

# Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging

Gang Wang<sup>a,b</sup>, Manabu Tanifuji<sup>a</sup>, Keiji Tanaka<sup>a,c,\*</sup>

<sup>a</sup> *RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama, 351-01, Japan*

<sup>b</sup> *Department of Physiology II, Faculty of Medicine, Kagoshima University, 8-35-1, Sakuragaoka, Kagoshima-shi, Kagoshima 890, Japan*

<sup>c</sup> *Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology (JST), 2-1 Hirosawa, Wako-shi, Saitama, 351-01, Japan*

Received 23 February 1998; accepted 6 May 1998

## Abstract

To investigate the functional organization in the monkey inferotemporal cortex, which is the last exclusively visual area along the ventral visual cortical pathway, optical imaging based on intrinsic signals was carried out. We first conducted single-cell recordings with microelectrodes and determined the features critical for the activation of single cells. For the subsequent optical imaging, each critical feature was presented, which evoked multiple dark spots. Individual spots were  $\sim 0.5$  mm in diameter and one of them covered the site of the electrode penetration at which the particular critical feature had been determined. The degree of stimulus selectivity varied from spot to spot, and from region to region even within a spot. Some regions were activated only by one of 12–16 stimuli, while others by more than three stimuli. There were spots specifically activated by faces, and the position of activation spot changed gradually along the cortical surface as the stimulus face was rotated in depth. The length of the overall region along the direction of shift of these spots was  $\sim 1$  mm. These results confirm the regional clustering of cells with similar stimulus selectivity and suggest larger units in which some parameters of object features are continuously mapped. © 1998 Elsevier Science Ireland Ltd. All rights reserved.

**Keywords:** Optical imaging; Intrinsic signals; Macaque monkey; Ventral visual pathway; Extrastriate visual cortex; Inferotemporal cortex; Object recognition; Face recognition

## 1. Introduction

The ventral visual cortical pathway, which interconnects the striate, prestriate, and inferotemporal areas, is critical for the visual identification of objects (Gross, 1973; Dean, 1976; Mishkin et al., 1983). The anterior part of the inferotemporal cortex, defined as cytoarchitectonic area TE, is the last exclusively visual area along the pathway. TE cells selectively respond to object attributes such as shape, color, texture or a combination of these (Gross et al., 1972; Desimone et al., 1984; Tanaka et al., 1991; Komatsu and Ideura,

1993; Kobatake and Tanaka, 1994; Sheinberg and Logothetis, 1997). The object features to which TE cells selectively respond are more complex than the features to which cells in the earlier stages respond (Gallant et al., 1993; Kobatake and Tanaka, 1994; Gallant et al., 1996). The size of the receptive fields increases towards TE along the ventral visual cortical pathway (Gattass et al., 1988; Boussaoud et al., 1991; Kobatake and Tanaka, 1994).

Based on single-cell recordings, Fujita et al. (1992) proposed that TE cells responsive to similar object features cluster vertically across the cortical layers to form a column. The width of individual columns was estimated to be  $\sim 0.4$  mm. To confirm the existence of such clustering of cells, and if present, to further study

\* Corresponding author. Tel.: +81-48-467-9342; Fax: +81-48-462-4651.

the spatial organization of the clustering, optical imaging based on intrinsic signals was carried out in the monkey area TE in the present study.

The optical imaging has been successfully used in the studies of the functional organization in V1, V2, MT, and more recently TE and V4 (Grinvald et al., 1988; Frostig et al., 1990; Ts'o et al., 1990; Bonhoeffer and Grinvald, 1991; Blasdel, 1992; Malonek et al., 1994; Wang et al., 1996; Ghose and Ts'o, 1997), and was shown to be superior to single-cell recordings for studying the functional organization. In optical imaging the local average of neuronal activities is measured simultaneously over a wide cortical region, whereas such measurement is difficult with single-cell recordings.

## 2. Materials and methods

Five adult Japanese monkeys (*Macaca fuscata*) were prepared for repeated experiments. The optical imaging was conducted in the dorsolateral part of TE (Fig. 1(A)) in eight hemispheres. Both hemispheres were used in succession in three of the five monkeys. For two of the eight hemispheres, single-cell recordings with microelectrodes were first conducted in several separate sessions to determine the visual stimuli effective for activation of the TE region to be subjected to the optical imaging, and clarify the correlation between

optical signals and neuronal activity. The object features critical for the activation of single cells were determined during the single-cell recordings, as described below in detail. For the final optical imaging session on these two hemispheres, we used the critical features as well as some simple stimuli which had been demonstrated to be not effective for the activation of the single cells. The interval between the single-cell recording sessions was 1 week, and that between the last single-cell recording session and the optical-imaging session was 2 weeks. Only optical imaging was conducted in the other six hemispheres.

### 2.1. Preparation

The initial surgery was carried out under anesthesia with pentobarbital sodium (35 mg/kg i.p., supplemented by 10 mg/kg if necessary). A brass block for head fixation was attached to the top of the skull, the zygomatic arch was trimmed, two stainless steel bolts for EEG recordings were implanted in the skull. For the hemispheres on which both single-cell recordings and optical imaging were carried out, the lateral surface of the skull was exposed and covered with resin at the initial surgery. After the last single-cell recording session and one week before the optical imaging session, a second surgery was carried out, in which a chamber 18 mm in inner diameter was fixed to the skull for the optical imaging. For the hemispheres on which only the optical imaging was carried out, the chamber was implanted at the initial surgery. The center of the chamber was placed 17.5 mm anterior to the ear bar position, and close to the lip of the superior temporal sulcus (Fig. 1(A)). The optics of the eyes was measured to select appropriate contact lenses, and photographs of the fundus were taken to determine the position of the fovea before the recordings.

For both single-cell recording and optical imaging, the pupils were dilated by local application of 0.5% tropicamide–0.5% phenylephrine. The corneas were covered with contact lenses of appropriate power to focus the image on a 20-inch CRT display (CMM 20-11, Shibasoku) placed 57 cm from the eyes, onto the retina. Throughout the recording session, the animals were immobilized with pancuronium bromide, and the anesthesia was maintained by artificial ventilation with a mixture of N<sub>2</sub>O and O<sub>2</sub> (70:30), supplemented with isoflurane (up to 2%). The level of anesthesia was monitored with the electrocardiogram. Atropine sulfate was subcutaneously administered to reduce salivation. The administration of pancuronium bromide was stopped at the end of recordings, and spontaneous respiration recovered a few hours after. The monkey was then returned to the home cage.

Monkeys were regularly monitored by a veterinarian and cared for in accordance with the Guiding Principles

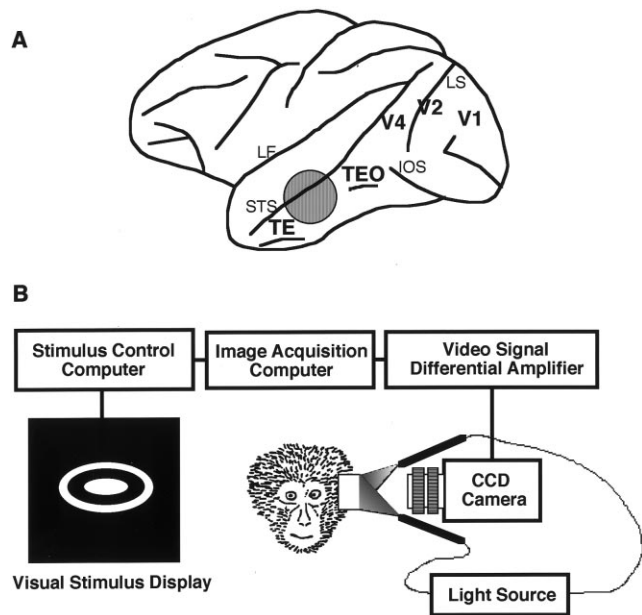


Fig. 1. The site of optical imaging and the schematic of the optical imaging system. (A) The shaded area indicates the position of recording chamber in the dorsolateral part of TE close to the lip of the superior temporal sulcus. (B) The optical imaging system consisted of a CCD camera, data acquisition system, stimulus presentation system, and a light source to generate 605 nm red light. STS: superior temporal sulcus; IOS: inferior occipital sulcus; LS: lunate sulcus; LF: lateral fissure.

