COGNITIVE NEUROPSYCHOLOGY, 2000, 17 (1/2/3), 13–34

EFFECT OF IMAGE ORIENTATION AND SIZE ON OBJECT RECOGNITION: RESPONSES OF SINGLE UNITS IN THE MACAQUE MONKEY TEMPORAL CORTEX

E. Ashbridge, D.I. Perrett, M.W. Oram, and T. Jellema
University of St Andrews, Scotland, UK

This study examined how cells in the temporal cortex code orientation and size of a complex object. The study focused on cells selectively responsive to the sight of the head and body but unresponsive to control stimuli. The majority of cells tested (19/26, 73%) were selectively responsive to a particular orientation in the picture plane of the static whole body stimulus, 7/26 cells showed generalisation responding to all orientations (three cells with orientation tuning superimposed on a generalised response). Of all cells sensitive to orientation, the majority (15/22, 68%) were tuned to the upright image. The majority of cells tested (81 %, 13/16) were selective for stimulus size. The remaining cells (3/16) showed generalisation across four-fold decrease in size from life-sized. All size-sensitive cells were tuned to life-sized stimuli with decreasing responses to stimuli reduced from life-size. These results do not support previous suggestions that cells responsive to the head and body are selective to view but generalise across orientation and size. Here, extensive selectivity for size and orientation is reported. It is suggested that object orientation and size-specific responses might be pooled to obtain cell responses that generalise across size and orientation. The results suggest that experience affects neuronal coding of objects in that cells become tuned to views, orientation, and image sizes that are commonly experienced. Models of object recognition are discussed.

INTRODUCTION

The visual system allows us to discriminate between objects of different orientation and size. The system also allows us to generalise across image transformations and identify an object as the same despite changes in view, orientation, and image size. How the visual system represents the appearance of objects to enable these recognition capacities is not resolved.

The study reported here investigates the effects of image transformation in orientation or size on object processing in the cortex of the anterior part of the superior temporal cortex (STSa) of the macaque monkey. This area was studied because it contains cells that are both selectively responsive to

Requests for reprints should be addressed to D.I. Perrett, School of Psychology, University of St Andrews, Scotland, KY16 9JU, UK.

E. Ashbridge is currently at the Division of Psychology, South Bank University, 103 Borough Road, London SE1 0AA, UK.

This research was funded by project grants from the UK MRC, BBSRC, the US ONR, and the HFSP. E. Ashbridge (née Wachsmuth) was supported by a UK SERC studentship. We acknowledge the contribution of N. J. Emery, L. K. Harrison, and J. K. Hietanen, who participated in some of the experiments. We are grateful for the support given by University technical and photographic staff and to Dr Walsh for comments.

the sight of one complex object—the face or head—and sensitive to viewing conditions (e.g. Bruce, Desimone, & Gross, 1981; Perrett, Rolls, & Caan, 1982; Perrett et al., 1985, 1991). The focus of studies on the face or head of previous studies (and the use of terms such as "face responsive" or even "face cells") is perhaps misleading as to the selectivity of the cells. A substantial proportion of cells that respond to the sight of the face are also sensitive to visual information arising from body regions other than the face (e.g. Wachsmuth, Oram, & Perrett, 1994). Cells may respond to the face when this is presented in isolation, but such cells often respond to the rest of the body when the head is occluded from sight. Not surprisingly, when both the head and body are visible the response is greater than that seen to the face alone. We therefore examined the effects of image transformations on whole body stimuli.

If an object's representation can be activated independent of the orientation or size of the object's image on the retina, then such a representation will here be referred to as "object-centred." On the other hand, if the object's representation is preferentially activated by a specific image orientation or size, then the representation will be referred to as viewer-centred" since activation depends on orientation and size relative to the viewer.

Scalp-recorded evoked potentials to the sight of faces show orientation and size specificity (Jeffreys, 1989, 1993; Jeffreys, Tukmachi, & Rockley, 1992), implying that faces are coded at a particular stage of the human visual system in a viewer-centred manner.

Neurophysiological studies of the ventral route of cortical processing have previously suggested that cells in early visual areas (V1, V2, V4) exhibit orientation-specific coding of elementary features of objects (Henry, Dreher, & Bishop, 1974; Kobatake & Tanaka, 1994). These cells project to inferotemporal (IT) cortex, where cells are selectively responsive to progressively more complex features but still exhibit orientation-specific responses (Kobatake & Tanaka, 1994; Tanaka, Saito, Fukada, & Moriya, 1991). IT cortex in turn projects to the cortex of the STSa (Seltzer & Pandya, 1978). This area contains cells selective for complex objects and which have previously been reported to respond irrespective of the stimulus orientation (Ashbridge, & Perrett, 1998; Perrett et al., 1982, 1985, 1988).

Studies of temporal cortex cell responses to faces have so far tested few orientations (often restricted to upright and inverted). The first aim of the present study is to determine the extent to which cells in the STSa show object-centred orientation invariance, or viewer-centred orientation specificity in their responses to whole bodies presented in multiple orientations. This allows us to address the question of how orientation in the picture plane is processed for one biologically important object and whether it is processed in a similar way to view (see Logothetis, Pauls, & Poggio, 1995; Perrett et al., 1991; Wachsmuth et al., 1994).

Furthermore, previous studies suggest mainly size-specific coding in V4 and IT and suggest a possible greater degree of size generalisation within STSa (Dobbin, Jeo, Fiser, & Allman, 1998; Ito, Tamura, Fujita, & Tanaka, 1995; Rolls & Baylis, 1986). The second aim of the study reported here is to measure the extent to which cells selective for complex objects in the cortex of the anterior STS generalise across different image sizes.

METHODS

Recordings of responses of single cells from five macaque monkeys (*Macaca mulatta*, two females wt. 4–8kg and three males wt. 5–8kg) were carried out. The techniques applied (including surgical and recording procedures) and results from previous cellular studies of these subjects have been previously described (Perrett et al., 1991; Wachsmuth et al., 1994).

Training and Fixation Task

Pre-surgical training: While in a primate chair, the subjects were trained to fixate on one of five LED lights presented on a white wall at eye level at a distance of 4 metres. For a block of 50 trials the position of the fixation LED light was constant. The

monkey's task was to discriminate the colour of the LED that followed a short signal tone to obtain the monkey's attention. Licking resulted in fruit juice reward for the green LED. The monkey was to withhold licking in order to avoid delivery of a weak saline solution to the red LED. The LED stimuli were presented in pseudorandom order under computer control.

Visual Stimuli

Pictures of eight different views (at 45° intervals rotating from the front view, see Wachsmuth et al., 1994) of the whole body were taken. For each view, eight orientations in the picture plane were constructed (resulting in 64 different stimuli); 0° (upright), 45°, 90° (horizontal), 135°, 180° (inverted), 225°, 270° (horizontal), and 315°. The images were then projected onto a white wall at a viewing distance of 4m, resulting in an image size of 24.4° (1.73m) head to toe. For some cells testing was additionally performed with different views of the head without the body in view. Testing was normally performed with the fixation light at the centre of the test image, subsidiary testing to examine effects of receptive fields was performed with the centre of the image presented 10° above, below, or to the side of the LED fixation spot.

For size stimuli the same range of eight views were presented at four magnifications ranging from 100° (1.73m, head to toe height, subtending 24.4°), 75% (1.3m, 18.5°), 50% (0.87m, 12.3°) to 25% (0.43m, 6.2°) at the viewing distance of 4m. During daily life the monkey subjects saw humans at distances between 0.5 and 5m. The head to toe size of humans encountered by the subjects, at the projection distance of 4m, ranged between 1.5m and 1.8m.

Control stimuli included complex 2D and 3D objects of different sizes, shapes, textures and orientations (broom, lab coats, chairs, pictures of different animals, etc.), simple 2D geometrical shapes (bars, spots and gratings), and simple 3D forms (balls, cylinders, boxes, etc.).

Testing Methods

Every cell from which neuronal activity was recorded was first tested in an exploratory way by presenting a series of static and moving 3D objects (including bodies), and tactile and auditory stimuli. Where cells were found responsive to the face or body, they were tested for (a) selectivity between objects and (b) selectivity between views.

Selectivity Between Objects. This was studied by comparing responses to slide images of faces, bodies, and a minimum of five control objects of approximately equivalent shape, size, and complexity (e.g. a fire extinguisher, lab coat). Test comparisons were also made using videodisk images of heads, bodies, and a variety of laboratory objects. Sensitivity to species of primate was tested (less systematically) by comparing responses to photographs of monkeys and humans.

Selectivity Between Views. Each cell was tested with four views of the head and body (face, left and right profile, and back view) or with eight views (the same four views and four intermediate views). Cells found to be responsive to static views of the whole body but not to control objects were then further investigated for sensitivity to stimulus orientation and size. Size and orientation testing was performed with stimuli presented using the cell's preferred view (front, back, etc.) once this had been established. Stimuli were presented in blocks of trials, with five trials for each stimulus condition in a computer-controlled pseudorandom order.

Data Analysis

Since most cells in the anterior STS respond with a latency of 100msec (± 30msec), the magnitude of cell activity on individual trials was assessed over the 250msec time period occurring 100–350msec after stimulus onset. For some cells [with late response onset (>200msec) or inhibitory responses (i.e. below S/A)] a 500msec time period (100–600msec post-stimulus) was used to assess cell activity.

Cell responses to the whole body presented in different orientations and sizes, control objects, and

S/A were compared on-line using a one-way ANOVA and post hoc tests [protected least significant difference (PLSD), Snedecor & Cochran, 1980] with a significance level of $P < .05$.

Orientation Tuning

A multiple linear regression analysis was performed to estimate the best relationship between response and second-order cardioid function of orientation of the stimulus (see Perrett et al., 1991). This regression analysis calculates the values of the coefficients b_{1-5} of the following equation for producing the highest correlation between cell responses and the angle of orientation.

$$R = b_1 + b_2 \cos(q) + b_3 \sin(q) + b_4 \cos(2q) + b_5 \sin(2q)$$

where R is the response, b_{1-5} are coefficients, and q is the angle of the body orientation. This equation was used to define the optimal angle of orientation (qmax), the maximum response at this orientation (Rmax), and the sharpness of tuning.

