JOURNAL OF NEUROPHYSIOLOGY
Vol. 72, No. 3, September 1994. *Printed in U.S.A.*

Translation Invariance in the Responses to Faces of Single Neurons in the Temporal Visual Cortical Areas of the Alert Macaque

MARTIN J. TOVEE, EDMUND T. ROLLS, AND PAUL AZZOPARDI
University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, United Kingdom

SUMMARY AND CONCLUSIONS

1. The responses of single neurons in the inferior temporal cortex and the cortex in the banks of the anterior part of the superior temporal sulcus of three awake, behaving macaques were recorded during a visual fixation task. Stimulus images subtending 17 or 8.5° were presented in the center of the display area, and fixation was either at the center of the display area, or at one of four positions that were on the stimulus, or several degrees off the edge of the test stimulus. The experiments were performed with face-selective cells, and the responses were compared for fixation at each position for both effective and noneffective face stimuli for each cell.

2. The firing rates of most neurons to an effective image did not significantly alter when visual fixation was as far eccentric as the edge of the face, and they showed only a small reduction when the fixation point was up to 4° from the edge of the face. Moreover, stimulus selectivity across faces was maintained throughout this region of the visual field.

3. The centers of the receptive fields of the cells, as shown by the calculated "centers of gravity," were close to the fovea, with almost all being within 3° of the fovea.

4. The receptive fields of the cells typically crossed the vertical midline for at least 5°.

5. Information theory procedures were used to analyze the spike trains of the visual neurons. Nearly six times more information was carried by these neurons' firing rate about the identity of an image than about its position in the visual field. Thus the information theory analysis showed that the responses of these neurons reflected information about which stimulus had been seen in a relatively translation invariant way.

6. Principal component analysis showed that principal component 1 (PC1) is related primarily to firing rate and reflected information primarily about stimulus identity. (For identity PC2 added only 14% more information to that contained in PC1.) Principal component 2 (PC2) was more closely related to neuronal response latencies, which increased with increasing eccentricity of the image in the visual field. PC2 reflected information about the position of the stimulus in the visual field, in that PC2 added 109% more information to that contained in PC1 about the position of the stimulus in the visual field.

7. These findings show that in the temporal cortical visual areas there exists a population of visual cells whose responses are largely translation invariant, and that stimulus selectivity is maintained independently of retinal position for at least several degrees of the visual field near the fovea.

INTRODUCTION

There are two main visual processing streams: the dorsal stream, which is concerned with the relative spatial position of an object, and the ventral stream, which is concerned with its identification (Mishkin et al. 1983). Damage to the dorsal stream, through which the striate cortex projects to the parietal lobes, produces impairments on visuospatial tasks (Mishkin et al. 1982; Pohl 1973). Damage to the ventral system, through which the striate cortex projects to the temporal lobes, does not impair performance of visuospatial tasks but does impair performance of object discrimination tasks (Cowey and Gross 1970; Gaffan et al. 1986; Weiskrantz and Saunders 1984). As visual information passes through the ventral system, a representation of objects is built up that becomes in the inferior temporal visual cortex and cortex in the anterior part of the superior temporal sulcus relatively invariant with respect to the size, spatial frequency, contrast, color, motion, and luminance of the object (Rolls 1992a; Rolls and Baylis 1986; Rolls et al. 1985; Sary et al. 1993). In contrast, cells in the striate cortex, V1, at the start of the processing stream, are much more finely tuned to size/spatial frequency, with a typical spatial frequency tuning of 1.5 octaves. It is hypothesized that along the ventral processing stream representations of objects are built that become by the inferior temporal cortex and cortex in the anterior part of the superior temporal sulcus invariant with respect to size, rotation, and view, because such invariant representations form appropriate inputs to associative neuronal networks in the hippocampus and amygdala, which appear to be involved in an intermediate term "on the fly" buffer store and in stimulus-reinforcement association memory (see, e.g., Rolls 1990a, 1992a,b; Treves and Rolls 1994). Consistent with this, lesions of the inferior temporal visual cortex impair the ability of monkeys to respond to objects irrespective of changes in size, lighting, and viewing angle (Weiskrantz and Saunders 1984).

It is likely that invariant representations across retinal position are also built along the same processing stream (Gross and Mishkin 1977; Rolls 1992a; Wallis et al. 1993). Consistent with this, the receptive fields of single cells in V1 are usually small (e.g., 0.5–1°). In contrast, the receptive fields of neurons in the inferior temporal visual cortex are very large, almost always including the fovea, and often extend to include large areas of the contralateral visual field (Desimone and Gross 1979; Desimone et al. 1985; Gross et al. 1972; Schwartz et al. 1983). However, these studies on inferior temporal neurons were performed under anesthesia and pharmacological immobilization, and it is not clear that the neuronal responses found will reflect those that occur when the cortex is operating normally, while the animal is awake and behaving (Rodman et al. 1991). Nor have previous studies focused explicitly on the translation invariance of face-selective neurons, for which there is now a great deal of evidence on their responsiveness and tuning

(see Rolls 1992a). This group of neurons is of especial interest in relation to the study of translation invariance, for, although it is probably desirable that their responses should be globally translation invariant, it is nevertheless the case that their responses must reflect the local spatial arrangement of features, which must be in the correct spatial arrangement for correct face recognition, and for the normal responses to be obtained from many of these face-selective neurons (Perrett et al. 1982; Yamane et al. 1988; Rolls et al. 1994). Further, we know that some face-selective neurons can respond to individual features present in a face, such as the eyes or mouth (Perrett et al. 1982). Does this mean that face-selective neurons would have best responses when the monkey is fixating one of the features in a face, such as the eyes, or would the responses of the neurons nevertheless be translation invariant? To answer these questions we investigated the extent to which face-selective neurons in the temporal cortical visual areas show translation invariance. The awake, behaving monkeys performed a visual fixation task, and eye position was measured with the search coil technique, to ensure that the stimuli were presented in known positions within the visual field.

The neurons investigated in this study were the temporal cortical neurons with responses that are different to different faces, because this enabled us to investigate whether the selectivity was maintained across retinal translation; because these neurons can be found regularly on tracks made into the temporal cortical visual areas; and because many of the response properties of these neurons have been described (Baylis et al. 1987; Rolls 1992a). This investigation is part of a series performed to understand the normal functioning of the cerebral cortex, and how disorders of its function lead to perceptual and cognitive dysfunctions (Rolls 1992a,c, 1994; Rolls et al. 1994; Rolls and Tovee, 1994; Tovee and Rolls 1992; Tovee et al. 1993).

METHODS

Recording techniques

The activity of single neurons was recorded with glass-insulated tungsten microelectrodes (after Merrill and Ainsworth 1972, but without the platinum plating) in three alert macaque monkeys (*Macaca mulatta,* weight 3.0 kg) seated in a primate chair with the use of techniques that have been described previously (Rolls et al. 1976, 1990; Tovee et al. 1993). The action potentials of single cells were amplified with the use of techniques described previously (Rolls et al. 1976), were converted into digital pulses with the use of the trigger circuit of an oscilloscope, and were analyzed on-line with the use of a MicroVaxII computer. The computer collected peristimulus rastergrams of neuronal activity for each trial and displayed, printed, and stored each trial, as well as computing the peristimulus time histogram by summing trials of a given type. Eye position was measured to an accuracy of 0.5° with the search coil technique (Judge et al. 1980), and steady fixation of a position on the monitor screen was ensured by use of a (blink version of a) visual fixation task. The timing of the task is described below. The stimuli were static visual stimuli presented at the center of the video monitor placed at a distance of 1.0 m from the eyes. A full-size face image typically subtended 17° in the visual field. The fixation spot position was either at the center of the screen, or at one of a set of four positions arranged diagonally (upper left, etc.) as shown in Fig. 2. The normal distance between F2 and F1 in Fig. 2 was 11°, but different values were tested in different experiments. The monitor was viewed binocularly, with the whole screen visible to both eyes.

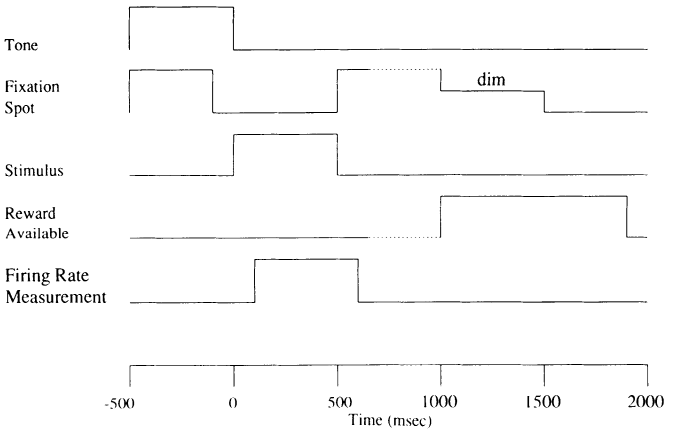

FIG. 1. Timing used in the visual fixation blink task. The fixation spot blinked off 100 ms before the visual stimulus was shown. Dotted lines represent the random interval before the fixation spot dimmed (see text).