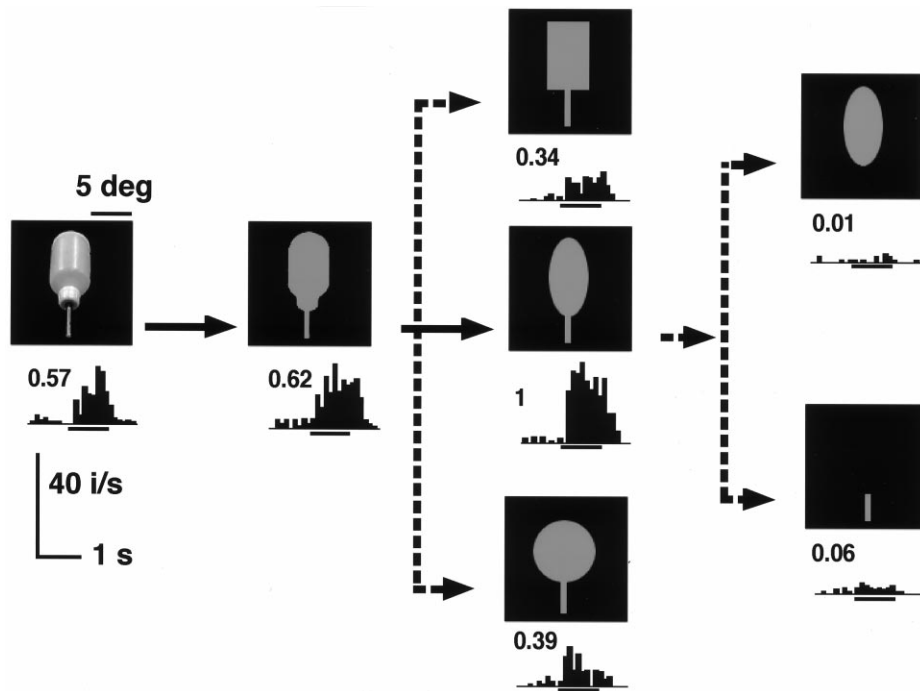


Fig. 2. An example of the reduction process to determine the critical feature for the activation of single cells. The stimulus images and responses of a TE cell are shown in combinations. The PST histograms were calculated for responses in ten trials of stimulus presentation. The numbers above histograms represent the magnitude of the responses normalized by that of the maximum response.

for the Care and Use of Animals in the Field of Physiological Science of the Japanese Physiological Society.

## 2.2. Single-cell recordings

For two of the eight hemispheres, extracellular recordings with microelectrodes of single-cell activities in TE were made according to the method described previously (Fujita et al., 1992; Ito et al., 1994; Kobatake and Tanaka, 1994; Ito et al., 1995). Recordings were limited to areas where the final optical imaging was planned for. The electrode was advanced through a pin hole made in the dura by a needle coated with black ink. The ink-coated needle was penetrated more deeply than usual so that the needle reached the cortex and the penetration site was marked directly on the cortical surface with ink. In the first hemisphere, in which we made 12 penetrations, we failed to recover the marks on the cortical surface, probably because the needle had not reached the cortical surface. In the second hemisphere, we recovered marks in 13 out of 18 penetration sites. At another penetration site, we found a depression marked with ink, which obscured the optical imaging. We could not find ink marks at the remaining four penetration sites.

After isolation of single-cell activity, the critical feature for maximal activation of the cell was determined. Dozens of three-dimensional objects were presented to

the monkey, first directly and then through a video camera (DXC-327, Sony) and the CRT display, to determine the most effective stimulus. We also determined the most effective view angle, position, size and background color (black or white). The video image of the most effective stimulus was then simplified step by step with the aid of a computer graphics system (Nexus 600-M4, Nexus), and the critical feature was determined as the feature that was the simplest but sufficient for maximal activation of the cell, as exemplified in Fig. 2 for one TE cell.

## 2.3. Optical imaging

Optical imaging was carried out without dye, based on the intrinsic signals. On the day of the optical imaging, the skull and dura in the chamber were removed to expose the region for imaging. The chamber was closed with a glass coverslip and filled with silicone oil to minimize movements of the brain during recordings. Response images were obtained by using light tuned to 605 nm wavelength (bandwidth, 10 nm). A schematic illustration of the apparatus used to record optical signals is shown in Fig. 1(B). Light from a tungsten lamp was directed through an interference filter, and guided to the chamber through a fiberoptic bundle. To illuminate the cortical surface evenly, two sets of such light devices were used. Reflected images from the cortical surface were obtained using an Imager

2001 (Optical Imaging) equipped with a charge-coupled device (CCD) camera. Two commercially available lenses with large numerical apertures were connected face to face to focus the image onto the detector chip of the CCD camera (Ratzlaff and Grinvald, 1991). The size of imaged area was adjusted by selecting an appropriate combination of lenses with different focal distances, 35 and 50 mm. The CCD camera was focused on a plane 300  $\mu\text{m}$  below the cortical surface.

A trial of recording was initiated by first synchronizing a certain phase of respiration by stopping and restarting the respirator, to a heartbeat which acts as a trigger to start data acquisition. This procedure let the recording always start at the certain phase of respiration in synchrony with a heartbeat. The visual stimulus was presented 1 s after the start of the recording, and continued for 4 s. The recording ended when the visual stimulus was withdrawn. Five to 28 visual stimuli were combined in one stimulus set and presented in a circular order. Several stimulus sets were usually presented in one experiment. In the cases in which single-cell recordings were conducted before the optical imaging, the critical features determined for single cells, simpler control features (bars, squares, and disks), and faces of different views were included in the stimulus sets. In the cases in which only the optical imaging was conducted, moderately complex features which had been determined as the critical features for the activation of TE cells in the past in this laboratory were used instead of the critical features. The interval between stimuli was set at 8–14 s to allow cortical relaxation after previous stimulation. The stimulus sequence was repeated 40 or  $28 \times$  to improve the signal/noise ratio. During the stimulus-off period, a homogenous gray was presented on the CRT display. The stimulus image suddenly appeared on the CRT display and then moved in a circular path (with a radius of  $0.4^\circ$  at a rate of 1 cycle/s) for 4 s. The background of the stimulus image on the CRT was the same as the homogeneous gray during the stimulus-off period. The homogenous gray image was presented instead of stimulus images after every six to eight stimulus to provide control cortical images without stimulation. Since it is the same as the gray during the stimulus-off period, the same homogeneous gray was continuously presented throughout the trials.

#### 2.4. Analyses

To overcome the limitation in the intensity resolution of the AD converter (8 bits), images sampled only for the 8-bit stem components of the intensity (stem images) were regularly taken. The last stem image was subtracted from the subsequent input images, and the differential image was sampled again by the AD converter to be registered. The intensity of pixels was later converted to the absolute intensity by adding the intensities of corresponding pixels in the stem image.

The images were first averaged over a 3.5-s period starting at 0.5 s after stimulus onset and ending at the stimulus withdrawal in each trial. This image was divided by the image obtained in the nearest trial without visual stimulation (gray image), to remove the spatial inhomogeneity in the background level of reflected light. The image obtained by a division between the images obtained in the 1-s periods just before the stimulus onset in the same two trials, was subtracted from the image obtained by the division between the 3.5-s period images, to remove possible slow signal changes over several minutes. The images were finally averaged across different cycles of trials to improve the signal/noise ratio.