Population Response Analysis

All cells, independent of their response pattern, were included in a population analysis. For each cell, the neuronal responses to different test conditions (averaged from five trials) were normalised by applying the following formula: (R-S/A)/(Rmax-S/A) where R = response to test condition, Rmax = maximum response to any orientation or size (depending on the analysis), and S/A = spontaneous activity. The population response to each test condition was then computed from the average of the normalised single cell responses and was displayed after renormalising such that the maximum population response = 1, and population S/A = 0.

Histological Reconstruction

After each recording session, frontal and lateral X-radiographs were taken to localise the electrode. Micro-lesions (10 microamp DC for 30sec), made at the end of some electrode tracks and subsequently identified using standard histological techniques, allowed reconstruction of the electrode position within the brain. In addition, reference markers were made by injection of RRP and the fluorescent dyes, true blue and diamadino yellow.

Once the last recording session had been completed, the monkey was given a sedating dose of ketamine followed by a lethal dose of barbiturate anaesthetic. After transcardial perfusion with phosphate buffered saline and 4% gluteraldehyde/paraformaldehyde fixative, the brain was removed and put into a series of sucrose solutions with increasing concentration (10, 20, and 30%), or alternatively 2% dimethylsulphoxide and 20% glycerol. Standard histological procedures followed.

RESULTS

From five subjects, 23% of anterior STS cells (1692 out of 7288 cells tested) were found to be visually responsive. These included cells selectively responsive to moving or static visual stimuli (see, e.g., Bruce et al., 1981; Oram, Perrett, & Hietanen, 1993). Of the visually responsive cells, a total of 26 cells were found to be selectively responsive to the whole body (i.e. with response to the whole human body significantly greater than that to control objects and S/A). These 26 cells were tested for selectivity to different whole body orientations in the picture plane. A total of 16 cells selectively responsive to the whole body were tested for size sensitivity. All of the cells included in this study were found unresponsive to the variety of control objects tested. Although sensitivity to identity was not tested systematically, no differences were noted in the response to the different experimenters. Moreover, the tested effects of image orientation and size were found to be comparable for images of humans and monkeys.

Generalisation Across Orientations

Seven cells (of 26 tested) responded to all orientations at a rate significantly above S/A and control stimuli. Four of these cells showed complete generalisation in that they responded without statistical difference to all orientations (see Fig. 1).

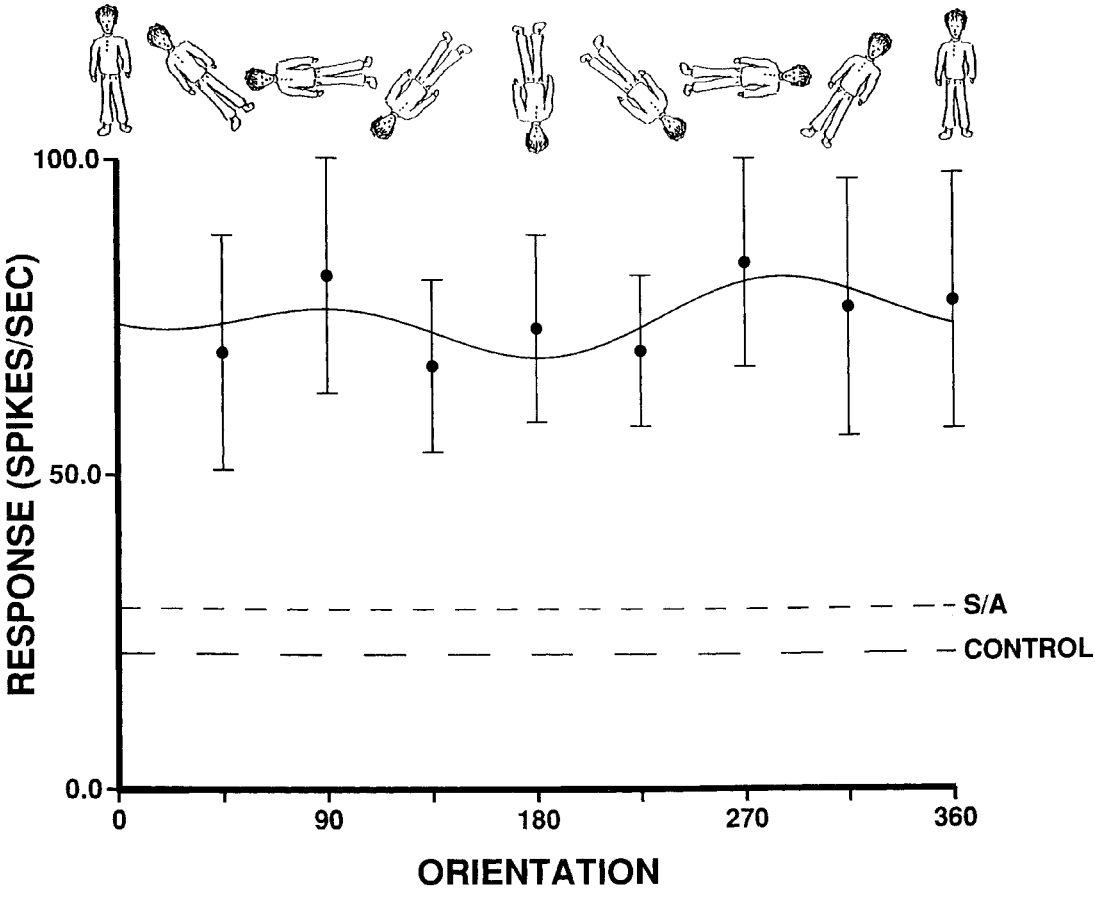

Fig. 1. *The responses of a cell displaying generalisation over body orientation. The mean responses (± 1SE) to the front view of the whole body presented in eight different orientations and spontaneous activity (S/A) are illustrated for one cell (E83_38.31). Orientation is defined as the angle of anti-clockwise rotation from upright (0° or 360°). The cell response showed no significant difference between different orientations tested (P >.5 PLSD each comparison), but the cell responded to all orientations at a rate greater than control stimuli (P <.05 each comparison) and S/A (P <.05 each comparison, except 45°, P <.06 and 135°, P <.067). Overall effect of conditions ANOVA: F(9,40) =2.28, P <.05.*

Sensitivity to Orientation

The majority of cells (19/26) responded to only some rather than all orientations (e.g. Fig. 2). For the cell displayed in Fig. 2 the presence of the whole body presented in its gravitational upright orientation elicited the strongest response. The lower part of Fig. 2 displays responses of the same cell recorded on five individual trials with the whole body presented in its upright, two horizontal, and inverted orientations. Responses to the upright whole body occurred at approximately 180msec after stimulus onset. Responses to horizontally presented bodies were reduced but remained at higher than pre-stimulus activity. When the whole body was presented inverted, however, there was no change in cell activity in comparison to the pre-stimulus period.

For the cell illustrated in Fig. 2, response latency changes with stimulus orientation. Most cells that exhibited orientation tuning did not exhibit changes in latency. Responses to suboptimal orientations were reduced in magnitude but began at the same latency as responses to preferred orientations.