Task timing

This is shown diagrammatically in Fig. 1. Each trial started at −500 ms (with respect to the onset of the test image) with a 500-ms warning tone to allow fixation of the fixation point, which appeared at the same time. At −100 ms the fixation spot was blinked off so that there was no stimulus on the screen in the 100-ms period immediately preceding the test image. The screen in this period, and at all other times including the interstimulus interval, was set at the mean luminance of the test images. At 0 ms the tone was switched off, and the test image was switched on for 500 ms. At the termination of the test stimulus, the fixation spot reappeared, and then, after a random interval in the range 150–3,350 ms, it dimmed, to indicate that licking responses to a tube in front of the mouth would result in the delivery of fruit juice. The dimming period was 500 ms, and after this, the fixation spot was switched off, and reward availability terminated 500 ms later. The monkey was required to fixate the fixation spot in that if he licked at any time other than when the spot was dimmed, saline instead of fruit juice was delivered from the tube; in that the dimming could only be detected if the monkey fixated the spot; and in that if the eyes moved by 0.5° from *time 0* until the start of the dimming period, then the trial was aborted. (When a trial aborted, a high-frequency tone sounded for 0.5 s, no reinforcement was available for that trial, and the intertrial interval was lengthened from 8 to 11 s.)

Stimuli

Faces were used as test stimuli. The faces included both macaque and human faces. Examples have been shown in Rolls et al. (1985) and Leonard et al. (1985). The stimuli were stored in digital form on a computer disk and displayed on a monochrome video monitor with the use of a video framestore (Advanced Electronic Design 512). The resolution of these images was 256 wide by 256 high with 256 gray levels. At all times the mean luminance of whatever was displayed on the monitor (even when it was a blank screen) was set to gray level 128, so that there were no luminance changes during each trial. The monitor provided maximum and minimum luminances of 300 and 15 cd/m^2, respectively, and was adjusted internally and by use of a lookup table for linearity to within 3% with the use of a photometer. The computer randomized the order of presentation of the test stimuli, switched the stimuli on and off for each trial, and synchronized its data

collection so that the stimulus was turned on at the start of the 21st bin of the peristimulus time histogram.

During initial investigation of the responsiveness of a temporal cortical cell, when digitized visual stimuli were being presented on the video monitor, 1 set of 4–12 visual stimuli were used at a time. Each set of stimuli was designed to provide neuronal response data relevant to one or several hypotheses. For example, one set included five different faces, to test whether the neuron responded differently to different faces, and some nonface stimuli such as a sine-wave grating, a boundary curvature descriptor, and a complex visual image (see Baylis et al. 1985, Fig. 1), to provide an indication of whether the neuron responded differently to face and to nonface stimuli. The computer randomized the sequence in which the members of the set were presented, and after it had presented the sequence once, it restarted the set with another random sequence. The computer was allowed to repeat the set 4–10 times to provide sufficient data for an analysis of variance so as to determine whether the neuron responded differently to the different stimuli within the set.

Procedure

As tracks were made into the cortex in the superior temporal sulcus, the responses of each neuron were measured to a standard digitized set of stimuli of different faces and of nonface stimuli (Baylis et al. 1985). If a neuron responded to one or more of the faces, but to none of the nonface stimuli in the set, then a wide range of digitized and real three-dimensional (3D) nonface stimuli were shown, to determine whether the response of the neuron was selective for faces. The criterion was that the response to the optimal face stimulus should be more than twice as large as to the optimal nonface stimulus. [In fact, the majority of the neurons in the cortex in the superior temporal sulcus classified as showing responses selective for faces responded much more specifically than this. For half these neurons, their response to the most effective face was 5 times as large as to the most effective nonface stimulus, and for 25% of these neurons, the ratio was 10:1. These ratios show that, although responding preferentially to faces, these neurons do not have absolute specificity for faces. Further information on and discussion of the extent to which these neurons have selective responses is given by Baylis et al. (1985) and by Rolls (1984, 1992a,c). The nonface stimuli from which the optimal was chosen included sine-wave gratings, boundary curvature descriptors, complex 2D stimuli, and complex 3D junk objects, as described above.]

If the neuron satisfied the criterion for face selectivity, then the following standard experiment was performed to assess how the fixation position relative to the stimulus affected the neuronal responses. Two or more face stimuli were selected, including one effective and one relatively ineffective in driving the cell, so that we could determine whether the cell maintained its stimulus selectivity when the stimuli were placed in different parts of the visual field. The order in which these stimuli were presented within a sequence for any fixation position was randomized. One of the five fixation positions was selected by the computer, with the use of a random order for the selection of fixation positions. When a fixation position had been chosen, all the stimuli were shown once in random order, in the center of the screen. Then a new fixation position was chosen from the random order, and the set of stimuli was again repeated. After the computer had selected all the five fixation positions (and tested all stimuli at each position), it ran further series of data collection trials, with new fixation position and stimulus order sequences. The trials followed each other with a fixed intertrial interval, and all the testing and neuronal spike collection was fully automatic. When 6–12 series had been run, so that each stimulus had been shown 6–12 times for every fixation position, that test was complete. All the neurophysiological data were saved to disk for subsequent analysis.

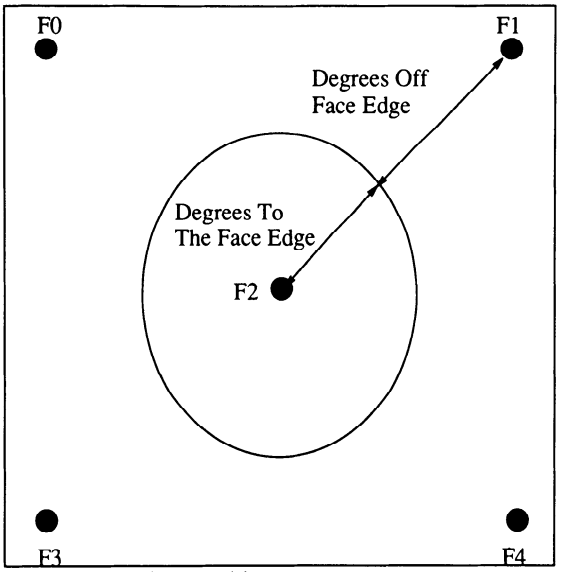

FIG. 2. Arrangement of the stimuli and fixation spots on the screen of the video monitor. The face stimulus extended from the center of the screen to an eccentricity of 8.5°. In some experiments the stimuli were ½ (approximately as illustrated) or ¼ (linearly) of the screen. The fixation points in the top left, etc., of the screen were normally 11° from the center of the stimulus, but this eccentricity was altered during some experiments.

Further tests were then run with the same cell. The first test typically involved a full-size face, which subtended 17° at the retina, and the fixation positions were as shown in Fig. 2, at the center of the face, and at the top left, top right, etc., edges of the face. Subsequent tests frequently involved half-size faces, so that the fixation points were several degrees off the edge of the face; or different fixation points on a full-size face that were still diagonally arranged as shown in Fig. 2, but were at a different angle of eccentricity from the center of the screen.

Data analysis

The default period for which the firing rate was calculated was a 500-ms period starting 100 ms after the onset of the target stimulus. (The period was chosen to start at 100 ms because all the neurons started to respond strongly by this time. Typical response latencies for the neurons were 70–90 ms.) The mean firing rate over the 6–12 trials for each fixation point/stimulus condition together with the standard error of the mean was calculated by the computer for each of the stimulus conditions for graphic presentation. In the figures, the response of the neurons, that is the firing rate minus the spontaneous firing rate, is shown. (The spontaneous firing rates of the neurons analyzed did not differ for different fixation positions. The spontaneous firing rate was measured in the 200-ms period preceding the stimulus and was averaged across all stimuli and trials.) In addition, a two-way analysis of variance (ANOVA) was performed over the same set of data, with one factor stimulus type (face 1, 2, etc.), and the other fixation position. In the ANOVAs, each category within a factor was one of five fixation positions, arranged as shown in Fig. 2.