The original raw images of individual trials were checked by visual inspection to remove the scrambled images in which the signals went out of the amplifier's dynamic range for unknown reasons. Binning ( $2 \times 2$  pixels were combined into a single pixel) in some cases, but no further spatial filtering, was carried out before the statistical evaluation of activation spots. In the quantitative comparison, the areas covering big blood vessels were excluded. Averaging (by  $3 \times 3$  pixels, with pixel sizes ranging from 35 to 57  $\mu\text{m}$ ) was carried out for some of the images shown in the figures.

### 3. Results

Data described here are based on the recordings from eight hemispheres of five monkeys. Ninety three successful images were obtained for moderately complex non-face features including the critical features, 63 for faces, and seven for simple control stimuli, besides many control images obtained without visual stimulation.

#### 3.1. Global darkening

Presentation of visual stimuli generated two kinds of optical changes in TE. One was a global darkening and the other were localized spots superposed on the global darkening. The cortical regions above the superior temporal sulcus, which are known to be occupied by the auditory association areas, were included as well as TE, which is below the sulcus, in the images in two hemispheres. Fifty three successful images were obtained, 43 for 26 moderately complex features, five for faces, and five for simple features, in these two hemispheres. All of the 53 images showed general darkening in the exposed TE area (one case is illustrated in Fig. 3). The mean intensity averaged over the whole imaged TE region was significantly darker than that of the image obtained without stimulation (paired *t*-test,  $t = 2.55$ – $4.05$  with  $df = 37$  for the 19 images in one hemisphere, all  $P < 0.01$ ; and  $t = 2.06$ – $3.56$  with  $df = 30$  for the 34 images

in the other hemisphere, all  $P < 0.025$ ). The global changes covered most of the imaged TE area: the pixels with significant darkening,  $P < 0.05$ , covered 81.1% (S.D. = 10.4%) of the imaged TE area (5.3 and 6.8 mm<sup>2</sup>). The activation pattern did not depend on the kind of visual stimuli: over 90% of the activated regions overlapped with those in images for different stimuli (mean = 96.8%; S.D. = 6.9%). The overlap was calculated by the overlapped area divided by the mean of each total area, as elsewhere in this paper.

In 49 out of the 53 images, the areas above the superior temporal sulcus did not get significantly darker

compared with the images obtained without visual stimulation ( $p > 0.05$ ). The darkening reached the significance level of 0.025 in one image, and 0.05 in the remaining three images.

### 3.2. Localized spots

The presentation of a visual stimulus evoked a few dark spots superposed onto the global darkening, distributed in the imaged TE region. To extract such localized spots, a cocktail reference image, averaged over the images obtained for all the different stimuli (excluding the images obtained without stimulation) in the series was subtracted from each of the images to remove the component of global darkening. This subtraction distributed the intensity of the images  $\sim 0$ . To describe the size and number of these local spots, we calculated  $1/e$  of the maximum intensity value in each image and delineated the spots at this value. We refer to the contours of spots as  $1/e$  contours. The maximum value used to delineate the  $1/e$  contours was the maximum over the whole image, but not the local maximum. Fig. 4 illustrates the map of  $1/e$  contours obtained with the same set of moderately complex features in two hemispheres.

To examine the reproducibility of the local dark spots, the  $1/e$  map obtained for the former half of the stimulus repetitions and that obtained for the latter half were compared. The overlap was as much as  $0.78 \pm 0.14$  (mean  $\pm$  S.D.). Considering the small number of repetitions (14 or 20) in one half of stimulus repetitions, the values indicate high reproducibility of the location of the dark spots. The two images illustrated in Fig. 3 were obtained from two recordings which were separated by 6 h.

The statistical significance of individual dark spots was further examined by combining two statistical significance tests. The first one was the ordinary  $t$ -test of the darkening of mean intensity within individual  $1/e$  contours based on the whole data obtained from all trials. In the second test,  $1/e$  contours were defined in the images averaged over the data obtained in a half of the total trials, and the darkening of mean intensity within the  $1/e$  contours was tested with  $t$ -test by using the data obtained in the other half of the trials. We used  $P < 0.01$  in the first test, and  $P < 0.075$  in the second test. Out of 251 spots (202 to non-face features and 49 to faces) defined by the  $1/e$  criterion, 218 spots (176 to non-face features and 42 to faces) satisfied both of the statistical significance criteria. In the following parts of the paper, values will be described in parallel for all the spots defined by the  $1/e$  criterion and for those which were first defined by the  $1/e$  criterion and verified by both of the statistical significance criteria.

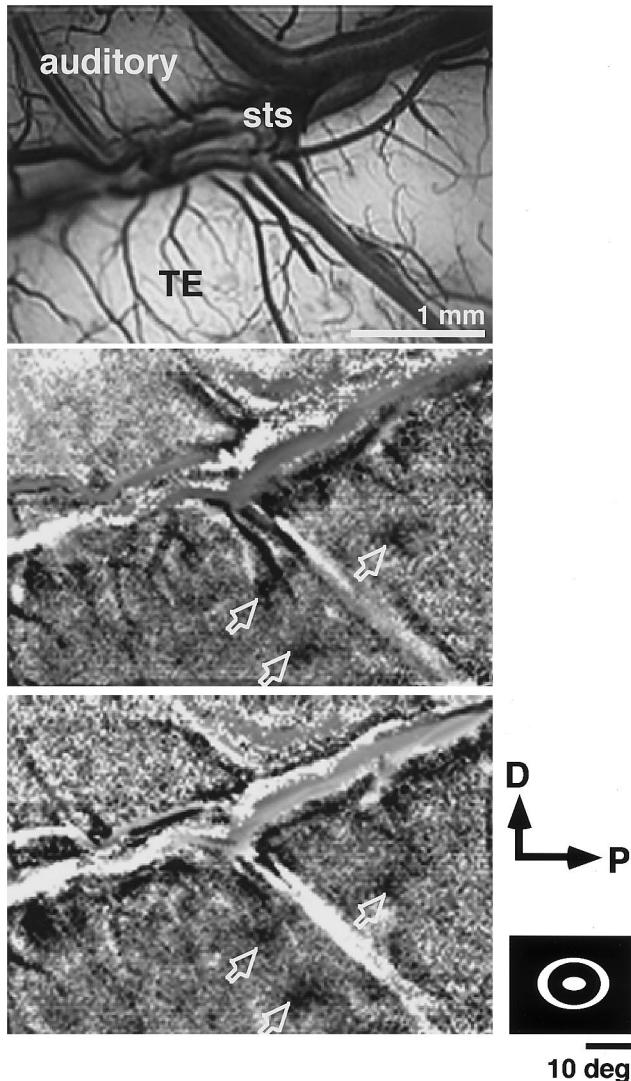


Fig. 3. Reproducibility of optical signals. Top: a view of cortical surface of the imaged area, obtained with 570 nm green light which provided the best contrast against the blood vessels. Middle and bottom: images obtained with 605 nm wavelength light during the presentation of the stimulus shown at the bottom right. The two images were obtained from the same cortical region, but the two experiments were separated by 6 h. The images were divided by those obtained during a blank screen period. Three localized dark spots were observed in TE (arrows) in addition to the general darkening. STS: superior temporal sulcus.