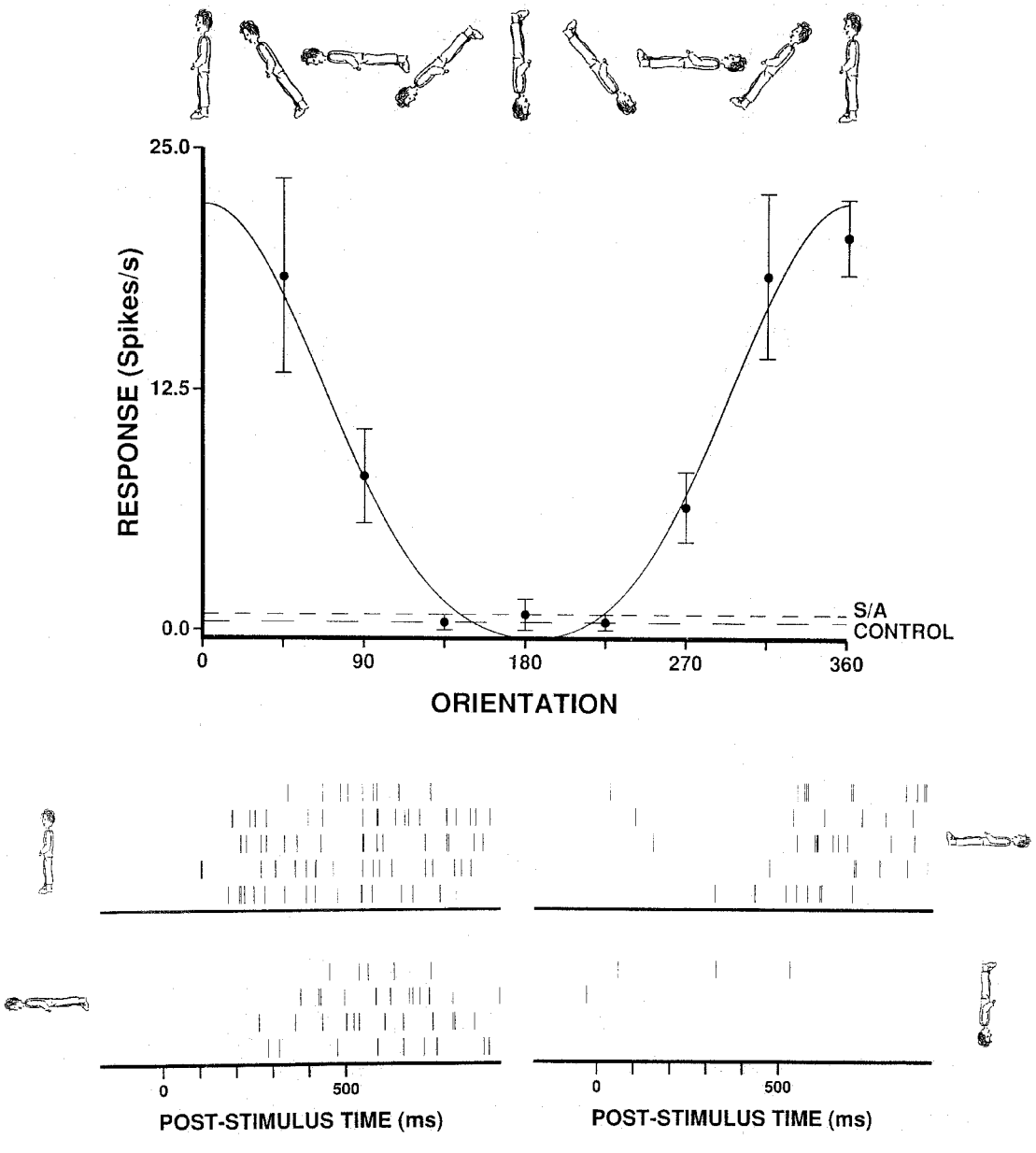

Fig. 2. *The responses of a cell displaying orientation tuning for upright bodies. Upper: The mean responses (± 1SE) to the side (90°) view of the whole body and spontaneous activity (S/A) are illustrated for one cell (E86_38.43). The curve is the best fit cardioid function, relating response to view. The cell exhibited tuning to the whole body stimuli in an upright orientation but responded to all orientations (except 135°, 180°, and 225°) at rates greater than to control stimuli and S/A. ANOVA: F(9,40) = 12.4, P < .0005. Lower: Rastergram displays of responses of the cell on five trials for four orientations. Each trial (originally in pseudorandom order) is represented by a single row of ticks, each tick indicates one action potential. Post-stimulus time is given at the figure base. The responses to the upright whole body (top left panel) were significantly different from responses to non-upright orientations (P < .005 each comparison).*

This behaviour is evident for the cell illustrated in Fig. 3, which responded to face images presented in a horizontal orientation at 200msec after stimulus presentation. The cell illustrated showed reduced responses to upright and inverted orientations but these responses also commenced approximately 200msec after stimulus onset.

Most cells (21/22) displayed an approximately monotonic tuning function. That is, the cells responded best to one orientation and showed a gradual decrease of response amplitude as the stimulus was rotated away from its optimal orientation (see following). One cell responded to two orientations of the front view of the body at a rate greater than intermediate orientations (see Fig. 4). This cell was unresponsive to the back view of the body; therefore responses to the front view were not simply attributable to orientation tuning for vertical stimuli.

Distribution of Orientation Tuning

The majority of cells (22/26, 82%) were selectively responsive for orientation in the picture plane. Although most of the tuned cells (16/22) were selective for orientations close to upright, cells were selective for non-upright orientations (e.g. Fig. 3) including three that were selective for inverted orientations. Figure 3 illustrates the responses of one cell to horizontal images of the face. The preference for this unusual orientation was found for both human and monkey faces. Figure 5 illustrates the responses of one cell tuned to orientations close to inverted. For the cell illustrated in Fig. 5 the sight of upright images of the body did not produce responses above spontaneous or controls.

Average Tuning Curve

The tightness of the tuning varied across the cells. Rotation of the stimulus away from optimal by 45° to 90° reduced most cell responses by half. To find the shape of orientation tuning in the picture plane a regression analysis for each cell was used to estimate the maximal response (Rmax) and the optimal angle of orientation (qmax). Figure 6 (thin lines) illustrates the tuning curves for individual cells maximally sensitive to one orientation in the picture plane and where regression analysis provided a significant correlation between response and second-order cardioid function of orientation of the stimulus. Figure 6 combines different individual tuning curves, each being normalised by setting the mean S/A neuronal firing rate of each cell to 0 and the derived Rmax to 1.0. The peak normalised responses of each cell are expressed as a function of angle from optimal orientation (0°). If a normalised response curve falls below zero (i.e. response < average S/A), then the cell response is inhibited to that orientation of the stimulus.

Having established tuning curves for each cell tested with eight different orientations of the stimulus, an average tuning curve for these cells was produced. This is done by averaging the coefficients obtained from the individual regression analysis for each cell (Fig. 6, thick line).

Population Response to Different Orientations

The average response of 18 cells tested with eight angles was computed to provide an estimate of the total activity amongst the total population of cells responding to the sight of the body within the cortex of the anterior STS, and also to assess how such population activity was affected by stimulus orientation (Fig. 7). Note that this population response profile is different from that shown in Fig. 6, which estimates the average tuning of cells selective for orientation. Figure 7 indicates that, for the population, the response to the upright orientation of a whole body stimulus is higher than to other orientations in the picture plane. A gradual decline away from the upright orientation is observed. Nonetheless, the cell population responded to inverted stimuli at a rate significantly greater than that to controls [PLSD, $P = .005$, ANOVA: $F(8,136) = 11.7$, $P < .0005$]. This is because the population contains cells responsive to all orientations and cells tuned to inverted orientations.

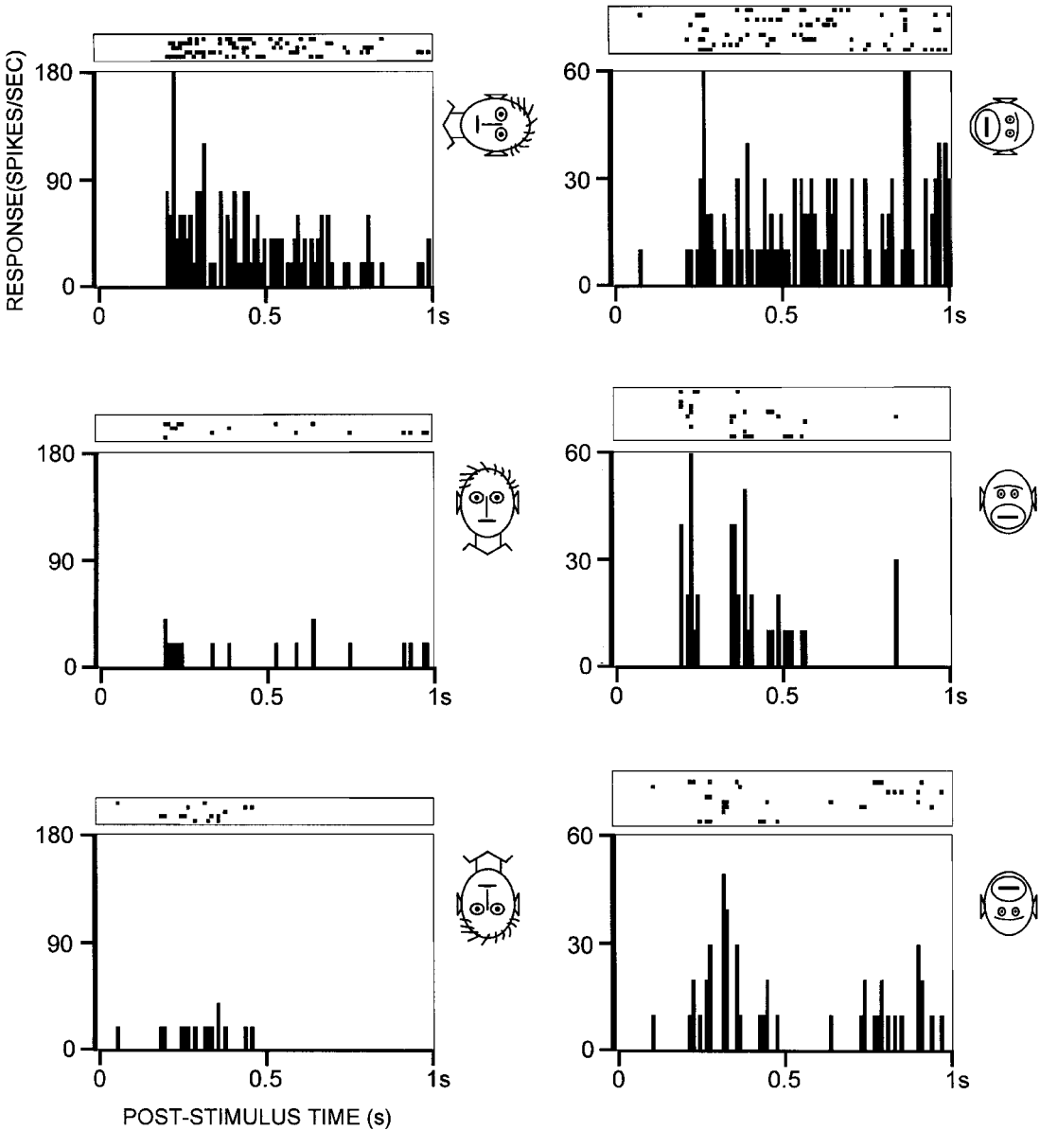

Fig. 3. *The responses of a cell displaying orientation tuning for horizontal faces. Rastergrams and peristimulus time histograms of the responses of one cell (S 128_2980) to different orientations of a human face (left) and a monkey face (right). Bin width is 10msec; spikes are indicated by black squares in rastergrams. The cell responded significantly more to a human face in the horizontal orientation than to upright or inverted orientations ($P < .001$, each comparison), ANOVA: $F(2,12) = 53.9$, $P < .00001$. Similarly for the monkey face, responses to the horizontal orientation were greater than to the upright or inverted orientations ($P < .001$, each comparison), ANOVA: $F(2,26) = 28.1$, $P < .002$.*