Further, the information about the stimulus type and about the fixation position contained in the responses of the cell was computed with the use of techniques that have been described fully previously (Optican et al. 1991; Tovee et al. 1993). The general procedure was to smooth the spike train for a single trial with a Gaussian filter (with a σ of 10 ms) to produce a spike-density

function. The spike-density function was then sampled every 10 ms over 400 ms starting at the onset of the visual stimuli, to produce a 40-point time series for each trial. At least five such time series were collected for each stimulus in the stimulus set. From these time series, the principal components were extracted. The principal components are extracted from all individual responses of a neuron to all stimuli. The principal components form a basis set such that the covariance matrix is diagonal, and are ordered so that each component accounts for more variance than any subsequent one. The response of a neuron to a particular stimulus can then be fully described as a weighted sum of these principal components. The proportion of the variance accounted for by each principal component can be taken as a measure of its importance. The information was calculated from the first few principal components, or from the firing rate of the neuron, with the use of a procedure that enables calculation of the average information $I(S,R)$ contained in a set of responses R of a particular cell about the set of stimuli, S, according to the following

$$I(S,R) = \sum_{s \subset S} \sum_{r \subset R} P(s,r) \log_2 \frac{P(s,r)}{P(s)P(r)}$$

where P is the probability of occurrence of a particular event. This raw information measure was then corrected for the limited number of data trials as described by Tovee et al. (1993). The information was calculated for each experiment with the use of data from responses to four different visual stimuli (all faces), and five different fixation positions, arranged as shown in Fig. 2.

Recording sites

X-radiographs were used to locate the position of the microelectrode on each recording track relative to permanently implanted reference electrodes and bony landmarks. The position of cells was then reconstructed from the X-ray coordinates taken together with serial 50-μm histological sections that showed the reference electrodes and microlesions made at the end of some of the microelectrode tracks (Feigenbaum and Rolls 1991).

RESULTS

The responses of a large sample of cells in the temporal cortical visual areas were recorded. It was possible to perform 116 experiments on the effects of fixation position on neuronal responses on 44 neurons in 3 monkeys.

An example of the results of an experiment in which the effects of different fixation positions on the responses of a neuron to full-size face stimuli that were as large as the screen are shown in Fig. 3. The face subtended 17° at the retina. The means and standard errors of the means of the neuronal responses at the different fixation positions are shown. The neuronal responses were as large when the monkey was fixating at the edge of an effective face stimulus for the cell as when he was fixating at the center of the same face (Fig. 3*A*). The stimulus selectivity of the cell was maintained across this area of visual field. A less effective face stimulus for the cell elicited only a small response when the monkey was fixating at the center of that face and at its edges (Fig. 3*B*). This considerable invariance of the neuronal response with respect to fixation position was confirmed by a two-way ANOVA, which showed a significant effect of face identity ($F(3,60) = 331.7$, $P < 0.0001$), no effect of fixation position ($F(4,60) = 2.12$, $P > 0.05$), and no significant interaction between face identity and fixation position ($F(12,60) = 1.78$, $P > 0.05$). (In the ANOVAs, each category within a factor was 1 of 5 fixation positions, arranged as shown in Fig. 2.)

Neuronal responses to both effective and noneffective stimuli that were independent of fixation position on the stimulus were found throughout the receptive-field area tested of the cells. Local areas of greater sensitivity in the receptive field were not found. This is illustrated (for another cell) by the data shown in Fig. 4, in which many fixation positions throughout the receptive field were tested while the cell was recorded. The edge of the face is indicated by a dotted line. It is clear that the neuron had responses that reflected which stimulus was shown consistently, with only a minor effect of fixation position up to angles as great as 12° from the center of the face. The effect of fixation position as tested by the ANOVAs for the cell was not significant more frequently than would be expected by chance in this set of experiments. Further examples of data from other cells that also show that the responses to the effective stimuli were relatively independent of the fixation position are shown in Fig. 5. The same pattern of responses is shown for another cell in Fig. 6 with the use of a half-size face with different fixation positions. The difference in the response rate to the two stimuli is maintained throughout the visual field, although at the more extreme eccentricities the overall neuronal response is attenuated.

The average response obtained from cells in 116 experiments (scaled to the response with fixation at the center of the face) as a function of the number of degrees the fixation point was away from the center of an effective face stimulus is shown in Fig. 7. The experiments were conducted on 44 cells. Each experiment consisted of a measurement for at least two different stimuli (typically 4) of the neuronal responses with fixation at five different positions as indicated in Fig. 2. The different experiments were for different eccentricities for F0, F1, F3, and F4 from the center of the face. (Different numbers of experiments were performed at each eccentricity, so the standard deviation of the distribution of each mean is shown in Fig. 7.) The face stimuli in these experiments generally subtended 17° at the retina. The cells' responses only tended to diminish when the fixation point was 11° from the center of the face (which was generally 2.5° beyond the edge of the face). The ANOVAs performed on individual cells showed that, when fixation was as far eccentric as 2–4° away from the edge of the face, then the attenuation of the neuronal response was in 40% of cases nonsignificant, in 40% significant, and in 20% there was a nonsignificant reduction for some but not other stimuli. The neurons described here had only little response to noneffective stimuli, regardless of where they were in the visual field.

These results indicated that the cells could respond differently to different faces when a point some degrees from the edge of the face was fixated. To test whether humans could also discriminate and identify the faces correctly when fixation was off the edge of the face, we carried out a series of psychophysical experiments on four human subjects (2 of whom were the authors M. J. Tovee and E. T. Rolls). Using the same experimental apparatus and protocol, we presented a series of trials with four different half-size faces in random order and asked the subjects to identify which image was shown. The fixation points 0, 1, 3, and 4 were at

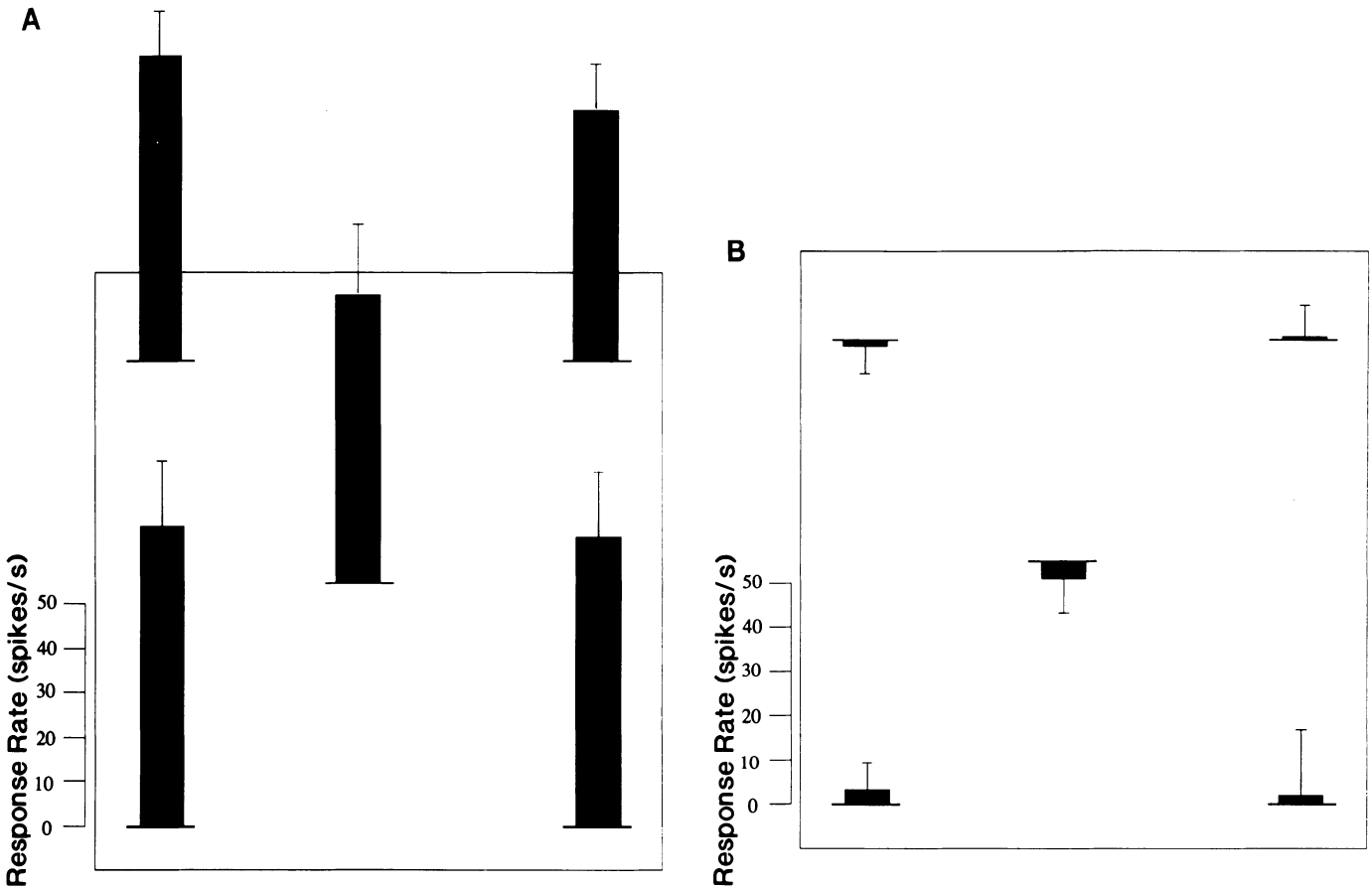

FIG. 3. Neuronal responses of 1 cell with different fixation positions when a full-size effective (*A*) and less effective (*B*) face stimulus was shown. The means and standard errors of the means of the neuronal responses (firing rate minus spontaneous rate) are shown in this and the following figures. These data were obtained for a full-size face (extending 8.5° from the center of the screen), with the fixation spots when not central 11° from the center.

an eccentricity of 11°, and the half-size faces extended out 4.25° from the center. The data showed that the subjects were able to perform the task almost perfectly (97% correct).