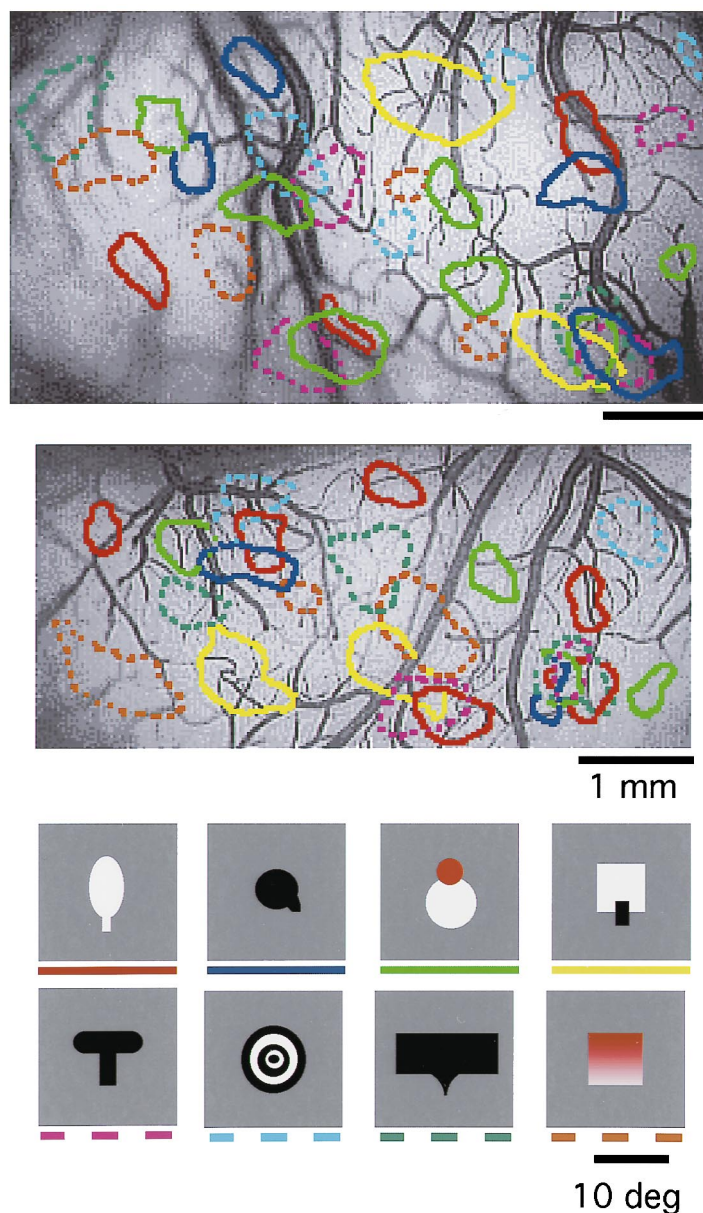


Fig. 4. Activation maps evoked by the presentation of eight moderately complex features. To obtain the map, each image was subtracted by the cocktail reference averaged over the images obtained for all the different stimuli combined in the experiment to remove the global darkening, the activation spots 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 and colors. The spots were obtained for the stimulus image underlined by bars of the same line types and colors. The two maps were obtained from two different hemispheres with the same set of stimuli.

The dark spots were usually circular or elliptic. Their sizes along the longer axis varied more from spot to spot than their sizes along the shorter axis. The size along the longer axis was  $0.62 \pm 0.55$  mm ( $0.69 \pm 0.31$  mm for the statistically significant spots), while the size along the shorter axis was  $0.53 \pm 0.15$  mm ( $0.57 \pm 0.12$  mm for the statistically significant spots) in the 93 images obtained for non-face features. The size along the longer axis was  $0.64 \pm 0.42$  mm ( $0.70 \pm 0.23$  mm for the statistically significant spots) and the size along the shorter axis was  $0.57 \pm 0.30$  mm ( $0.56 \pm 0.21$  mm for the statistically significant spots)

in the 38 images obtained for faces. These values are not significantly different between the images obtained for non-face features and those for faces.

There were usually multiple spots in the imaged TE region for a single stimulus. The number of spots per unit cortical surface area ( $1 \text{ mm}^2$ ) was  $0.35 \pm 0.29$  (all spots) or  $0.29 \pm 0.17$  (statistically significant spots) for the 93 images for moderately complex non-face features and  $0.21 \pm 0.29$  (all spots) or  $0.18 \pm 0.16$  (statistically significant spots) for the 38 images obtained for faces. The two sets of values were not significantly different.

### 3.3. Stimulus selectivity of the localized spots

The distribution of the localized spots was different for different stimuli (Fig. 4). The overlap of the  $1/e$  contours between images obtained for different non-face features was as small as  $0.34 \pm 0.16$  ( $0.31 \pm 0.15$  for the statistically significant spots).

The distribution of localized spots correlated well with the results of single-cell recordings. As we described in Section 2, we conducted single-cell recordings with microelectrodes prior to the optical imaging in two hemispheres. We marked the penetration site with ink, and determined the critical features for the activation of single cells. We failed to recover any of the markings in one hemisphere, but successfully recovered the marks of 13 penetrations in the other hemisphere. One of the dark spots evoked by each critical feature covered the penetration site at which the particular critical feature was determined for a single cell in most of the cases. In ten out of the 13 cases, the penetration site was included within the  $1/e$  contour (defined at  $1/e$  of the maximum intensity in the whole image) of one of the dark spots. Moreover, the drop in intensity from the peak of the individual spot to the penetration site was smaller than  $1/e$  of the peak in all these ten cases. In another case, the intensity at the penetration site was just outside of the  $1/e$  contour and at  $2/e$  drop from the peak of one dark spot. The dark spots in all these 11 cases were statistically significant. In the other two cases the penetration site was at a distance from dark spots. Fig. 5 illustrates the contours of the dark spots for nine critical features and the positions of eight penetrations. The other five recorded penetrations were located outside of the region shown in Fig. 5. The contours shown in Fig. 5 were delineated by the significance of darkening in individual pixels ( $t$ -test,  $P < 0.05$ ) compared with the cocktail reference image ( $t$ -test contours). Here, the  $t$ -test was applied to individual pixels, differently from the above described statistical significance test of the dark spots, where the mean intensity within the  $1/e$  contours was examined.

To further demonstrate the correlation between optical signals and cell activity, the results of single-cell recordings and optical imaging were put together in Fig. 6 for two penetrations. The critical features determined for single cells were a combination of white and green rectangles at the penetration site shown in the upper half and a combination of horizontal white and dark bars in the penetration site shown in the lower half. The critical features evoked strong responses in the cells, while simpler control stimuli evoked much weaker responses. In the optical imaging, the critical feature evoked a clear dark spot covering the penetration site at which the critical feature had been determined (marked by the red crosses), while the simpler control stimuli did not evoke dark spots around the penetration site.

Figs. 4 and 5 show that some of the TE regions were activated by multiple different visual stimuli while other regions were activated only by one particular stimulus. To quantify this difference, we counted the number of stimuli which evoked activation in individual pixels. The size of the pixels was  $54 \times 54$ – $75 \times 75 \mu\text{m}^2$  (after the binning was applied). The experiments in which successful images were obtained with no fewer than 12 stimuli (up to 16) were used in this part of data analysis, and the pixel was regarded as activated when its intensity was above  $1/e$  of the maximal intensity in individual images. As shown in Table 1, 24.4% of the pixels responded only to one stimulus, 5.1% to two stimuli, 0.5% to three stimuli, and 2.2% to more than three stimuli (up to ten). The pixels which were not activated by any of the stimuli constituted 67.8%. Slightly different values were obtained by excluding the spots which did not satisfy the statistical significance criteria (the values in parentheses in Table 1), but the above described conclusion is supported by either set of values.

### 3.4. Continuous mapping of different face views

Faces of five different views were included in some of the stimulus sets in all of the eight hemispheres. To analyze the positional relationship among the dark spots evoked by the different views, we first analyzed the images obtained by subtracting the cocktail image averaged only over the images for the faces of five different views (face cocktail) from individual images. As shown in Fig. 7 for one hemisphere, one of the dark spots considerably overlapped among the images for the different views, and the center of the spot shifted in one direction with the rotation of the face. The overlapping of dark spots evoked by five different views and the continuous shift with the rotation of the face was obtained in four hemispheres (three out of the five left hemispheres and one out of three right hemispheres). The size of individual spots defined at  $1/e$  of the maximum was  $0.58 \pm 0.08$  mm in longer axis and  $0.39 \pm 0.08$  mm in shorter axis, whereas the mean length of the overall region along the direction of the shift was  $1.28 \pm 0.16$  mm. All of these dark spots were statistically significant. In the other four hemispheres, a few dark spots were evoked by the faces, but the systematic shift with the face rotation was not observed. Moreover, there were no regions where all the face views activated nearby locations.