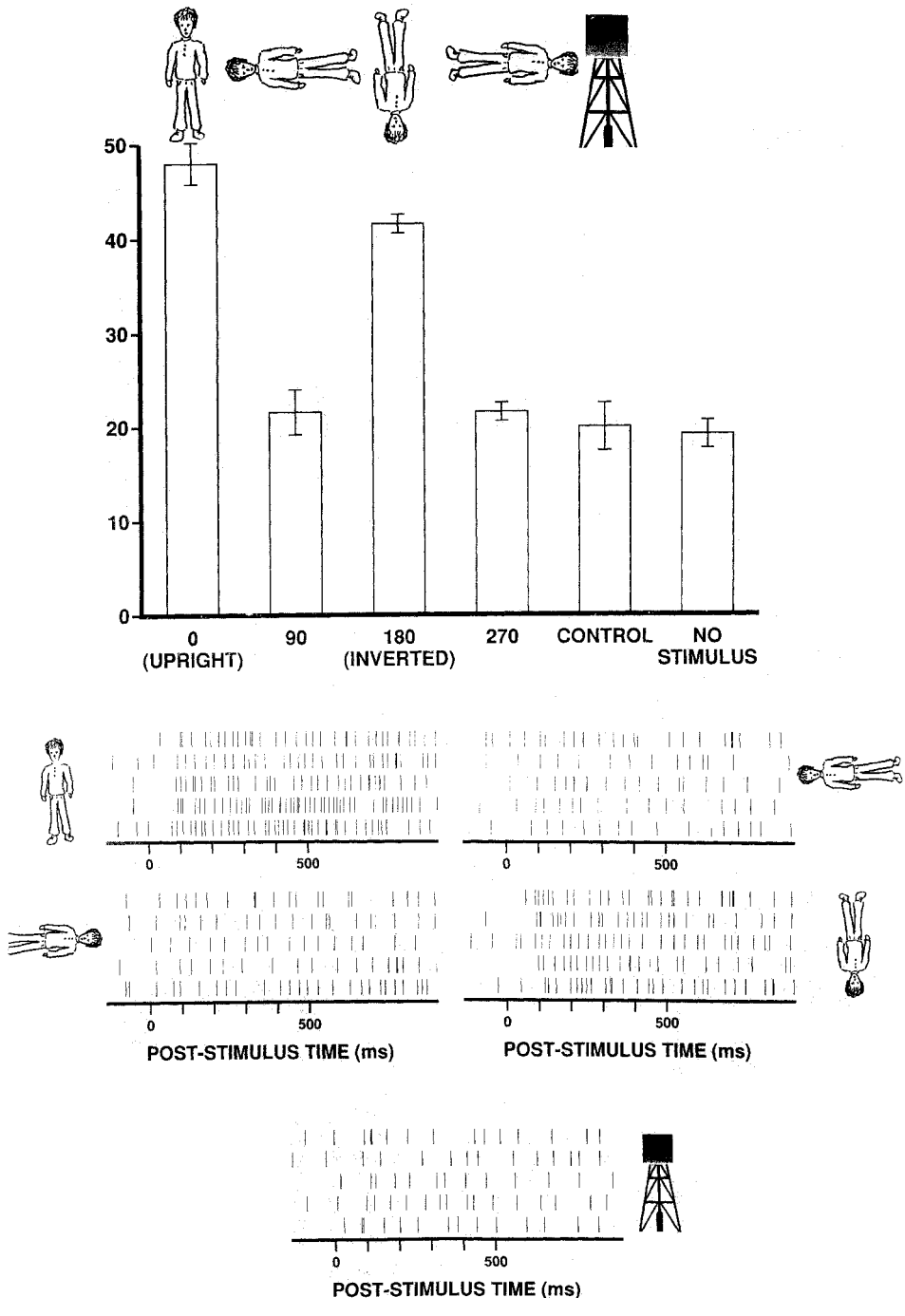

Fig. 4. *The responses of a cell with bimodal tuning for body orientation. Histogram (±1SE) of the responses of a single cell (J69_29.04) to four orientations of the front view of the whole body. The cell responded more to inverted and upright orientations than to other orientations, control objects, or spontaneous activity (P < .0005 each comparison), ANOVA: F(5,24) = 45.8, P < .0005. Lower: Rastergram displays of cell activity (conventions as Fig. 2).*

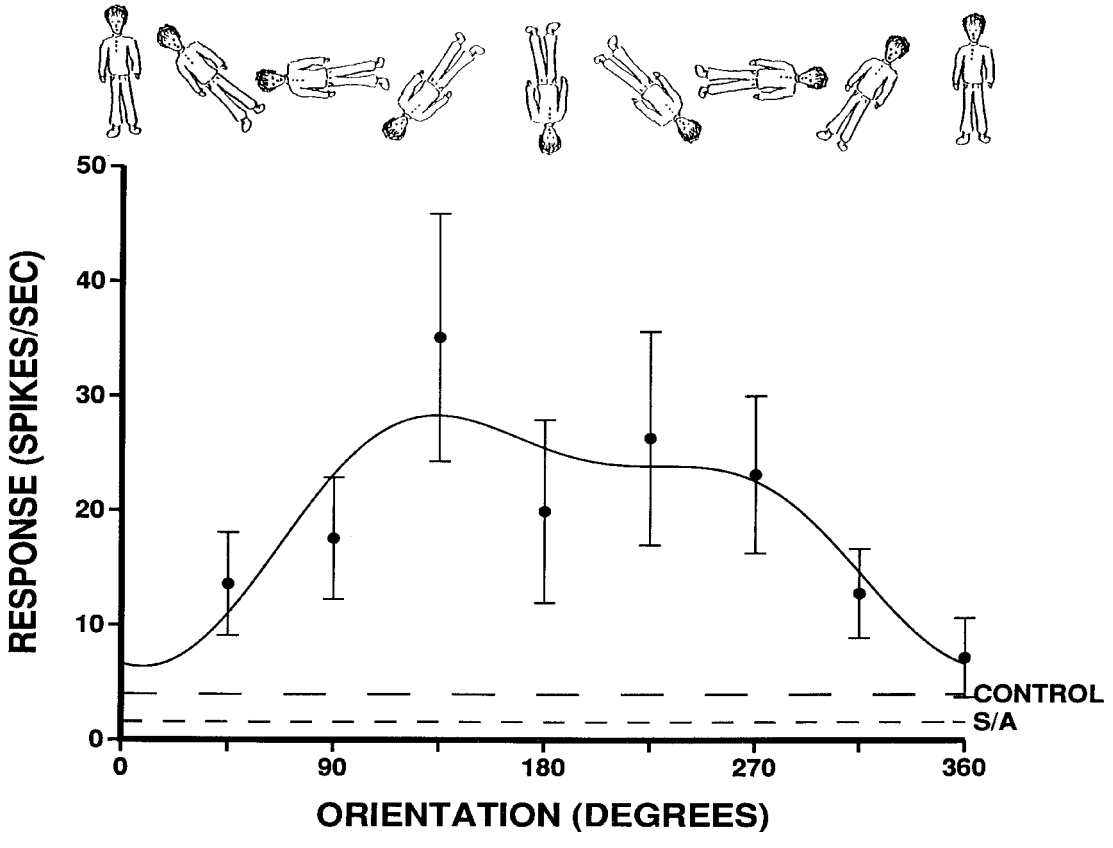

Fig. 5. *The responses of a cell tuned to non-upright body orientations. Histogram (±1SE) of the responses of a single cell (E92_38.20) to eight orientations of the front view of the whole body. The cell responses to orientations close to inverted (135° and 225°) were greater than response to upright (0°), control objects, or spontaneous activity (P < .05 each comparison), ANOVA: F(9,40) = 2.7, P < .02.*

Generalisation Across Image Size

Of the 16 cells tested for size sensitivity, three cells showed size generalisation and responded to all image sizes (100% to 25%) at a rate greater than to S/A and control objects. Of these three cells, one showed no difference in response amplitude to different image sizes, whereas two cells displayed some size tuning superimposed on a generalised response (e.g. Fig. 8).

Sensitivity to Image Size

The remaining 13 cells (81%) were responsive to one (or more) but not all image sizes tested at a rate greater than S/A and control objects, and hence showed size specificity (e.g. Fig. 9).

Thus, the majority of cells (15/16) showed sensitivity to size (including cells that displayed some size tuning superimposed on a generalised response, e.g. Fig. 8). Interestingly, all tuned cells were maximally responsive for the whole body at the largest (100%) or second largest (75%) projection size tested. These two sizes would correspond to images of real humans encountered at the projection distance. As for orientation, the response tuning to image size varied across cells. Some cells showed narrow tuning and responded to only one size of the body at a rate significantly different than to control objects and S/A, others showed broader tuning with the responses declining gradually as the

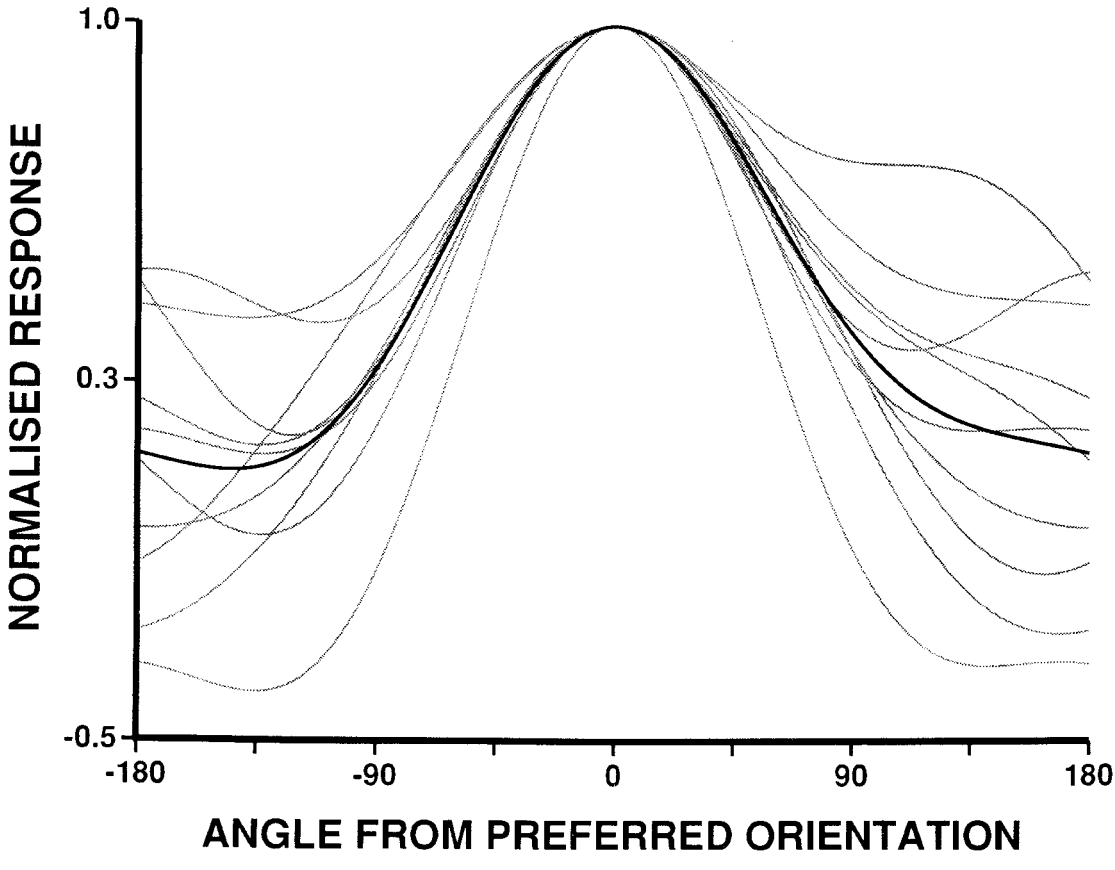

Fig. 6. *Tuning curves for cells sensitive to body orientation. Normalised tuning curves for cells exhibiting orientation sensitivity. The tuning curves are estimated from the best fit cardioid function relating response to an orientation. Spontaneous activity = 0 and maximal response = 1.0. Orientation is expressed as an angle of rotation from optimal orientation (ρmax). Thick line: the average tuning curve from 10 orientation sensitive cells that gave a significant and good fitting regression curve.*

image size progressively decreased from the optimal size.