The neurophysiological data show that the responses of the cells were relatively independent of fixation position. Analysis of the neuronal spike trains using information theory was next used to *quantify* the extent to which information was made *explicit* in the cells' responses about which stimulus (face) was shown, rather than information about where the stimulus was on the retina.

We were able, using the methods developed from those of Optican et al. (1991) by Tovee et al. (1993), to measure for each neuron the information about stimulus type and fixation position contained in the firing rate (measured over the poststimulus period 0–400 ms). The information theoretic analysis also enabled us to assess how the information about image type and fixation position was carried by the temporal properties of the spike train. The preliminary step in the analysis involves computation of the principal components of the spike trains elicited when different images and fixation positions were being tested. The principal components were calculated for 40 time bins of the neuronal response taken in the poststimulus time of 0–400 ms (Richmond and Optican 1987; Tovee et al. 1993). The results are shown in Fig. 8 (using the correction 2 procedure of Tovee et al. 1993). It is clear from the information measure based on firing rate that this population of cells had considerable information about which of two to four face stimuli were shown (on average 0.203 bits), and only about one-sixth as much information about where the fixation point was (with 5 possible positions) on the stimuli (on average 0.035 bits). (The mean ratio of the information about stimulus identity to that about location of the stimulus in the visual field available in the responses of these neurons was 0.203/0.035 = 5.8.) [Given that there were more fixation positions, 5, than stimuli, 2–4, the maximum information a cell *could* have conveyed (about 4 stimuli) is 2 bits, and about 5 fixation positions is 2.3 bits, so the cells *could,* if they had been tuned to position, have provided more information about fixation position than about identity.] For identity the bulk of the information is coded by the first principal component (which is highly correlated with the firing rate). Addition of the second principal component added only 14.6% more information, and addition of the third principal component 3.6%. The situation is complementary for fixation position. The first principal component carried comparatively little information about fixation position, the addition of the second principal component added 108.6% more information, and the third component 15.7% more information. The amount of information about identity and about fixation position encoded in the first three prin-

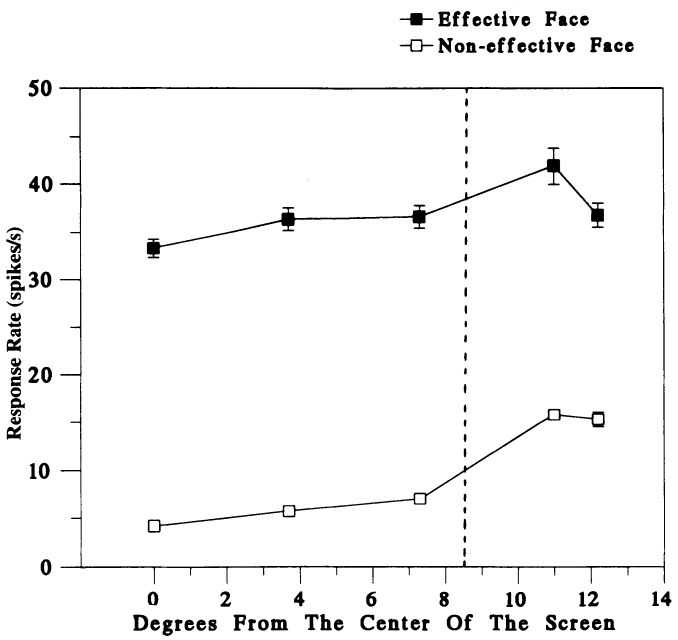

FIG. 4. Responses (firing rate minus spontaneous rate) of a single cell as a function of angular distance of the fixation point from the center of the face, for stimuli that were effective or noneffective for that cell. Data from fixation points along the different diagonals (1 of which is indicated in Fig. 2) are combined. The means and standard errors of the neuronal responses are shown. Dotted line indicates the edge of the face.

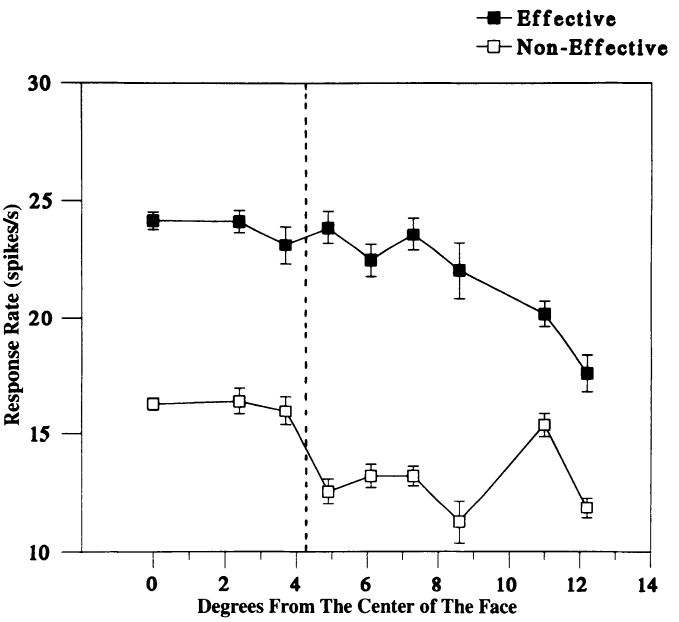

FIG. 6. Responses of a single cell as a function of angular distance of the fixation point from the center of the face, for 1/2-size stimuli that were effective or noneffective for that cell. The 1/2-size images extended 4.25° from the center. Data from fixation points along the diagonals indicated in Fig. 2 are combined. The means and standard errors of the neuronal responses are shown. Dotted line indicates the edge of the face.

cipal components of the spike train of each cell analyzed is shown in Fig. 9.

We have shown previously that the latency for the neuronal responses when the fixation position is away from the center of the face is 20–30 ms longer than for fixation on the center of the face, and further, that this latency difference is for the majority of the cells reflected in the second principal component (Tovee et al. 1993). [This is consistent with the point that in earlier visual processing areas (V1, V2, and V4) when fixation is on the center of the stimulus, more cells will be active and carrying information about the stimulus than when fixation is eccentric. The input to temporal cortical cells from many earlier cells may cause the cell to reach its firing threshold more quickly.] Thus the fact that the second (and 3rd) principal components shown in Fig. 8 convey proportionally more information about image position may well be related to the fact that different fixation positions can often alter the latency of the neuronal responses. This is confirmed by Fig. 10, which shows that the (limited amount of) information that is available in the

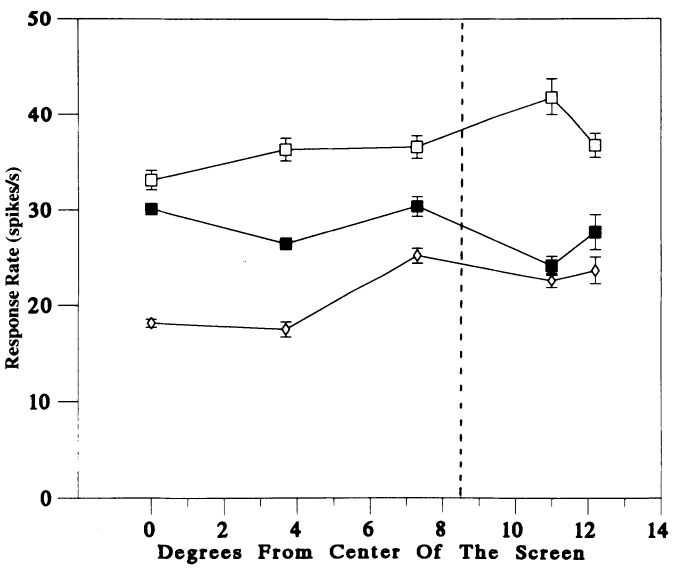

FIG. 5. Responses as a function of angular distance of the fixation point from the center of the face for effective face stimuli for *3 different cells*. Each set of data is for 1 cell tested to an effective face stimulus for that cell. Data from fixation points along the diagonals indicated in Fig. 2 are combined. The means and standard errors of the neuronal responses are shown. Dotted line indicates the edge of the face.