To further examine the nature of this shift of dark spot, we analyzed the images with the cocktail reference averaged only over the images obtained for the non-face features included in the same experiment (non-face cocktail). It is expected that the subtraction of the face cocktail emphasizes the differences in the images for different views, whereas the images obtained by the



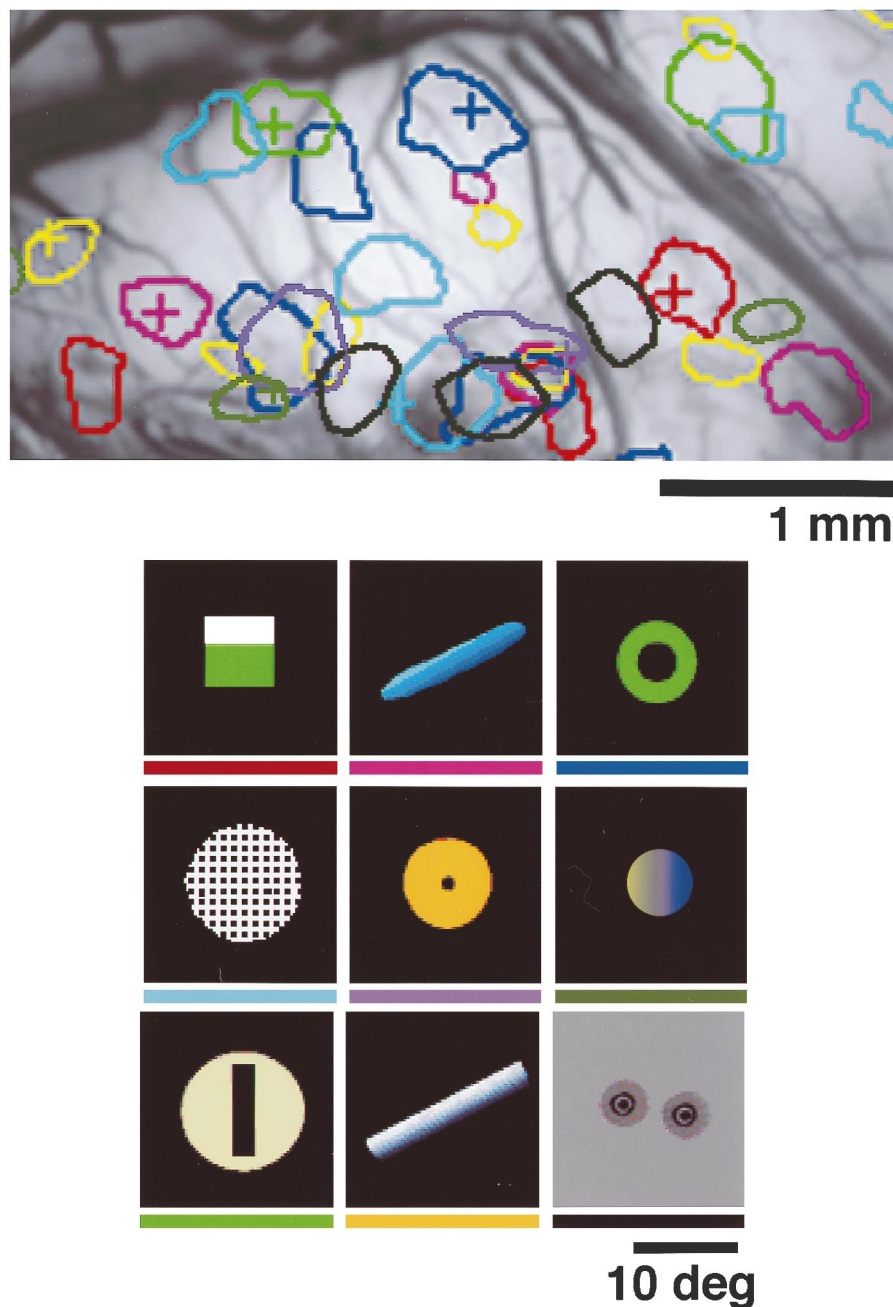


Fig. 5. Correlation between optical signals and single-cell activity. Contours of significantly activated regions for nine critical features were plotted with different colors. The contours circumscribe the pixels which had an intensity significantly ( $P < 0.05$  with  $t$ -test) greater than that of the corresponding pixels in the cocktail reference. The sites of the penetrations at which the critical features were determined are indicated by the cross marks of corresponding colors.

subtraction of the non-face cocktail demonstrate the activation components common to the different views. The analysis with non-face cocktail was applied only to the two cases in which a sufficient number of non-face features were included in the stimulus set. The results are shown in Fig. 8(A,B). In the images obtained for the non-face cocktail, although the  $1/e$  contours defined at  $1/e$  of the maximum intensity (Fig. 8(A,B) leftmost) showed a systematic shift, the contours defined at lower

levels ( $1/e^2$  for example, second from left) almost completely overlapped with one another. The  $t$ -test contours defined by the statistical significance of the activation in individual pixels at  $P < 0.05$  (Fig. 8(A,B) third from left) were similar to those defined at  $1/e^2$ . These results suggest that the activation was composed of two components: a completely overlapping, common component and a spatially differentiated component. The common component was big enough to exceed  $1/e^2$



and the threshold of  $t$ -test at  $P < 0.05$ , while the second component was big enough to show the shift in  $1/e$  contours. The  $1/e$  contours in the images obtained for the face cocktail (not shown) were very similar to those obtained for the non-face cocktail, but the  $t$ -test contours were smaller and the shift was more prominent