For comparison with other studies, reclassification was carried out with an effective response defined as greater than half the difference of the cell's maximal response and spontaneous activity. This showed that the majority of cells (75%, 12/16) that were selectively responsive to the body tolerated a size change of less than one octave. Of these, eight cells responded to only one particular size, either 75% or 100%, whereas four cells responded to both 75% and 100% size stimuli but not to the 25% size. The remaining 4/16 cells exhibited a more tolerant response pattern to stimulus size change. Two of the cells tolerated a size change of one octave (from e.g. 50%–100%) and two cells tolerated a stimulus size change of at least two octaves (25%–100%).

Population Response to Different Image Sizes

All 16 cells were included in the population response analysis for Fig. 10. For each cell the response to different sizes was first normalised by setting the largest response (regardless of size) to 1.0 and S/A to 0. Two cells exhibited inhibitory responses and gave activity rates to the body stimuli that were lower than spontaneous activity or control stimuli. For these cells, response normalisation was

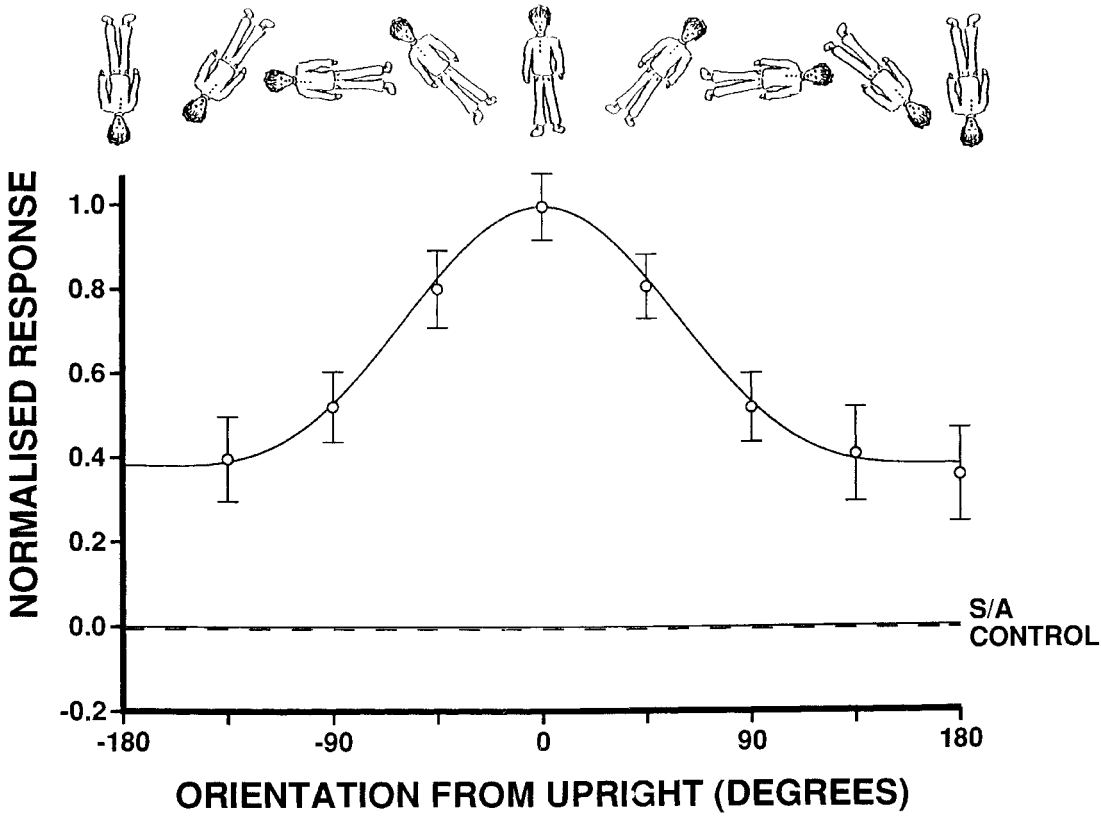

Fig. 7. *Population response to different body orientations. The mean response (± 1SE) for the population of cells responsive to the body at different orientations. The population response to each test condition was computed from the average of the normalised single cell responses (see Method) and was displayed after renormalising, such that the maximum population response = 1 and population spontaneous activity = 0.*

performed in an analogous manner except that the Rmax was defined as the response to the stimulus size that gave the greatest reduction from S/A. The graph displays a smooth degradation of response from maximum (100%) to minimum (25%). Even for the smallest image size of the body (25%), the population response was greater than the response to control, which was approximately zero (i.e. equal to S/A) [PLSD, $P = .021$, ANOVA: $F(4,60) = 24.8$, $P < .0005$].

Histological Reconstruction

The histological reconstruction showed that all cells reported in the present study were located in the upper and lower banks and the fundus of the anterior STS between 5 and 17mm anterior to the interaural plane (see Fig. 11). There was no observed clustering of cell types and no tendency for cells recorded at more anterior sites to show greater generalisation. For subject E, orientation-sensitive cells were found at anterior locations where the sulcus terminates in the temporal pole (see Fig. 11, lower left). Cells exhibiting size selectivity for the body were mainly located in the upper and to some extent in the lower bank of the anterior part of the anterior STS. For subject E, the size-sensitive cells were found in anterior regions of the STS (see Fig. 11, lower right).

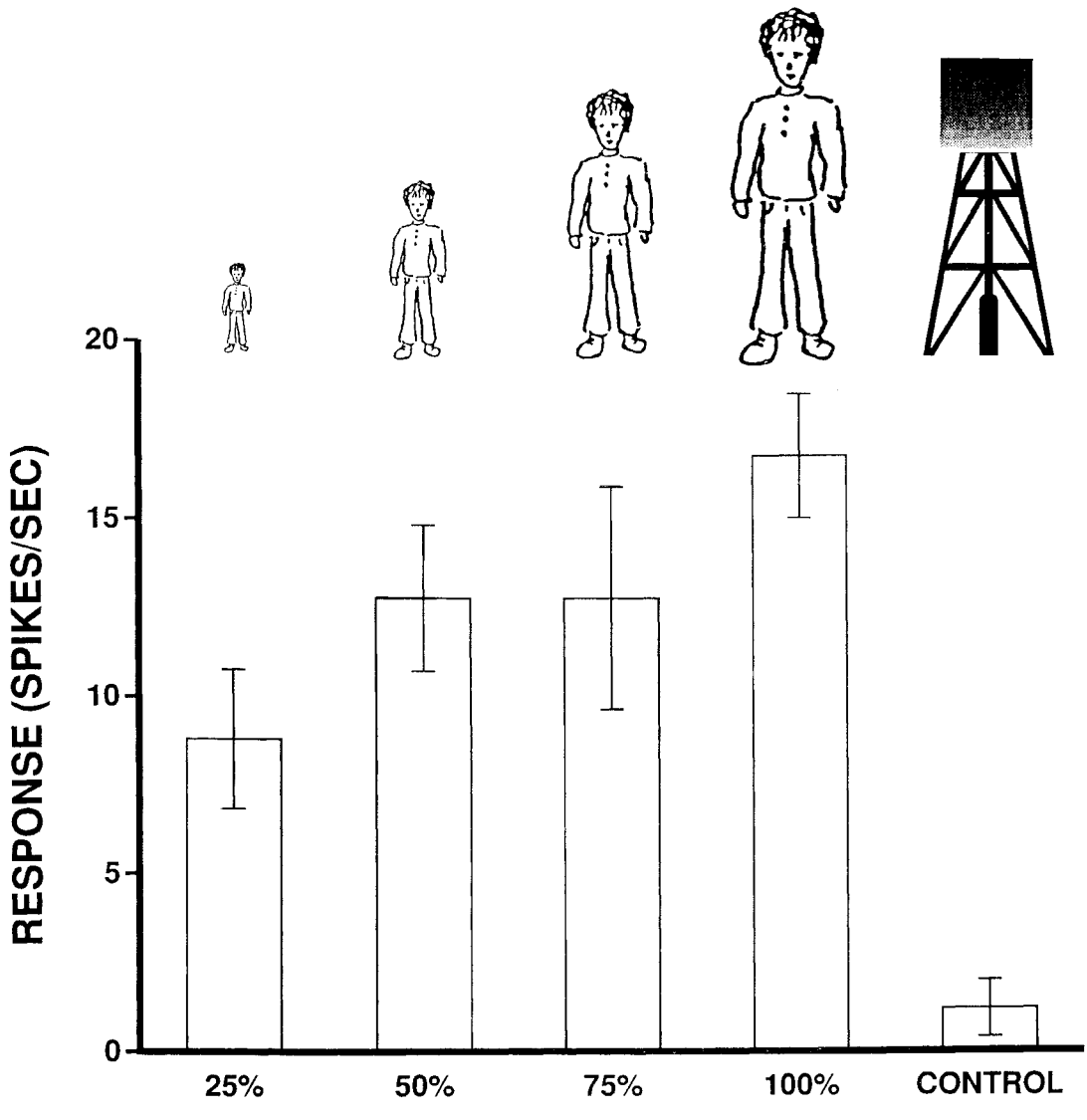

Fig. 8. *The responses of a cell displaying generalisation over body image size. The mean responses (± 1SE) to the whole body (front view) and to control objects are illustrated for one cell (E99_40.16). The cell showed generalisation in that all image sizes of bodies elicited a higher response than spontaneous activity (not shown) and control objects (P < .04 each comparison). The response also indicated size tuning since responses to largest three image sizes (100%–50%) were greater than that to the smallest image size (25%, P < .04 each comparison), ANOVA: F(5,24) = 9.6, P < .0005.*