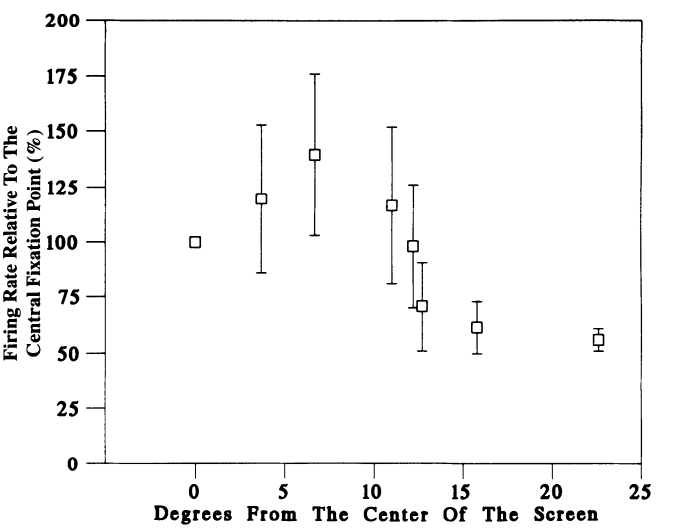

FIG. 7. Mean response of the population of 44 neurons analyzed for different distances from the center of an effective face stimulus that extended 8.5° from the center. Data are shown relative to the response with fixation at the center of the face, with standard deviations indicated.

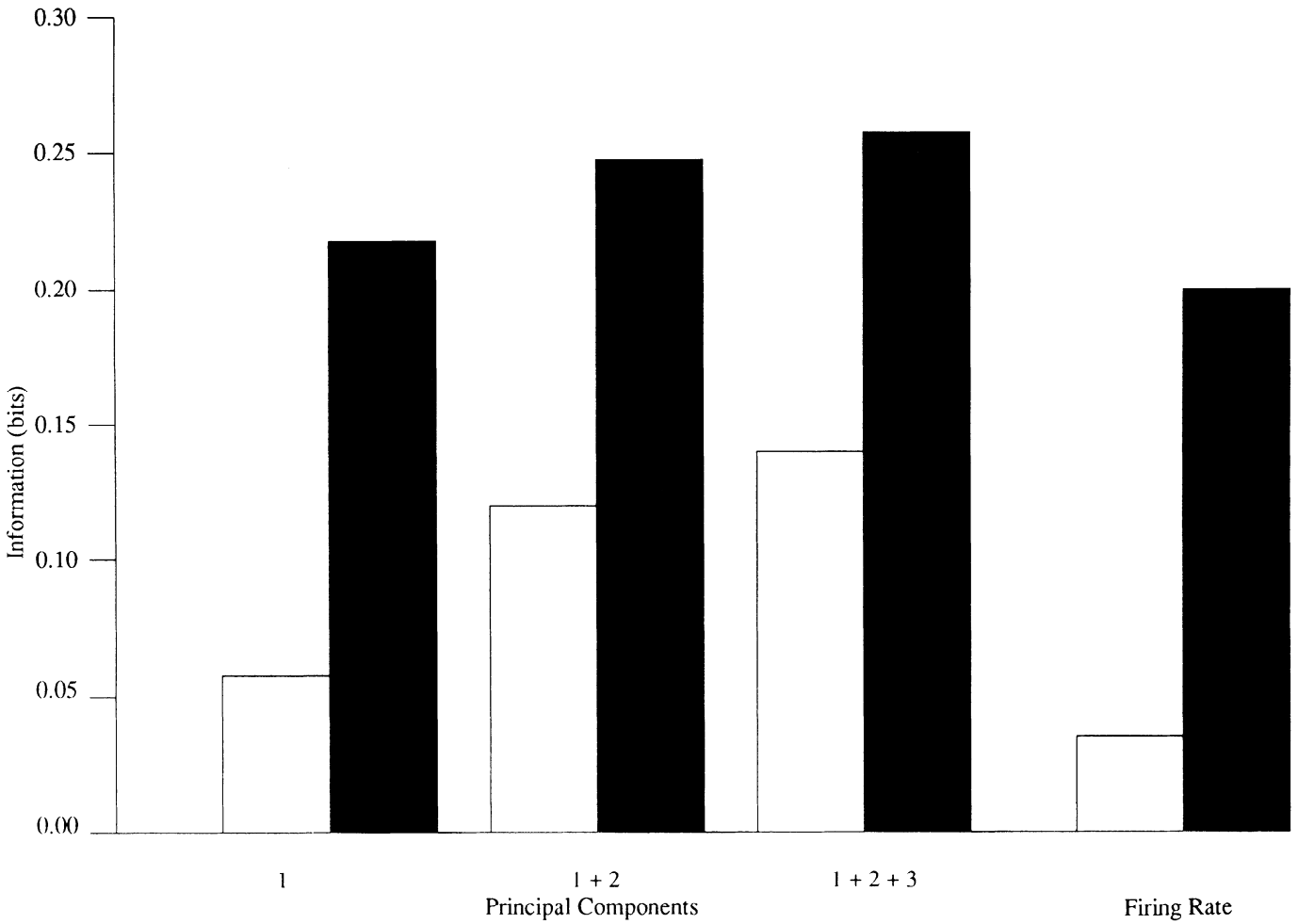

FIG. 8. Average information per cell in principal components 1–3 and in the firing rate about image identity (■) and fixation position (□) across the population of 44 neurons analyzed (see text).

responses of these neurons about fixation position is present primarily in time bins near the start of the neuronal response (with the time bin for the period 80–100 ms carrying much more of the information about fixation position than any other time bin). (Each time bin in Fig. 10 was 20 ms long.) In contrast, as shown in Fig. 10, considerable information about image identity is present in all the 20-ms time bins in the period 100–220 ms. The information available in each of these periods of the spike train about image identity accumulates to result overall in the cells conveying much more information about image type than fixation position, as shown in Fig. 8 (see further Tovee et al. 1993).

The results described so far show that the responses of individual neurons were relatively independent of fixation position on or close to the stimulus. However, for some cells, there was some tendency for the responses to be greater when, for example, the fixation was on the lower part of the face. To provide an indication of whether the centers of the receptive fields of the cells were systematically offset with respect to each other, so that there might be a distributed representation of information about where a stimulus was with respect to the fovea, we computed the center of gravity of the responses of each neuron, to examine whether the centers of gravity were systematically offset from the fovea. The center of gravity was calculated as

$$x = \langle f_i \cdot x_i \rangle / \langle f_i \rangle$$
$$y = \langle f_i \cdot y_i \rangle / \langle f_i \rangle$$

where x and y are the x and y coordinates of the center of gravity, f_i is the firing rate at position x_i, y_i, and $\langle \cdot \rangle$ indicates the ensemble average. (The firing rates were measured with fixation in the 5 positions shown in Fig. 2. The standard deviation was also calculated.)

The distances of the centers of gravity from the fovea of the different cells are shown in Fig. 11, scaled so that a distance of 1 corresponds to 11°. The centers of gravity of the cells were grouped relatively close to the fovea, with the great majority having centers of gravity within 3° of the fovea.

If the face cells recorded from in this study have bilateral responsive fields, receiving inputs from both the contralateral and ipsilateral visual fields, the ratio of the responses to stimuli on the left side to stimuli on the right side of the fovea should be one (given that the recordings were in the right hemisphere). The magnitude of 44 cells' response to full-size face stimuli on the left side of the visual field (contralateral to the recording sites in the right hemisphere; fixation positions 0 and 3, see Fig. 2.) compared with the response to stimuli on the right side of the visual field (fixation positions 1 and 4) is shown in Fig. 12*A*. Each square

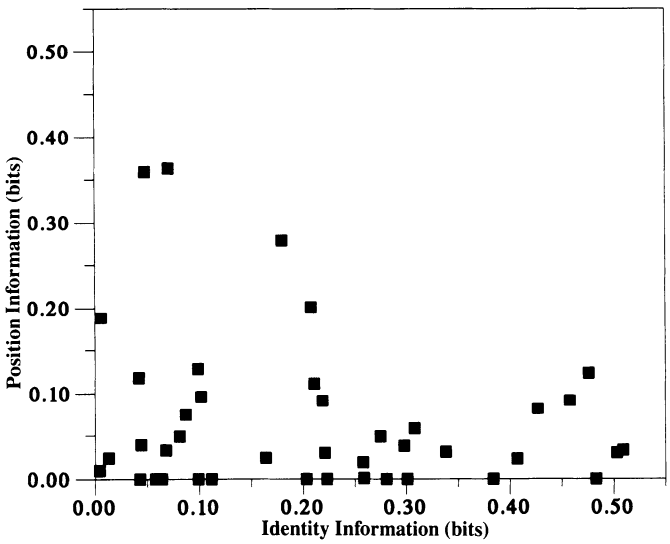

FIG. 9. Information contained in the neuronal response of each of 44 cells about which stimulus had been shown (abscissa), and about the fixation position on the stimulus (ordinate), for the different cells. Each point shows the data from a single cell. The information shown was that available based on principal component analysis of a 400-ms period of the neuronal spike train starting 0 ms after the stimulus appeared (see text). The information was calculated from the 1st 3 principal components.