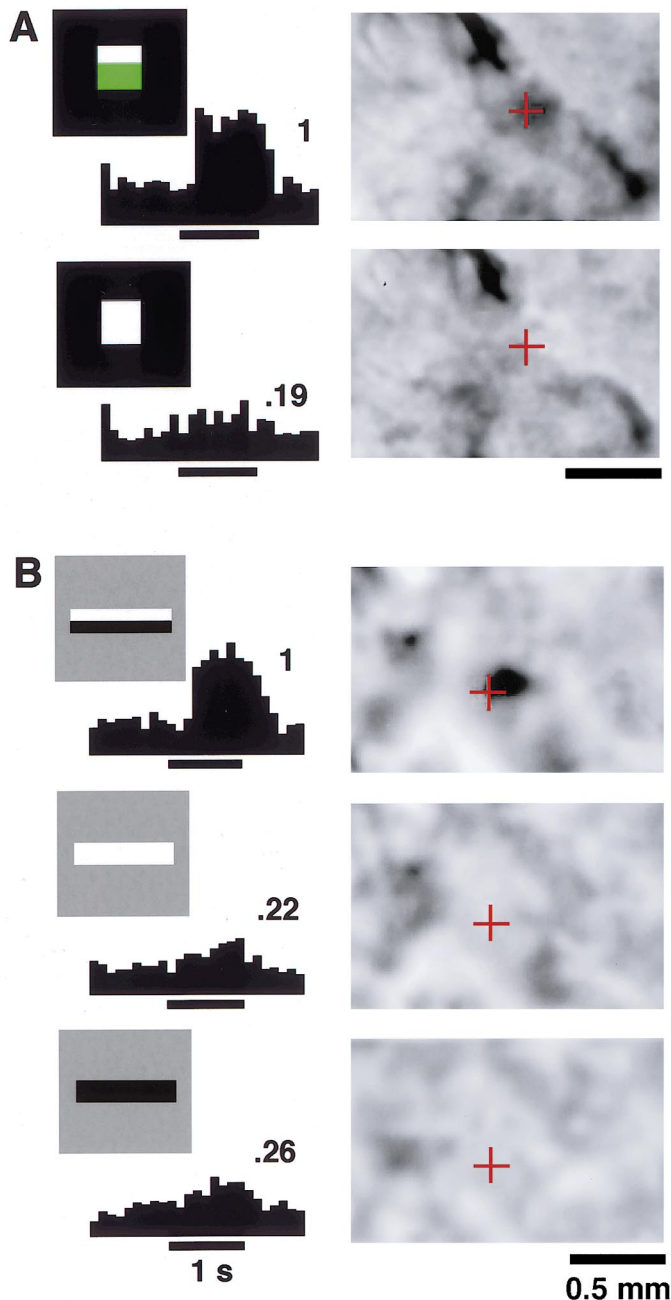


Fig. 6. Further demonstration of the correlation between optical signals and single-cell activity at two cortical sites. The stimulus images of the critical feature and simple control stimuli, responses of the cell for which the critical feature was determined, and optical images obtained with the same stimuli are aligned in lines. The red crosses in the optical images indicate the site of the penetration at which the cell was recorded. The optical images are original images without the subtraction of the cocktail reference (American Association for the Advancement of Science, 1996).

compared with the sizes of individual spots (Fig. 8(A,B) rightmost).

The spots which showed the systematic shift were more specific to faces than the other face-evoked spots which did not show a systematic shift. The averaged overlap of  $1/e$  contours between one of the face spots included in the systematic shift and spots evoked by one non-face feature was  $0.18 \pm 0.05$  ( $0.16 \pm 0.05$  with statistically significant non-face spots), whereas that between the other spots evoked by one face view and the spots evoked by one non-face feature was  $0.46 \pm 0.10$  ( $0.44 \pm 0.09$  with statistically significant non-face spots) (the overlap was calculated here by the ratio of the overlapped region to the total extent of the dark spot evoked by the face, differently from elsewhere in this paper).

The results obtained in one of the four hemispheres (Fig. 8(A)) correlated with the results of single cell recordings. All of the five cells recorded at one penetration site appeared to selectively respond to the sight of a face. A combination of the eyes and nose was sufficient for the activation of one cell, but the combination of all the essential features of the face was apparently necessary for the activation of the other four cells. Three of them responded maximally to the front view of the face, whereas the other two responded maximally to the left profile. Although we failed to mark the position of this penetration, the reference points on the bone suggested that the penetration was close to the position of the face spots which showed the systematic shift.

In one of the four experiments, different views of a doll face and those of a human face were included in the stimulus set. As shown in Fig. 8(D), the systematic shift of the dark spot was observed for both sets of faces, and the  $1/e$  contours of the dark spots evoked by the corresponding views coincided with each other, with an averaged overlap of 0.93 (S.D. = 0.04).

To exclude the possibility that the shift of the dark spots was caused by the position change of some part of the face, for example the nose, a same front view of the face was presented at three different horizontal positions. As shown in Fig. 9, essentially the same region was activated by the face regardless of the position. The commonly occupied region of the three dark spots constituted 91.8% of the average of the total extent of a single spot.

#### 4. Discussion

The visual stimuli evoked two kinds of optical signals, global darkening and localized spots, in the present study. Previous studies have shown that the intrinsic optical signals measured around 600 nm mainly reflect the concentration of deoxygenated

Table 1  
The stimulus selectivity of the activation

Number of effective stimuli	None	1	2	3	4	5	6	7	8	9	10	> 10
Proportion of pixels (%)	67.8 (72.3)*	24.4 (20.2)	5.1 (4.9)	0.5 (0.4)	0.2 (0.2)	0.1 (0.1)	0.1 (0.1)	0.2 (0.2)	0.8 (0.8)	0.7 (0.7)	0.1 (0.1)	0 (0)

\* The activation areas were delineated at 1/e of the maximum intensity in individual images, and pixels were classified according to the number of stimuli which activated the individual pixels. The number of stimuli used in these experiments were 12–16. The values in parentheses were the number of effective stimuli calculated after the spots which did not satisfy the statistical significance criteria (see the text) were excluded from the activation areas.

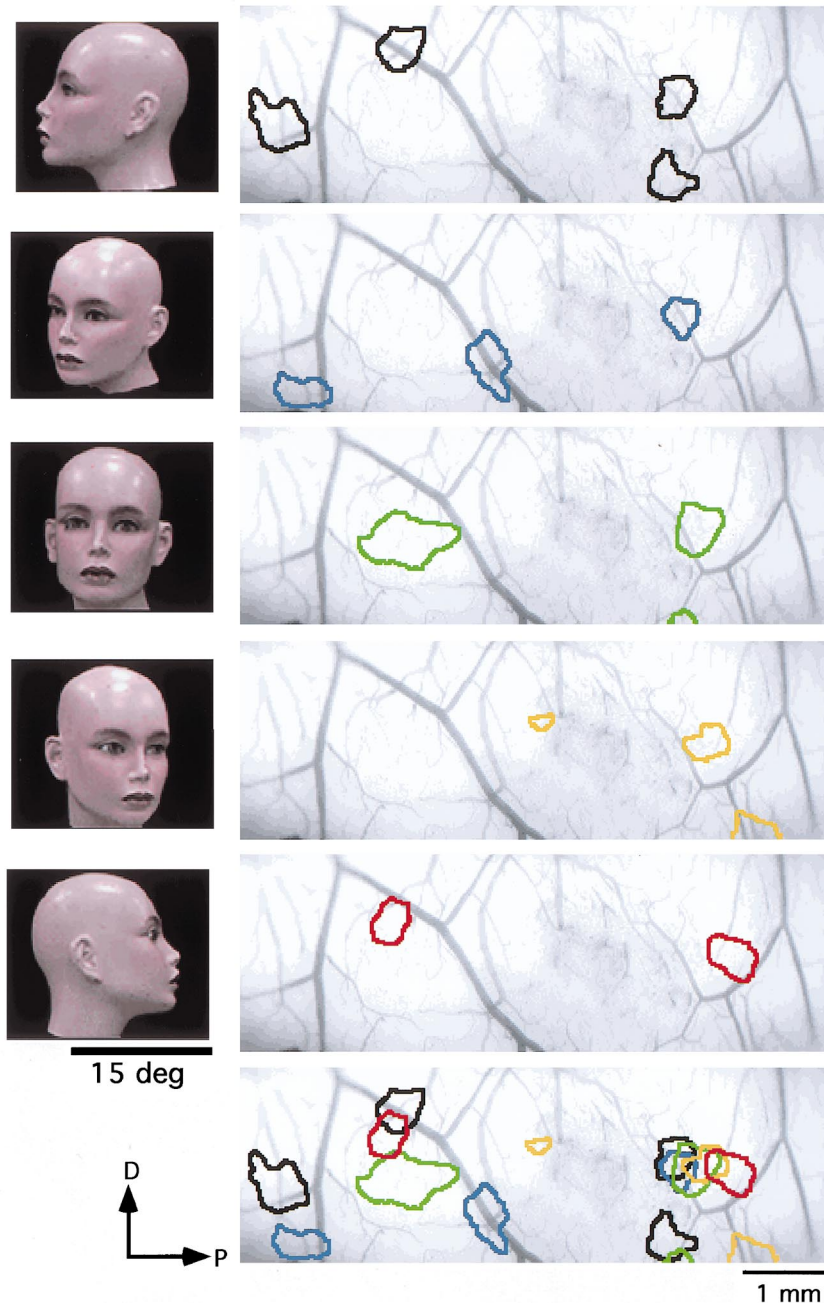


Fig. 7. Activation spots for five different views of the same doll face. The contours circumscribing the pixels with  $t$ -values giving  $P < 0.05$ , compared with the cocktail reference image, are shown. The contours of spots for the different views are superimposed at the bottom.

hemoglobin in capillaries (Frostig et al., 1990; Malonek and Grinvald, 1996), which in turn reflects the average level of synaptic activity in the particular area (Das and Gilbert, 1995). Because: (1) the global darkening covered more than 90% of the exposed TE area; (2) the location of the localized spots was largely different for different stimuli; and (3) the localized spots were well correlated with the single-cell activity, we concluded that the localized spots, but not the global darkening, are correlated with the suprathreshold synaptic activity which evokes spiking activity. The global darkening

likely represents subthreshold synaptic activity, although movements of deoxyhemoglobin and expansion of the blood volume increase from activated regions to non-activated regions may partially contribute to it.