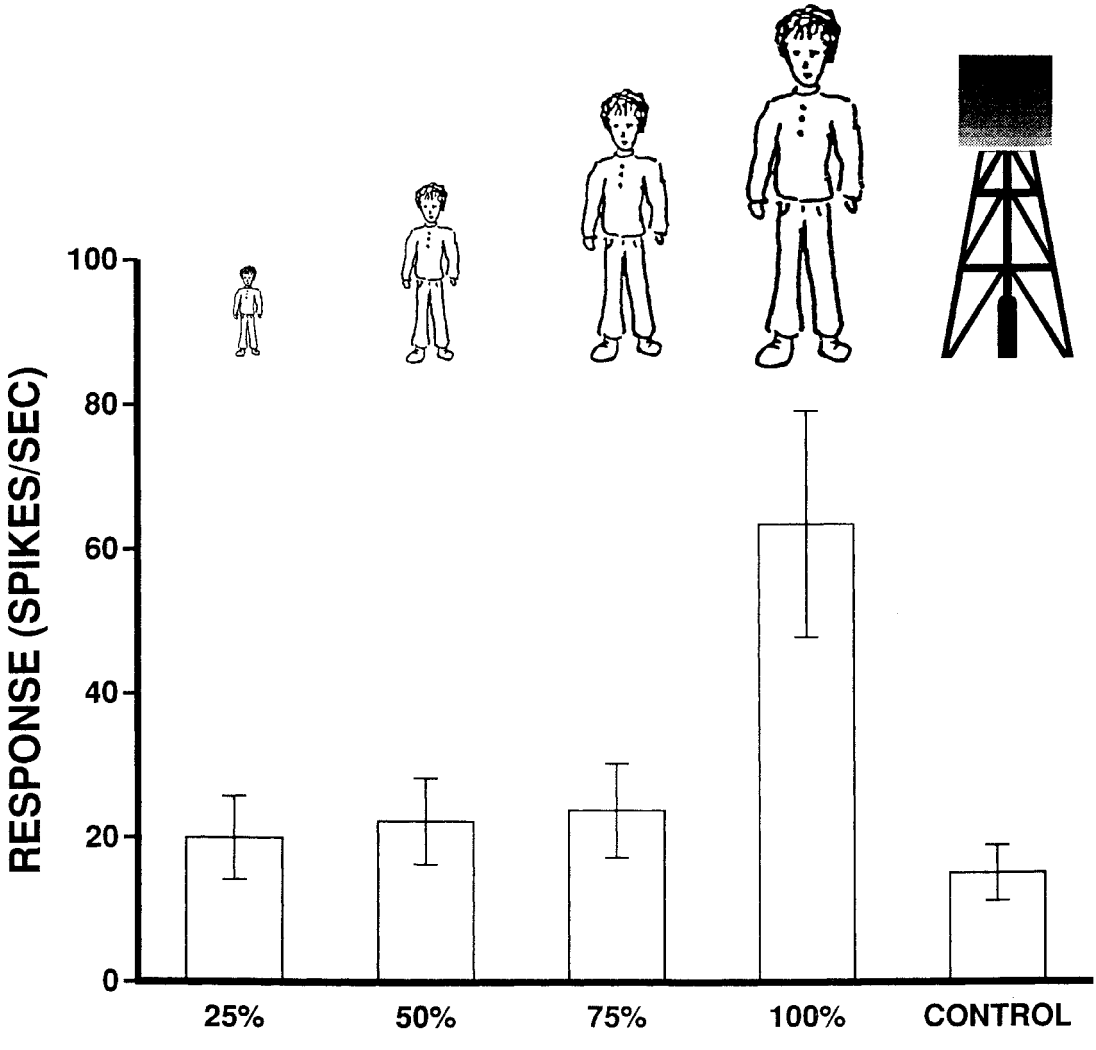

Fig. 9. *The responses of a cell displaying size tuning for large body image size. The mean responses (±1SE) to the whole body (front view) and control objects are illustrated for one cell (E83_38.31). The cell's response to the largest body image (100%) was significantly greater than the response to smaller images of bodies (75%–25%) and controls (P <.003 each comparison). Responses to smaller body stimuli (75%–25%) were not significantly different from control stimuli (P >.4 each comparison), ANOVA: F(4,18) = 5.6, P = .004.*

DISCUSSION

Generalisation Along the Ventral Cortical Stream

It has been argued from previous observations that the further along the ventral stream one investigates the neuronal activity of cells, the greater the percentage of cells that generalise across viewing conditions (such as orientation and size) in which an object is seen (Perrett et al., 1982, 1985, 1988). Studies of single cells in visual areas of the ventral stream leading up to the STSa support the contention that, while the selectivity between objects rises, the sensitivity to particular viewing conditions declines. As the ventral stream is progressed, cells become less sensitive to position (Desimone & Gross, 1979; Ito et al., 1995; Kobatake & Tanaka,

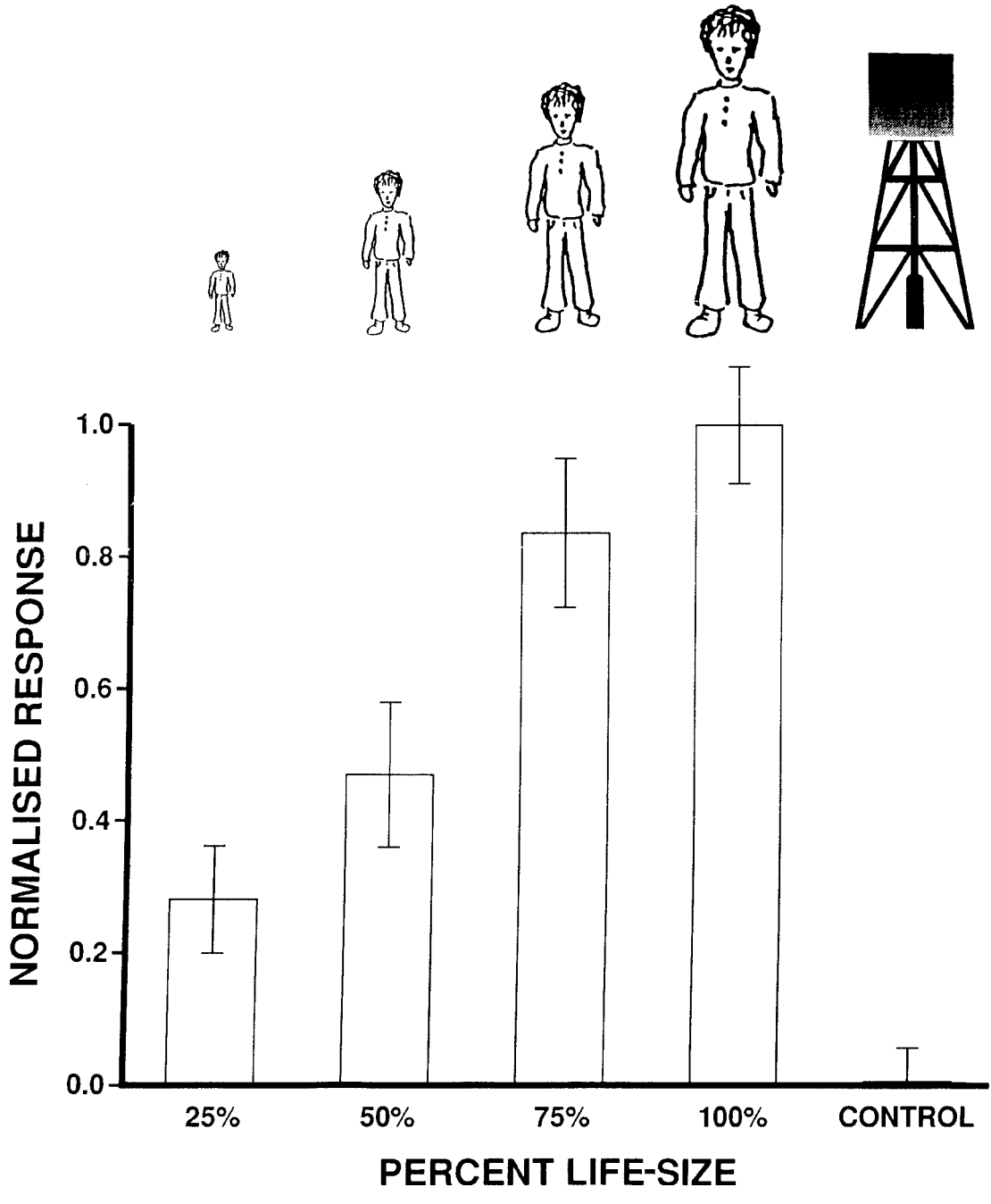

Fig. 10. *Population response to different image sizes of body stimuli. The mean response (±1SE) for the population of 16 cells responsive to the body at different sizes. The population response to each test condition was then computed from the average of the normalised single cell responses (see Method) and was displayed after renormalising such that the maximum population response = 1, and population spontaneous activity = 0.*