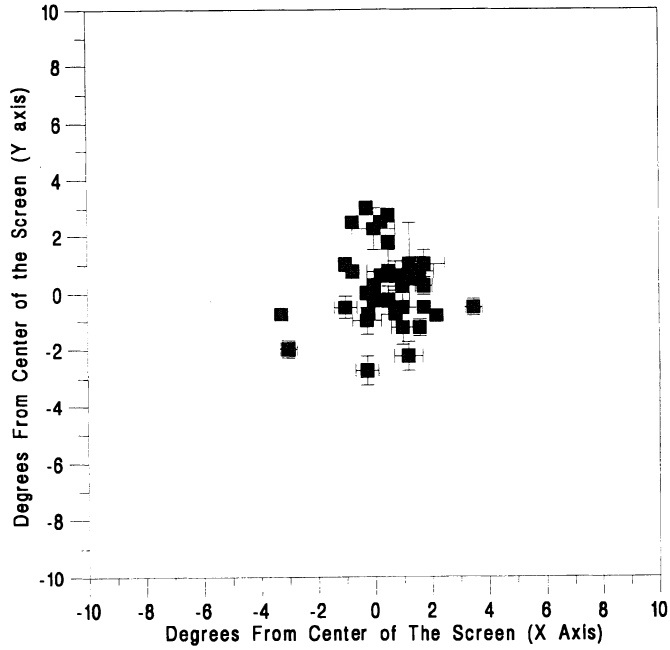

FIG. 11. Center of gravity (see text) of the receptive area of each of the cells analyzed, together with its standard deviation estimated from the standard deviation of the firing rate.

represents the response of a single cell. The stimuli were presented at an eccentricity of 11°, so that the closest part of the face to the fovea was 2.5° away. Cells with equal responses in the two visual fields would fall along the dashed line of Fig. 12*A*. The majority of the cells did not significantly differ in position from this line.

The magnitude of 10 cells' response to half-size face stimuli on the left side of the visual field is compared with their response to stimuli on the right side of the visual field shown in Fig. 12*B*. The cells were again recorded on the right side of the brain. The stimuli were presented at an eccentricity of 11°, so that the closest part of the face stimulus was 6.5° from the fovea. The distribution of the relative

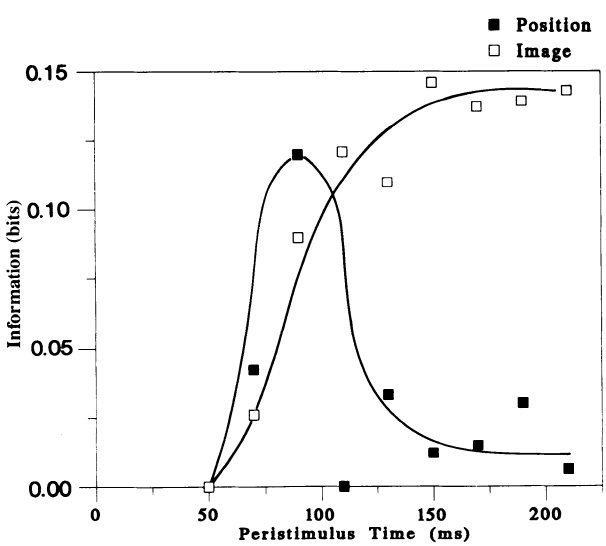

FIG. 10. Information that is available in 20-ms time bins of the firing rate about fixation position (■) and image identity (□) for the population of 44 cells analyzed.

responses was still not significantly different from one, so that the cells had large, bilateral receptive fields.

The recording sites of the neurons analyzed in this study are shown in Fig. 13. The neurons recorded were in both the cortex in the superior temporal sulcus and the inferior temporal cortex.

DISCUSSION

These neurophysiological results show that there is a population of neurons in the temporal cortical areas of the awake behaving macaque that respond with stimulus selectivity to faces that is relatively independent of where the fixation position is on the face. The invariance extends across many degrees of visual angle, with no diminution until the fixation position is 10° from the center of a face that extends 8.5° from its center (see Fig. 7). It was shown moreover that information about the identity of the stimulus was made explicit in the responses of the single neurons, and that there was relatively little information about where the fixation point on the face was (Figs. 8–10). The mean ratio of the information about stimulus identity to that about location of the stimulus in the visual field available in the responses of these neurons was 5.8.

This form of translation invariant representation is appropriate as an output from the ventral visual system to memory systems that must learn about objects, rather than about where patterns of light happen to fall at any one time on the retina. The inferior temporal cortex projects to at least three potentially important regions for memory: the amygdala, the hippocampus, and the prefrontal cortex. Cells in the amygdala have been found that are responsive to complex visual stimuli, such as faces (Leonard et al. 1985; Rolls 1992b), and the output of the temporal cortical areas to the amygdala may be important for associating which stimuli (in this case faces) are associated with pri-

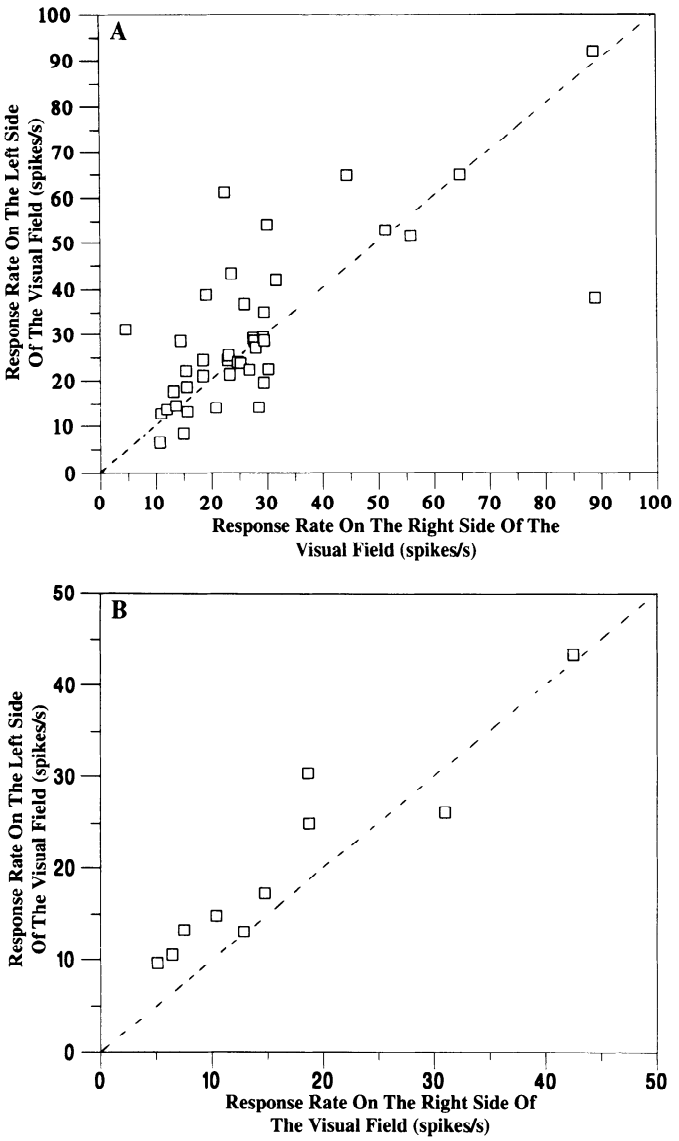

FIG. 12. *A*: magnitude of the responses of 44 cells to full-size face stimuli shown contralaterally (on the left side of the visual field) or ipsilateral (on the right side of the visual field) to the recording site (in the right hemisphere). Each square represents the response of a single cell. Cells with equal responses would appear along the dashed line. The monkey fixated at an eccentricity of 11° from the center of the face, and the face extended 8.5° from its center. *B*: magnitude of the responses of 10 cells to ½-size face stimuli shown contralaterally (on the left side of the visual field) or ipsilateral (on the right side of the visual field) to the recording site (in the right hemisphere). Each square represents the response of a single cell. Cells with equal responses would appear along the dashed line. The monkey fixated at an eccentricity of 11° from the center of the face, and the face extended 4.25° from its center.

mary reinforcement (Rolls 1984, 1990a, 1992b). The output from the inferior temporal cortex (via the parahippocampal gyrus and entorhinal cortex) to the hippocampal system may be important for forming intermediate term memories of episodes, such as who was present on a particular previous occasion (Rolls 1990b, 1991; Treves and Rolls 1994). The inferior temporal cortex also projects to the inferior prefrontal convexity (Barbas 1992), and the responses of some neurons in this region closely resemble those of inferior temporal cortex neurons (Wilson et al. 1993). The prefrontal neurons have large visual fields, are responsive to complex stimuli, such as faces, and their responses are invariant over transformation of size, inversion, and color. These neurons in the inferior prefrontal convexity are suggested to play a role in working memory for object identity (Wilson et al. 1993). For all three memory systems, what must be entered to the memory is which particular object was present, rather than where it was on the retina. The value of this is that once something has been stored in the memory about the object, no further separate memory, or special generalization process, is required when the object appears on a different part of the retina.