The stimulus-specific localized activation spots confirm the regional clustering of TE cells with similar selectivity, which was suggested on the basis of the results of single-cell recordings (Fujita et al., 1992). The observable metabolic changes should have been evoked only when a large proportion of neuronal cells in the local region were activated. Although the spatial resolu-

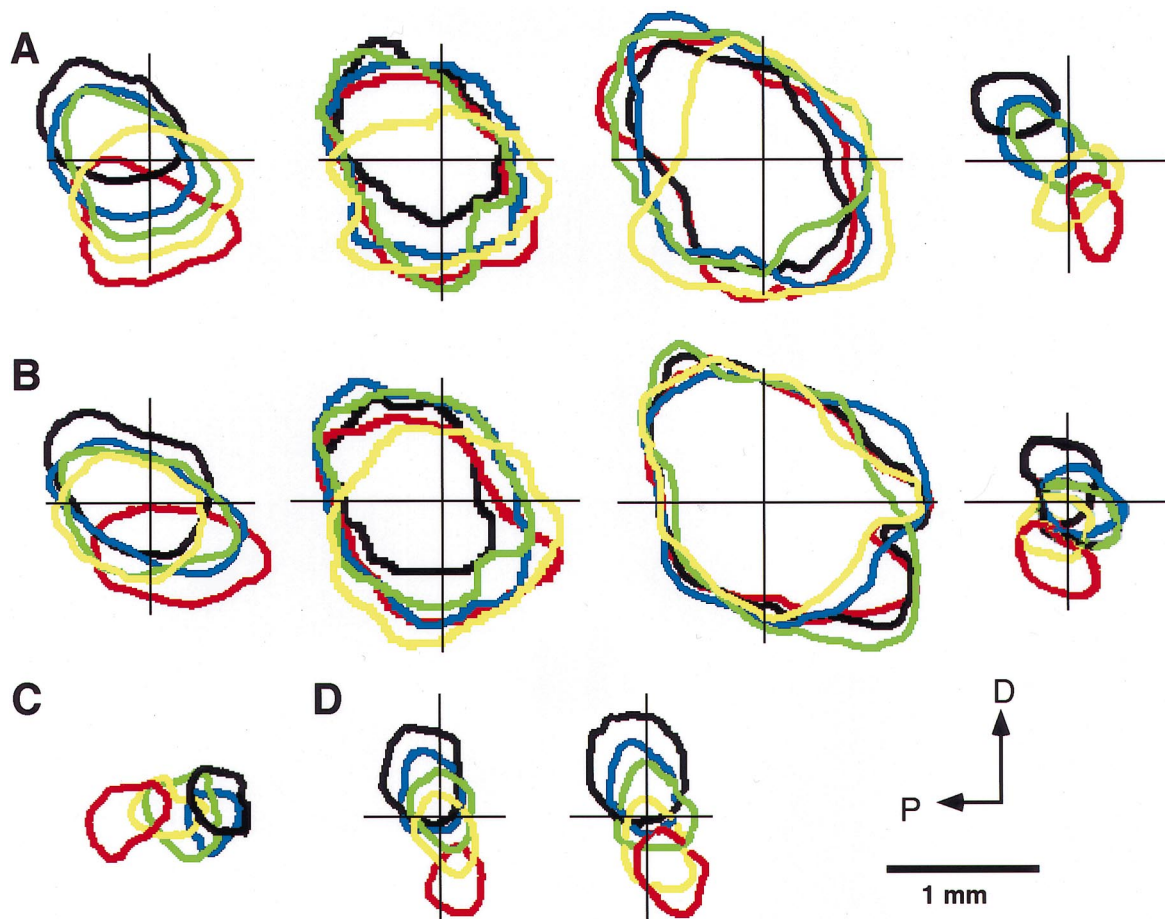


Fig. 8. Systematic shift of activation spot with the rotation of the face. The contours of dark spots are shown for four cases. (A,B) The  $1/e$  contours defined at  $1/e$  of the maximum intensity in the images obtained by subtraction of the non-face cocktail references (leftmost), the contours defined at  $1/e^2$  of the maximum intensity in the images obtained by subtraction of the non-face cocktail references (second from left), the  $t$ -test contours circumscribing the pixels with intensity significantly ( $P < 0.05$  with  $t$ -test) greater than that of the corresponding pixels in the non-face cocktail reference averaged only over the images for non-face features included in the stimulus set (third from left), and the  $t$ -test contours obtained in reference to the face cocktail reference averaged over the images obtained for the five different views of the doll face (rightmost). (C) The  $t$ -test contours obtained in reference to the face cocktail reference. The same case as that shown in Fig. 7. (D) The  $t$ -test contours for the face cocktail reference, obtained for a set of human faces (left) and the set of the doll faces (right). Only the doll faces were used in the cases shown in parts A, B, and C. Correlation between the different sets of contours can be examined by aligning the straight lines.

tion of the optical imaging with intrinsic signals is limited by the density of capillaries (their intervals are  $\sim 50 \mu\text{m}$ ) and blurring of optical images from defocused depths, structures larger than  $100 \mu\text{m}$ , like columns, could be sufficiently well resolved.

The greatest advantage of optical imaging over single-cell recordings is that the activity over a large cortical region can be monitored at a single point in time. By using this advantage, we showed several properties of the columnar organization in TE. First, a single feature evoked multiple activation spots within the imaged TE region. Second, there were regions with different levels of stimulus selectivity: some regions were activated by only one of 12–16 stimuli, whereas some other regions by more than three stimuli. Thirdly, different views of faces activated partially overlapping regions, and these regions were arranged to compose a

continuous map of the view angle along the cortical surface.

The presence of multiple spots does not necessarily indicate that all of them are specifically evoked by complex features. There is the possibility that some spots were activated by simpler features contained in the complex feature. Fig. 6 shows that this was the case for at least some of the spots. The presence of regions activated by multiple stimuli also suggests the presence of such less selective regions. Even within a single spot, the degree of selectivity may vary from part to part, because the overlapping was usually partial. The face-evoked spots which did not show systematic shift with the face rotation were likely activated by partial features of the faces common to non-face images, because they showed large overlaps with spots evoked by non-face features. The face-evoked spots which showed



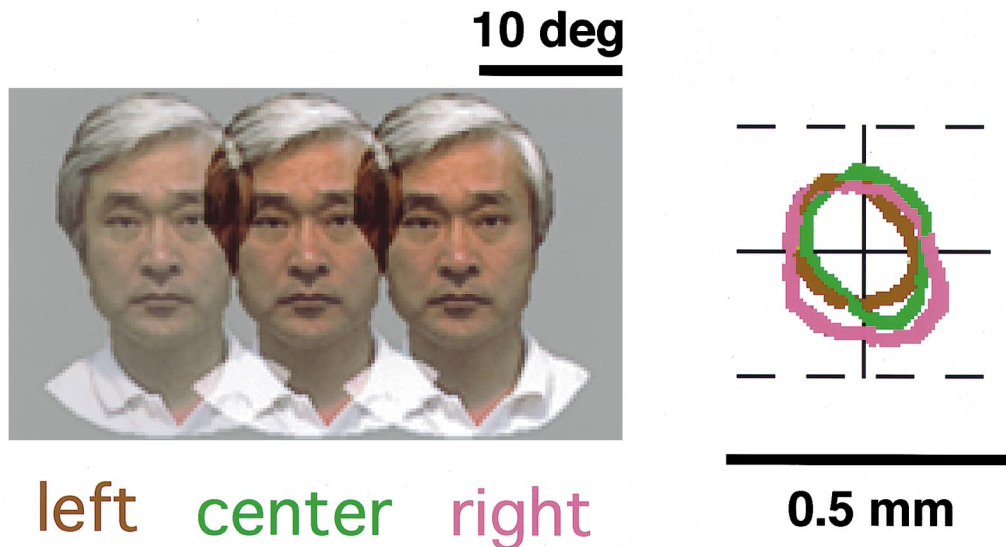


Fig. 9. Spots activated by the same front view of the face but at three different horizontal positions separated by 10°. The green spot was activated by the face at the center, the brown at the left, and the pink at the right. These images were obtained at the same cortical site as that shown in Fig. 8(D), and the green spot here is the same as the green spot in Fig. 8(D) left.

systematic shift were likely the machinery specifically developed to process faces, because of the minimum overlapping with the spots evoked by non-face features and the systematic shift with the face rotation.

