the monkey’s performance improved for a period of >1 mo after the introduction of the second stimulus set. The mechanisms that underlie the changes observed in the present experiment may differ from those previously observed over shorter intervals in which single cells were continuously recorded (Li et al. 1993; Miller et al. 1991; Rolls et al. 1989).

The transformations that underlie the selectivity changes observed in inferotemporal cells may occur either in TE itself or at earlier stages in its afferent pathway. Kobatake and Tanaka (1994) found that TEO and V4 contain cells that selectively respond to complex features, although the proportion of such cells is smaller in these earlier areas than in TE. It is possible that the increase in cells responsive to the training stimuli might be observable in these earlier areas. The change might first occur in TE and later pervade the earlier areas. It is also possible that changes first occurred in areas downstream from TE, e.g., the perirhinal cortex, and were then transferred to TE.

The question remains as to whether the neuronal changes described in this paper required the recognition performance obtained in training. The present results, as well as most previous studies, cannot exclude the possibility that frequent exposure to the stimuli for several months is enough to cause these changes. However, in the case of developmental plasticity found in the cat primary visual cortex, there are several lines of evidence implicating that not only the passive viewing but the utilization of visual stimuli in the frame of sensorimotor interaction is necessary to evoke plastic changes in cell selectivity (Singer 1978, 1979). Although specific experiments might rule out this possibility, it is not very likely that passive exposure could effect the changes we observed in the selectivity of inferotemporal cells.

This work was supported by the Frontier Research Program of the Institute of Physical and Chemical Research. Address for reprint requests: K. Tanaka, Laboratory for Cognitive Brain Mapping, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan.

Received 25 August 1997; accepted in final form 6 April 1998.

REFERENCES