These results extend earlier findings. Desimone and Gross (1979) showed that under anesthesia, the receptive fields of inferior temporal neurons are typically large (the average was $625°^2$ in area) and include the fovea. Desimone and his colleagues obtained further evidence on translation invariance of inferior temporal cortex cells with the use of boundary curvature descriptor stimuli (Desimone et al. 1985; Schwartz et al. 1983). They showed, for example, that some neurons had similar profiles of responses across a set of boundary curvature descriptors when they were shown at different positions. However, many processes involved in normal perception are unlikely to operate normally under anesthesia (Rodman et al. 1991). For example, the responses of inferior temporal neurons under anesthesia are sluggish and show habituation. It is therefore important to examine the degree of translation invariance shown by neurons in this region when it is operating normally. In this paper we have extended evidence on translation invariance not only to the awake behaving state, but also for the population of face-selective neurons recorded mainly in the cortex in the anterior part of the superior temporal sulcus as well as in the inferior temporal visual cortex. Moreover, we showed that they maintain their selectivity for different faces relatively independently of the fixation position on the face. We also quantified the extent of the translation invariance by applying information theory and showed that much more information was available in the responses of these neurons about stimulus identity than about stimulus position.

It was of interest that these neurons often continued to show responses to the face shown when the fixation point was several degrees away from the edge of the face. This emphasizes the translation invariance of these neuronal responses. However, given that perception of stimuli, and visual acuity, does decline with increasing eccentricity from the fovea, it is very likely that the selectivity between different faces of the neuronal responses cannot be maintained for fixation very far beyond the edge of a face. For most cells the responses to all stimuli tend to decrease at larger eccentricities, and thus the difference in a cell's response between different stimuli also tends to become less. Because we found it remarkable that the neurons carried some information about face identity even when fixation was several degrees off the edge of the face, we investigated this capacity psychophysically in humans. We found that under the same test conditions, humans could also discriminate between (and could identify) the faces. However, the subjects did report some differences in their perception of faces depending on whether they fixated centrally or peripherally.

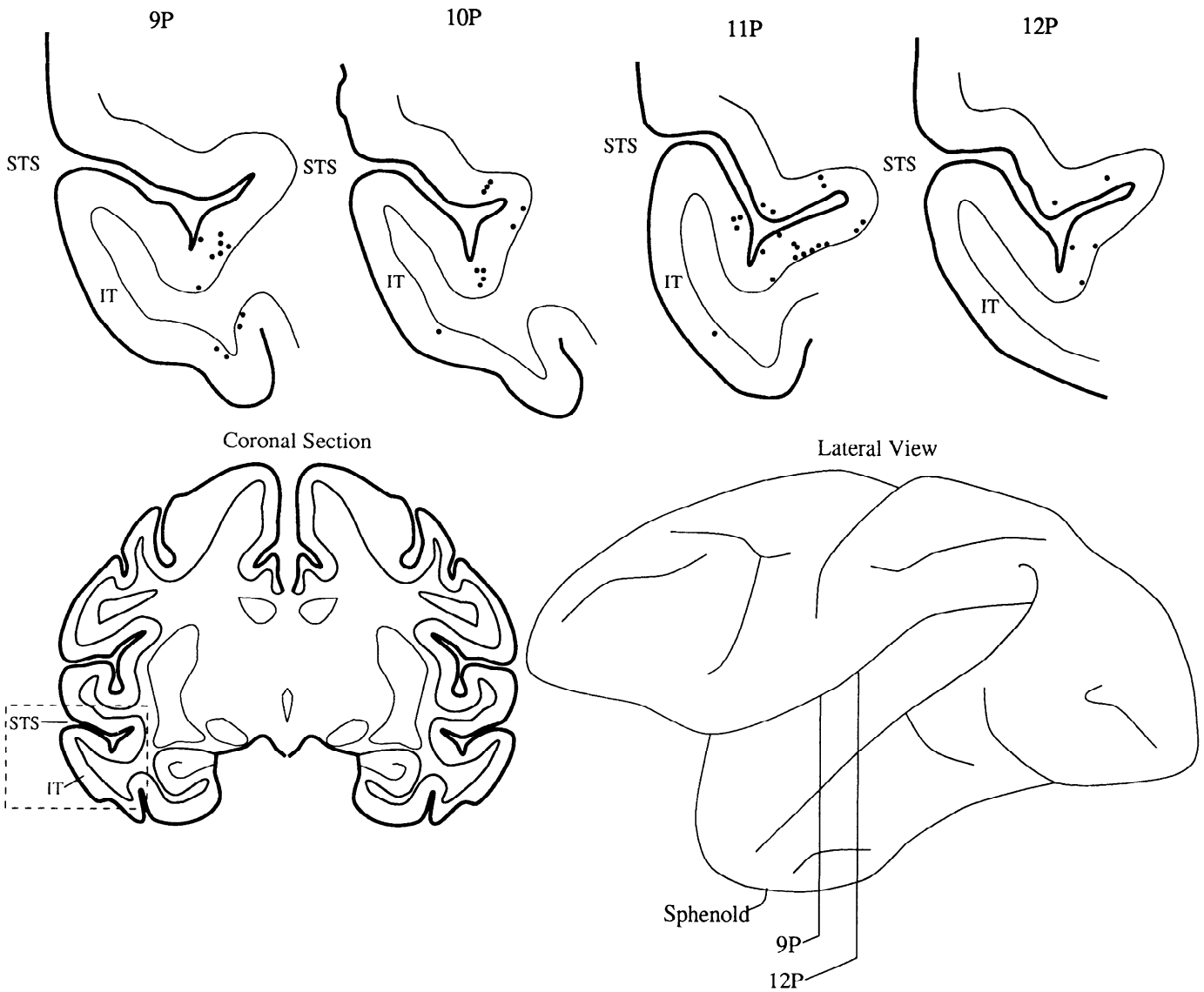

FIG. 13. Recording sites shown on coronal sections of the neurons included in this study. Positions of the coronal sections are shown on a lateral view of the macaque brain. Distances refer to millimeter posterior (P) to the sphenoid reference plane (see text). IT, inferior temporal visual cortex; STS, superior temporal sulcus.

The subjects reported a noticeable afterimage when fixating peripherally (6.5° away from the nearest part of the face) under the test conditions, and this was not noticeable when fixating on the center of the face. This presumably reflects the longer integration times of the peripheral visual mechanisms. Also, the subjects reported that when fixating this far off the edge of a face that the general outlines of one or several distinctive features, such as hair style or the size and shape of the nose, appeared to be particularly useful for identification. This is consistent with the reduced acuity that is available peripherally. When fixation was on any part of the face, or was close (within 1 or 2°) to the face, not only was translation invariance found for the cells, but also perceptually the face appeared to have its normal integration and resolution.

As shown in Fig. 7, there was a tendency for the cells to respond a little more when fixation was a few degrees from the center of a face. As shown by the standard errors, this effect did not reach statistical significance across the whole population of cells, but for some individual cells the effect was significant. If it were an effect that was consistently found, it might be related to the fact that, for the sizes of face stimuli used, the fixation positions 3–7° from the face center might have brought the gaze onto a part of the face such as an eye, which was particularly effective for a particular cell, or that the cell was less inhibited by information in another part of the face.

The observation that some of the cells had longer response latencies when fixation was away from the center of a face (see also Tovee et al. 1993) could be related to the fact that with eccentric fixation from an effective stimulus, there will be fewer cells at earlier stages of processing (in, e.g., V1 and V2) that are activated by the stimulus. The smaller number of relevant active afferents to cells in higher order cortical areas might then mean that they would take longer to accumulate sufficient activation to cause firing, thus producing a small lengthening in the neuronal response latency.

The results described here were obtained when the monkey was performing a fixation task. The fixation spot was turned off 100 ms before the test stimulus appeared. Under these conditions, without any further attempt to control the spatial locus of the monkey's attention, the temporal cortex single neurons showed good translation invariance. It was, however, the case that there was only one visual stimulus present in the visual field during these experiments. It will be of interest in future experiments to investigate whether translation invariance continues to operate perfectly under more complicated stimulation conditions.