The analyses of the face-specific activation with two kinds of cocktail references suggested that the activation was composed of two components. One was common to all face views, and the other was spatially discrepant among the different views. It was previously reported that responses of single TE cells have selectivity for the view. Some cells responded maximally to the front view of the face and others to the left or right profile (Harries and Perrett, 1991; Perrett et al., 1991; Tanaka et al., 1991). However, the selectivity was usually not complete: the cells which maximally responded to the front view of the face, for example, responded to the profile images with smaller but non-zero responses. These submaximal responses may have caused the common activation component. It is also possible that subthreshold excitations conveyed through mutual excitatory connections between the view-specific regions underlay the first component (Das and Gilbert, 1995).

The continuous mapping of different face views may be involved in the view-invariance of face recognition. One face view may be generalized to neighboring views through the horizontal excitatory interactions, and one of the generalized views is selected through the inhibitory interactions. Although the corresponding views of different faces activated mostly overlapping regions, this does not rule out the possibility of differential representation of individual faces at the level of single cells. Previous single-cell recordings have shown that individual TE cells have sensitivity to individual differences of faces, although the tunings were broad (Perrett

et al., 1984; Baylis et al., 1985; Yamane et al., 1988; Young and Yamane, 1992). If we assume specificity in the horizontal excitatory connections, the generalization will preserve the specificity for individuals.

This argument for faces cannot be extended to the view-invariant recognition of other non-face objects. Previous single-cell recordings have shown that non-face objects are not represented by single groups of cells, but by a combination of multiple groups of cells representing partial (both holistic and local) features of the object images (Tanaka et al., 1991; Fujita et al., 1992; Ito et al., 1994; Kobatake and Tanaka, 1994; Ito et al., 1995; Logothetis et al., 1995). Thus, the presence of a single continuous map does not adequately explain the mechanisms of the view invariance for non-face objects.

### Acknowledgements

This study was supported by the Frontier Research Program/RIKEN, Brain Foundation/Kowa Life Science Foundation, and Brain Science Foundation.

### References

- American Association for the Advancement of Science, 1996. Science 272, 1665–1668.
- Baylis, G.C., Rolls, E.T., Leonard, C.M., 1985. Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. Brain Res. 342, 91–102.
- Blasdel, G.G., 1992. Orientation selectivity, preference, and continuity in monkey striate cortex. J. Neurosci. 12, 3139–3161.

- Bonhoeffer, T., Grinvald, A., 1991. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature* 353, 429–431.
- Boussaoud, D., Desimone, R., Ungerleider, L.G., 1991. Visual topography of area TEO in the macaque. *J. Comp. Neurol.* 306, 554–575.
- Das, A., Gilbert, C.D., 1995. Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. *Nature* 375, 780–784.
- Dean, P., 1976. Effects of inferotemporal lesions on the behavior of monkeys. *Psychol. Bull.* 83, 41–71.
- Desimone, R., Albright, T.D., Gross, C.G., Bruce, C., 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* 4, 2051–2062.
- Frostig, R.D., Lieke, E.E., Ts'o, D.Y., Grinvald, A., 1990. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high-resolution optical imaging of intrinsic signals. *Proc. Natl. Acad. Sci. USA* 87, 6082–6086.
- Fujita, I., Tanaka, K., Ito, M., Cheng, K., 1992. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360, 343–346.
- Gallant, J.L., Braun, J., 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., 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.
- Gattass, R., Sousa, A.P.B., Gross, C.G., 1988. Visuotopic organization and extent of V3 and V4 of the macaque. *J. Neurosci.* 8, 1831–1845.
- Ghose, G.M., Ts'o, D.Y., 1997. Form processing modules in primate area V4. *J. Neurophysiol.* 77, 2191–2196.
- Grinvald, A., Frostig, R.D., Lieke, E., Hildesheim, R., 1988. Optical imaging of neuronal activity. *Physiol. Rev.* 68, 1285–1366.
- Gross, C.G., Rocha-Miranda, C.E., Bender, D.B., 1972. Visual properties of neurons in inferotemporal cortex of the macaque. *J. Neurophysiol.* 35, 96–111.
- Gross, C.G., 1973. Visual functions of inferotemporal cortex. In: Jung, R. (Ed.), *Handbook of Sensory Physiology*, vol. 7. Springer, Berlin, pp. 451–482.
- Harries, M.H., Perrett, D.I., 1991. Visual processing of faces in temporal cortex: physiological evidence for a modular organization and possible anatomical correlates. *J. Cogn. Neurosci.* 3, 9–24.
- Ito, M., Fujita, I., Tamura, H., 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., Tanaka, K., 1995. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* 73, 218–226.
- 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.
- Komatsu, H., Ideura, Y., 1993. Relationships between color, shape, and pattern selectivities of neurons in the inferior temporal cortex of the monkey. *J. Neurophysiol.* 70, 677–694.
- Logothetis, N.K., Pauls, J., Poggio, T., 1995. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563.
- Malonek, D., Tootell, R.B.H., Grinvald, A., 1994. Optical imaging reveals the functional architecture of neurons processing shape and motion in owl monkey area MT. *Proc. R. Soc. Lond. B* 258, 109–119.
- Malonek, D., Grinvald, A., 1996. Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy, implications for functional brain mapping. *Science* 272, 551–554.
- Mishkin, M., Ungerleider, L.G., Macko, K.A., 1983. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417.
- Perrett, D.I., Smith, P.A.J., Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., Jeeves, M.A., 1984. Neurons responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. *Hum. Neurobiol.* 3, 197–208.
- Perrett, D.K., Oram, M.W., Harries, M.H., Bevan, R., Hietanen, J.K., Benson, P.J., Thomas, S., 1991. Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Exp. Brain Res.* 86, 159–173.
- Ratzlaff, E.H., Grinvald, A., 1991. A tandem-lens epifluorescence macroscope, hundred-fold brightness advantage for wide-field imaging. *J. Neurosci. Methods* 36, 127–137.
- Sheinberg, D.L., Logothetis, N.K., 1997. The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. USA* 94, 3408–3413.
- Tanaka, K., Saito, H., Fukada, Y., Moriya, M., 1991. Coding visual imaging of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* 66, 170–189.
- Ts'o, D., Frostig, R.D., Lieke, E.E., Grinvald, A., 1990. Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science* 249, 417–420.
- Wang, G., Tanaka, K., Tanifuji, M., 1996. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272, 1665–1668.
- Yamane, S., Kaji, S., Kawano, K., 1988. What facial features activate face neurons in the inferotemporal cortex of the monkey? *Exp. Brain Res.* 73, 209–214.
- Young, M.P., Yamane, S., 1992. Sparse population coding of faces in the inferotemporal cortex. *Science* 256, 1327–1331.