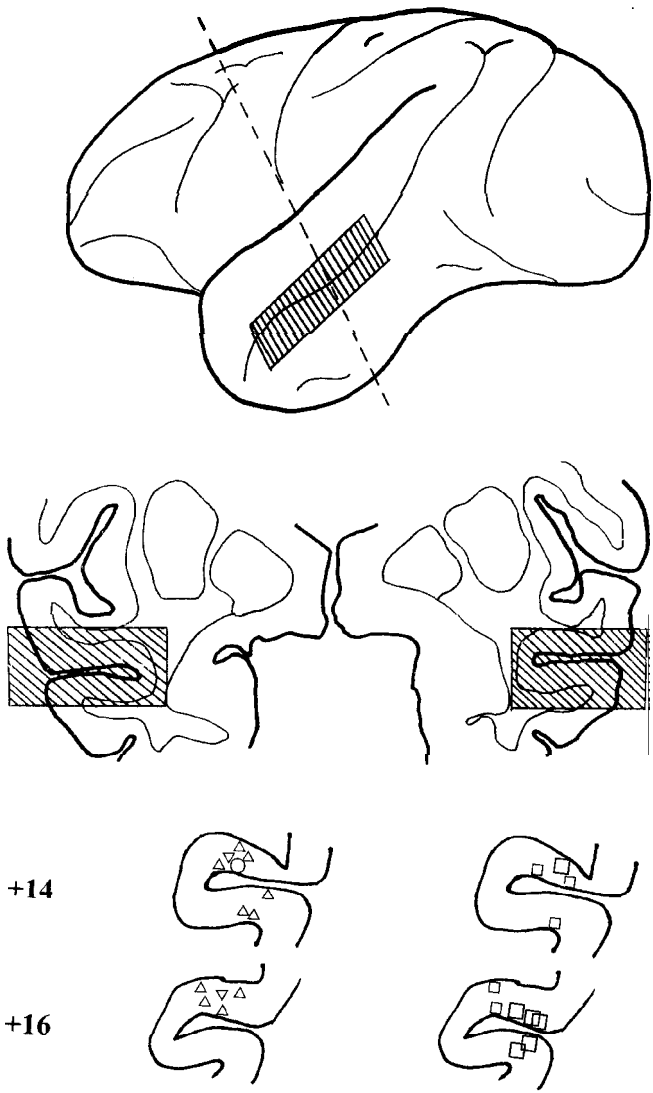

Fig. 11. *Histological reconstruction of position of cells sensitive and insensitive to body orientation and size. Upper: Side view of the brain. Study region (the anterior part of the superior temporal sulcus, STS) indicated by crosshatching, dashed line indicates plane of sections. Middle: A coronal section of the brain 12mm anterior to the interaural plane, anterior STS crosshatched. Lower: Frontal sections of the STSa in the right hemisphere of subject E, 14 and 16mm interior to the interaural plane. Left: The location of cells tuned to upright orientation of the body (triangles pointing up), cells tuned to non-upright orientations of the body (triangles pointing down), and cells generalising across all orientations of the body (circles). Right: The location of cells tuned to the largest (100%) image tested (large squares) and cells tuned to the 75% image size (small squares).*

1994; Tanaka et al., 1991) and show a greater tendency to generalise across object parts (Wachsmuth et al., 1994).

In contrast to the greater generalisation across position and components, the present investigation showed no marked tendency for cells to exhibit greater generalisation across orientation or size at more anterior recording sites. Indeed, cells selectively responsive to the sight of the body, yet showing orientation and size tuning, were found over a range of positions including very anterior sites within the STSa close to the temporal pole. Coding along the ventral pathway, therefore, does not appear to become independent of all viewing conditions.

Building View-general from View-sensitive Representations

Models of object recognition based on physiological evidence suggest the construction of view-independent representations by pooling the outputs across appropriate view-specific representations (e.g. Logothetis et al., 1995; Perrett & Oram, 1993). Such pooling operations can be based on simple learned association between different views of the same object seen in temporal succession as the object is rotated or as the viewer moves around the object (e.g. Perrett & Oram, 1993). Although these pooling operations were first conceived to account for generalisation across perspective view (Perrett et al., 1984, 1985), the same operations could account for generalisation in many domains (Földiák, 1991) including change in lighting conditions (Hietanen, Perrett, Oram, Benson, & Dittrich, 1992) or visibility of object part (Wachsmuth et al., 1994). In this same way responses of neurones selective for the head or body and capable of generalising across orientation (e.g. Fig. 1) could be formed by pooling outputs of orientation-sensitive cells (e.g. Fig. 2–5). Likewise, cells exhibiting size-tolerant responses to one object (e.g. Fig. 8) may owe their response characteristics to the pooled inputs of several cells with size-specific responses to the same object (e.g. Fig. 9). Experiencing the change in image size that happens when an object is approached may provide the basis for the formation of cells with size-tolerant responses through the temporal association of inputs from cells selective for different image sizes.

Orientation-specific Neuronal Coding and Experience

The study reported here has shown that different STSa cell populations are selective for different object orientations of the same complex object (face/body). However, the majority (71%, 15/21) of cells that were sensitive to orientation of the face/body were found to be tuned to the upright orientation. This neuronal response pattern is similar to that found in IT for cells responsive to the face (Tanaka et al., 1991). It is therefore suggested that experience with one (upright) orientation of an object may enhance the neuronal representation for that object in the familiar viewing condition. Monkeys spend the majority of their time with their heads in an upright position and therefore their perceptual experience of heads and bodies will be biased (as in humans) to the gravitational upright. This biased experience may explain the predominance of cells tuned to the upright orientation.

It is interesting to note that the average *orientation* tuning curve for STSa cells responsive to body images is similar to the average *view* tuning curve obtained for STSa cells responsive to the head (Perrett et al., 1991), as well as the average *direction* tuning curve of motion-sensitive STSa cells (Oram et al., 1993).

Advantage of Orientation-specific Coding

What is the function of orientation-specific information in the ventral stream? First, computing such information is a useful and possibly essential stage in the processing leading to the identification of an object, even if later stages of the identification process discard orientation information through generalisation. Second, information about orientation is important in its own right. This is particularly evident for information about the face and body. The STSa appears to be a site that integrates information relevant to the interpretation of social

signals, such as where other individuals are directing their attention and actions (Perrett et al., 1992; Walsh & Perrett, 1994). To recognise where an individual is attending requires orientation-specific and view-specific information about the head posture. The left profile view of the head in its upright orientation may indicate that an individual is attending to the viewer's left, but the same profile view rotated in the picture plane by 45° to 135° from upright (in an anti-clockwise direction) indicates that the individual's attention is directed towards the ground. More generally, to interpret the significance of an individual's posture and actions one needs to specify both the view and the orientation of the face, body, limbs, and hands.

Size-specific Neural Coding and Experience

The majority of cells studied here showed little tolerance to changes in size of body images. Only 4/16 cells tolerated a reduction of stimulus size by one octave (from 100% to 50%) and only two of these cells tolerated a change of two octaves (100% to 25%). Unfortunately, it was not possible to test a greater range of sizes in this study, since at the fixed projection distance of 4m, larger images of whole bodies would not fit onto the projection screen. If larger stimuli were tested, then only part of the body (e.g. a very large head or two extremely large eyes) would be visible on the screen. Previous findings (Perrett et al., 1991) indicate, however, that many cells selective for faces do respond to both very large faces (12.5° visual angle) and life-sized faces (4°) presented at 4m. Thus, the current study might have underestimated size tolerance because extra-large images were not tested. Nonetheless, the study has determined limits in the size tolerance of cells for stimuli decreasing from life-size at a constant viewing distance.

Previous studies have reported greater generalisation to image size in IT cortex. Ito et al. (1995) found 57% (16/28) of cells tolerated a size change of more than two octaves. The study of Ito et al., however, focused on cells responsive to simple geometrical forms. For such stimuli there is no defined life-size. The high degree of size-specific coding reported here may be related to the response selectivity for complex biological stimuli. Life-size can be specified for images of biological objects at a given viewing distance. Studies of cells selectively responsive to faces and bodies show that the majority of size-specific cells are tuned to life-size images of these objects. Rolls and Baylis (1986) reported that 79% of size-specific face-responsive cells in IT and STSa were tuned to large faces (close to real life-size). In the present study, 16/17 cells selective for the head and body were maximally activated by the 75% or the 100% size images. At the projection distance these stimuli were life-sized. Figure 10 shows that the collective response of the cell population studied declines as image size is reduced from life size.

As with orientation tuning, the results for size tuning suggest a role of experience in shaping the response selectivity of cells in temporal cortex. Experience of an object over a particular range of distances and image sizes appears to result in a neuronal representation of that object at that range of image sizes experienced. This could explain the behavioural findings that images of familiar objects close to life-size (or to the image size experienced during training) are recognised faster than non-life-sized images (Besner, 1983; Jolicoeur & Besner, 1987). Indeed the greater the size change of stimuli between training and test, the greater the recognition impairment (Jolicoeur, 1987).

Size and Distance

The sensitivity to size exhibited by STSa cells appears to reflect tuning for "life-size" or absolute size. If this is true, then the cells' sensitivity to angular size (subtended at the retina) should change with distance. A cell selective for life-size bodies and optimally responsive to 18.5° stimuli at a test distance of 4m would be expected to show selectivity for 37° stimuli at a distance of 2m, and 9.75° stimuli at 8m to maintain the same selectivity for absolute size (Rolls & Baylis, 1986). The present study did not examine the possible interactions between sensitivity to angular size and distance but it would be an interesting focus for future work.

This is particularly the case because it is becoming increasingly apparent that information about object position is coded within the ventral stream of cortical processing. Dobbins et al. (1998) recently reported that cell responses within area V4 depend on the distance to visual stimuli. We have also found recently that many of the STSa cells responsive to faces, bodies and their movements are sensitive to distance and position of stimuli (Baker, Keysers, Jellema, & Perrett, in press; Baker, Keysers, Jellema, Wicker, & Perrett, 1999). Although one cannot present a real head at greater than life-size at any distance, with adjustments to viewing optics (e.g. in virtual reality environments) one could begin to isolate cues used by cells in the analysis of size of familiar objects.