It was of interest that for identity the bulk of the information is coded by the first principal component (which is highly correlated with the firing rate). In contrast, the first principal component carried comparatively little information about fixation position, the addition of the second principal component added 108.6% more information, and the third component 15.7% more information (see Fig. 9). This is consistent with the evidence that the bulk of the information about stimulus identity is carried in the first principal component, so that temporal encoding within the spike train is not very important for conveying information about stimulus identity, and that at least some of the information in higher principal components than the first is related to the onset latency of the neuronal response (see Tovee et al. 1993). The onset latency in this study tended to be longer for stimuli not at the fixation point.

In this paper we have shown that, when the cortex is operating normally, temporal cortex neurons can have translation invariant responses that reflect the identity of the stimulus seen. Such translation invariance is a far from trivial computational property of a connectionist system. Given that this translation invariance is a real property of the visual system, it is worthwhile to continue investigation of biologically plausible algorithms that can produce translation invariant representations of objects (e.g., Wallis et al. 1993), to better understand how the visual system operates.

We thank Dr. A. Treves for advice on the center of gravity calculations.

This research was supported by the Medical Research Council, PG8513790 to Dr. E. T. Rolls.

Present address and address for reprint requests: M. J. Tovee, Dept. of Psychology, University of Newcastle, Ridley Bldg., Newcastle upon Tyne, NE1 7RU, United Kingdom.

Received 26 August 1993; accepted in final form 14 March 1994.

REFERENCES

BARBAS, H. Architecture and cortical connections of the prefrontal cortex in the rhesus monkey. In: *Advances in Neurology,* edited by P. Chauvel, A. Delgado-Escueta, et al. New York: Raven, 1992, p. 91–115.

BAYLIS, G. C., ROLLS, E. T., AND LEONARD, C. M. 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, 1985.

BAYLIS, G. C., ROLLS, E. T., AND LEONARD, C. M. Functional subdivisions of temporal lobe neocortex. *J. Neurosci.* 7: 330–342, 1987.

COWEY, A. AND GROSS, C. G. The effects of foveal prestriate and inferotemporal lesions on visual discrimination by rhesus monkeys. *Exp. Brain Res.* 11: 128–144, 1970.

DESIMONE, R. AND GROSS, C. G. Visual areas in the temporal cortex of the macaque. *Brain Res.* 178: 363–380, 1979.

DESIMONE, R., SCHEIN, S. J., MORAN, J., AND UNGERLEIDER, L. G. Contour colour and shape analysis beyond the striate cortex. *Vision Res.* 25: 441–452, 1985.

FEIGENBAUM, J. D. AND ROLLS, E. T. Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. *Psychobiology* 19: 21–40, 1991.

GAFFAN, D., HARRISON, S., AND GAFFAN, E. A. Visual identification following inferotemporal ablation in the monkey. *Q. J. Exp. Psychol.* 30B: 5–30, 1986.

GROSS, C. G. AND MISHKIN, M. The neural basis of stimulus equivalence across retinal translation. In: *Lateralization in the Nervous System,* edited by S. Harnad, R. Doty, J. Jaynes, L. Goldstein, and G. Krauthamer. New York: Academic, 1977, p. 109–122.

GROSS, C. G., ROCHA-MIRANDA, C. E., AND BENDER, D. B. Visual properties of neurons in inferotemporal cortex of the macaque. *J. Neurophysiol.* 35: 96–111, 1972.

JUDGE, S. J., RICHMOND, B. J., AND CHU, F. C. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res.* 20: 535–538, 1980.

LEONARD, C. M., ROLLS, E. T., WILSON, F. A. W., AND BAYLIS, G. C. Neurons in the amygdala of the monkey with responses selective for faces. *Behav. Brain Res.* 15: 159–176, 1985.

MERRILL, E. G. AND AINSWORTH, A. Glass-coated platinum-plated tungsten microelectrodes. *Med. Biol. Eng.* 10: 662–672, 1972.

MISHKIN, M., LEWIS, M. E., AND UNGERLEIDER, L. G. The equivalence of parieto-preoccipital subareas for visuospatial ability in monkeys. *Behav. Brain Res.* 6: 57–77, 1982.

MISHKIN, M., UNGERLEIDER, L. G., AND MACKO, K. A. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6: 414–417, 1983.

OPTICAN, L. M., GAWNE, T. J., RICHMOND, B. J., AND JOSEPH, P. J. Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. *Biol. Cybern.* 305–310, 1991.

PERRETT, D. I., ROLLS, E. T., AND CAAN, W. Visual neurons responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* 47: 329–342, 1982.

POHL, W. Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys. *J. Comp. Physiol. Psychol.* 82: 227–239, 1973.

RICHMOND, B. J. AND OPTICAN, L. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J. Neurophysiol.* 57: 147–161, 1987.

RODMAN, H. R., SKELLY, J. P., AND GROSS, C. G. Stimulus selectivity and state dependence of activity in inferior temporal cortex of infant monkeys. *Proc. Natl. Acad. Sci. USA* 88: 7572–7575, 1991.

ROLLS, E. T. Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum. Neurobiol.* 3: 209–222, 1984.

ROLLS, E. T. A theory of emotion, and its application to understanding the neural basis of emotion. *Cognition and Emotion* 4: 161–190, 1990a.

ROLLS, E. T. Theoretical and neurophysiological analysis of the functions of the primate hippocampus in memory. *Cold Spring Harbor Symp. Quant. Biol.* 55: 995–1006, 1990b.

ROLLS, E. T. Neural organisation of higher visual functions. *Curr. Opin. Neurobiol.* 1: 274–278, 1991.

ROLLS, E. T. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335: 11–21, 1992a.

ROLLS, E. T. Brain mechanisms for invariant visual recognition and learning. *Behavioural Processes.* In press.

ROLLS, E. T. Neurophysiology and functions of the primate amygdala. In: *The Amygdala,* edited by J. P. Aggleton. New York: Wiley-Liss, 1992b, chapt. 5, p. 143–165.

ROLLS, E. T. The processing of face information in the primate temporal lobe. In: *Processing Images of Faces,* edited by V. Bruce and M. Burton. Norwood, NJ: Ablex, 1992c, chapt. 3, p. 41–68.

ROLLS, E. T. AND BAYLIS, G. C. Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp. Brain Res.* 65: 38–48, 1986.

ROLLS, E. T., BAYLIS, G. C., AND LEONARD, C. M. Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus. *Vision Res.* 25: 1021–1035, 1985.

ROLLS, E. T., BURTON, M. J., AND MORA, F. Hypothalamic neuronal responses associated with the sight of food. *Brain Res.* 111: 53–66, 1976.

ROLLS, E. T. AND TOVEE, M. J. The neurophysiology of backward masking. *Proc. Roy. Soc. Lond. B.* In press.

ROLLS, E. T., TOVEE, M. J., PURCELL, D. G., STEWART, A. L., AND AZZO-

PARDI, P. The responses of neurons in the temporal cortex of primates, and face identification and detection. *Exp. Brain Res.* In press.

ROLLS, E. T., YAXLEY, S., AND SIENKIEWICZ, Z. J. Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey. *J. Neurophysiol.* 64: 1055–1066, 1990.

SARY, G., VOGEL, R., AND ORBAN, G. A. Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science Wash. DC* 260: 995–997, 1993.

SCHWARTZ, E. L., DESIMONE, R., ALBRIGHT, T. D., AND GROSS, C. G. Shape recognition and inferior temporal neurons. *Proc. Natl. Acad. Sci. USA* 80: 5776–5778, 1983.

TOVEE, M. J. AND ROLLS, E. T. Oscillatory activity is not evident in the primate temporal visual cortex with static stimuli. *NeuroReport* 3: 369–372, 1992.

TOVEE, M. J., ROLLS, E. T., TREVES, A., AND BELLIS, R. P. Information encoding and the responses of single neurons in the primate temporal visual cortex. *J. Neurophysiol.* 70: 640–654, 1993.

TREVES, A. AND ROLLS, E. T. A computational analysis of the role of the hippocampus in memory. *Hippocampus.* In press.

WALLIS, G., ROLLS, E. T. AND FOLDIAK, P. Learning invariant responses to the natural transformations of objects. *Int. Joint Conf. Neural Networks* 2: 1087–1090, 1993.

WEISKRANTZ, L. AND SAUNDERS, R. C. Impairments of visual object transforms in monkeys. *Brain* 107: 1033–1072, 1984.

WILSON, F. A. W., O'SCALAIDHE, S. P., AND GOLDMAN-RAKIC, P. S. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science Wash. DC* 260: 1955–1958, 1993.

YAMANE, S., KAJI, S., AND KWANO, K. What facial features activate face neurons in the inferotemporal cortex of the monkey? *Exp. Brain Res.* 73: 209–214, 1988.