Advantage of Size-specific Coding

To identify an object, which is argued to involve the ventral stream of visual processing, it is important to generalise across the different image sizes that occur when the object is seen from different distances. Recognising the specific size of an object, on the other hand, is important for interaction with objects. For example, the gap between the fingers must be adjusted appropriately to match the size of objects in order to pick them up. Cells in parietal cortex which are sensitive to information about object size and orientation (Taira et al., 1990; Sakata & Taira, 1994; Sakata et al., 1997) could be involved in guiding motor movements to enable such prehensile interactions with objects (Jeannerod, Arbib, Rizzolatti, & Sakata, 1995).

However, size can also carry information for social interactions between animals and therefore does play a role in object recognition. If, for instance, a monkey or human face is visible but very small, it does not constitute a threat. The face is either very small in absolute size, or it is a great distance from the viewer; neither constitute immediate danger. The size selectivity of the neuronal responses of cells located in the anterior part of the ventral stream may reflect this advantage of size-specific coding.

Behavioural Reaction Time and Neural Response Latency to Different Orientations or Sizes

The present study revealed that the population of STSa neurones activated by the sight of the head and body was less responsive the more the image orientation departed from upright (see Fig. 7) or was reduced from life-size (see Fig. 10). That is, more cells selectively responded to familiar orientations or sizes.

Such preferential representation of familiar orientations/sizes is reflected in the way activity accumulates in the entire cell population (Perrett, 1996; Perrett, Oram, & Ashbridge, 1998). Activity evoked by the face and body in familiar orientations and image sizes accumulates faster than that evoked by unusual orientations and sizes. Therefore at any given response threshold of the whole cell population, which might for example trigger behavioural reactions, responses to upright and life-size stimuli result in the shortest latencies. In other words the more cells there are responding to one orientation or size the less they have to "compete" with background noise and hence will reach the threshold faster. This argument is made considering the cells as a population. It is also true for the majority of cells considered individually: Weak responses to suboptimal image orientations or sizes commence at the same latency as strong responses to effective orientations and sizes (e.g. Fig. 3).

The decreased neural representation of unusual orientations or sizes may be sufficient to account for the decreased efficiency of recognition at unusual orientations (Jolicoeur, 1985: McMullen & Jolicoeur, 1992; Perrett et al., 1988; Phelps & Roberts, 1994; Tarr, 1995; Tarr & Pinker, 1989). Faces and bodies shown in unusual orientations would take longer to recognise because cell activity to these object orientations or sizes would be weaker and would take longer to exceed any threshold set for indicating the presence of a face/body. This account of recognition does not need to invoke "mental rotation" or "size normalisation" (Besner, 1983; Bundesen & Larsen, 1975; Larsen, 1985; Shepard & Metzler, 1971; Ullman, 1989) to explain increased time for processing

unusual orientations (Perrett, 1996; Perrett et al., 1998).

Further, human scalp recordings of evoked potentials selective for face stimuli show an increase in latency and a decrease in amplitude of potentials as the face is rotated away from the upright orientation to the horizontal orientation (Jeffreys, 1989, 1993) and a small decrease in amplitude and latency for smaller faces (Jeffreys, 1989; Jeffreys et al., 1992). These evoked potentials presumably reflect large populations of neurones activated by the sight of faces and again indicate that neural activation is weaker to progressively more unusual orientations and sizes.

In conclusion, the physiological findings suggest a novel account for the recognition of familiar objects from unusual views. This account does not rely on "mental adjustment of size" (Besner, 1983; Bundesen & Larsen, 1975; Larsen, 1985; Shepard & Metzler, 1971) or "mental rotation" (Perrett, 1996; Perrett et al., 1998). Changes in the speed of recognition are taken to reflect the strength of size-, orientation-, and view-specific, neural representations of objects rather than the operation of transformational processes.

REFERENCES

Ashbridge, E., & Perrett, D.I. (1998). The physiology of shape generalisation (size and orientation). In: *Perceptual constancies: Why things look as they do* (pp. 192–209). Cambridge: Cambridge University Press.

Baker, C.I., Keysers, C., Jellema, T., & Perrett, D.I. (in press). Coding of spatial position in the superior temporal sulcus of the superior temporal sulcus of the macaque. *Current Psychology Letters.*

Baker, C.I., Keysers, C., Jellema, T., Wicker, B., & Perrett, D.I. (1999). *Temporal cortex and object permanence: Cells responsive to visual stimuli occluded from sight.* Manuscript submitted for publication.

Besner, D. (1983). Visual pattern recognition: Size preprocessing re-examined. *Quarterly Journal of Experimental Psychology, 35,* 209–216.

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

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 214–220.

Desimone, R., & Gross, C.G. (1979). Visual areas in the temporal cortex of the macaque. *Journal of Neuroscience, 4,* 2051–2062.

Dobbins, A.C., Jeo, R.M., Fiser, J., & Allman, L.M. (1998). Distance modulation of neural activity in the visual cortex. *Science, 281,* 552–555.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computing, 3,*194–200.

Henry, G.H., Dreher, B., & Bishop, P.O. (1974). Orientation specificity of cells in cat striate cortex. *Journal of Neurophysiology, 37,* 1394–1409.

Hietanen, L.K., Perrett, D.l., Oram, M.W., Benson, P.J., & Dittrich, W.J. (1992). The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Experimental Brain Research, 89,* 157–171.

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

Jeannerod, M., Arbib, M.A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends in Neuroscience, 18,* 314–320.

Jeffreys, D.A. (1989). A face-responsive potential recorded from the human scalp, *Experimental Brain Research, 78,* 193–202.

Jeffreys, D.A. (1993). The influence of stimulus orientation on the vertex-positive scalp potential evoked by faces. *Experimental Brain Research, 96,* 163–172.

Jeffreys, DA, Tukmachi, E.S.A. & Rockley, G.G. (1992). Evoked potential evidence for human brain mechanisms that respond to single, fixated faces. *Brain Research, 91,* 315–362.

Jolicoeur, P.E. (1985). The time to name disorientated natural objects. *Memory and Cognition, 13,* 289–303.

Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. *Memory and Cognition, 15,* 531–543.

Jolicoeur, P., & Besner, D. (1987). Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 478–487.

Kobatake, E., & Tanaka, K. (1994). Neural selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology, 71,* 856–867.

Larsen, A. (1985). Pattern matching: Effects of size ratio, angular difference in orientation, and familiarity. *Perception and Psychophysics, 38,* 63–68.

Logothetis, N.K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5,* 552–563.

McMullen, P.A., & Jolicoeur, P. (1992). The reference frame and effects of orientation on finding the top of rotated objects. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 807–820.

Oram, M.W., Perrett, D.I., & Hietanen, J.K. (1993). Directional tuning of motion-sensitive cells in the anterior superior temporal polysensory area of the macaque. *Experimental Brain Research, 97,* 274-294.

Perrett, D.I. (1996). View-dependent coding in the ventral stream and its consequences for recognition. In: R. Caminiti, K.-P. Hoffmann, & A.J. Lacquaniti (Eds.), *Vision and movement mechanisms in the cerebral cortex* (pp. 142–151). Strasbourg: HFSP.

Perrett, D.I., Hietanen, J.K., Oram, M.W., & Benson, P.J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London, B, 335,* 23–30.

Perrett, D.I., Mistlin, A.J., Chitty, A.L, Smith, P.A.J., Potter, D.D., Broennimann, R., & Harries, M.H. (1988). Specialised face processing and hemispheric asymmetry in man and monkey: Evidence from single unit and reaction time studies. *Behavioural Brain Research, 29,* 245–258.

Perrett, D.I., & Oram, M. (1993). The neurophysiology of shape processing. *Image Visual Computing, 11,* 317–333.

Perrett, D.l., Oram, M.W., Harries, M.H., Bevan, R., Hietanen, J.K., Benson, R.J., & Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86,* 159–173.

Perrett, D.I., Oram, M.W., Hietanen, J.K., & Benson, P.J. (1994). Issues of representation in object vision. In M.J. Farah & G. Ratcliff (Eds.), *The neuropsychology of high-level vision* (pp.33–61). Lawrence Erlbaum Associates Inc.

Perrett, D.l., Oram, M.W., & Ashbridge, E (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition, 67,* 111–145.

Perrett, D.l., Rolls, E.T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329–342.

Perrett, D.l., 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 organisation, sensitivity to identity, and relation to perception. *Human Neurobiology, 3,* 197–208.

Perrett, D.l., Smith, P.A.L., Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., & Jeeves, M.A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society, London B, 223,* 293–317.

Phelps, M.T., & Roberts, W.A. (1994). Memory for pictures of upright and inverted primate faces in humans (*homo sapiens*), squirrel monkeys (*saimiri sciureus*), and pigeons (*columba livia*). *Journal of Comparative Psychology, 108,* 114–125.

Rolls, E.T., & Baylis, G.C. (1986). 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. *Experimental Brain Research, 65,* 38–48.

Sakata, H., & Taira, M. (1994). Parietal control of hand action. *Current Opinion in Neurobiology, 4,* 847–856.

Sakata, H., Taira, M., Murata, A., Galese, V., Tanaka, Y., Shitake, E., & Kusunoki, M. (1997). Parietal visual neurons coding 3-D characteristics of objects and their relation to hand action. In O. Karnarth & P. Their (Eds.), *Parietal lobe contributions to orientation in 3D space.* New York: Springer.

Seltzer, B., & Pandya, D.N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Research, 149,* 1–24.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Snedecor, G.W., & Cochran, W.G. (1980). In *Statistical methods* (7th ed., pp. 215–237). Ames, IA: Iowa State University Press.

Taira, M., Mine, S., Georgopoulos, A.P., Murata, A., & Sakata, H. (1990). Parietal cortex neurons of the monkey related to the visual guidance of hand movements. *Experimental Brain Research, 83,* 29–36.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66,* 170–189.

Tarr, M.J. (1995). Rotating objects to recognise them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin Review, 2,* 55–82.

Tarr, M.J., & Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology, 5,* 233–282.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32,* 193–254.

Wachsmuth, E., Oram, M.W., & Perrett, D.A. (1994). Recognition of objects and their component parts: Responses of single units in the temporal cortex of the macaque. *Cerebral Cortex, 4,* 509–522.

Walsh, V., & Perrett, D.I. (1994). Visual attention in the occipitotemporal processing stream of the macaque. *Cognitive Neuropsychology, 11,* 243–263.