Exp Brain Res (1991) 86:159–173

Experimental Brain Research

Viewer-centred and object-centred coding of heads in the macaque temporal cortex

D.I. Perrett, M.W. Oram, M.H. Harries, R. Bevan, J.K. Hietanen, P.J. Benson, and S. Thomas

Psychological Laboratory, University of St Andrews, Fife, KY16 9JU, UK

Summary. An investigation was made into the sensitivity of cells in the macaque superior temporal sulcus (STS) to the sight of different perspective views of the head. This allowed assessment of (a) whether coding was 'viewer-centred' (view specific) or 'object-centred' (view invariant) and (b) whether viewer-centred cells were preferentially tuned to 'characteristic' views of the head. The majority of cells (110) were found to be viewer-centred and exhibited unimodal tuning to one view. 5 cells displayed object-centred coding responding equally to all views of the head. A further 5 cells showed 'mixed' properties, responding to all views of the head but also discriminating between views. 6 out of 56 viewer and object-centred cells exhibited selectivity for face identity or species. Tuning to view varied in sharpness. For most (54/73) cells the angle of perspective rotation reducing response to half maximal was 45–70° but for 19/73 it was >90°. More cells were optimally tuned to characteristic views of the head (the full face or profile) than to other views. Some cells were, however, found tuned to intermediate views throughout the full 360 degree range. This coding of many distinct head views may have a role in the analysis of social signals based on the interpretation of the direction of other individuals' attention.

Key words: Viewer-centred – Object-centred – Characteristic views – Face coding – Single unit – Macaque

Introduction

Viewer and object-centred coding in models of recognition

Visual recognition of objects is a process of comparing sensory information with internal representations of objects. Representation is used here to refer to the neural code or description of an object's attributes and appearance. The type of representation involved must be able to account for the phenomenon of object constancy, that is the ability to extract knowledge of the unchanging three dimensional structure of an object from a changing two dimensional retinal image. Two major types of stored representations (or descriptions) have been suggested which could account for this. These have been termed viewer-centred and object-centred (for discussion see Marr 1982; Marr and Nishihara 1978; Feldman 1989; Hinton and Parsons 1988; Rock and di Vita 1987).

Viewer-centred coding depends on the position of the viewer relative to the object being recognized. A viewer-centred description of an object is specific to the particular viewpoint from which the object is seen. Separate viewer-centred representations are therefore needed to enable recognition of the object from different perspective views. Such coding poses the problem that different views of a particular object would have to be treated as separate objects. Learning associations between one view of an object and some property would not enable the retrieval of this property when a different view of the object is encountered.

These problems are avoided using an object-centred representational system. Under this system features of the object are related not to the viewer but to some major part of the object itself (such as the longest axis). Although the apperance of features of an object change relative to the viewer when the angle of view is changed, their orientation relative to a point of reference on the object itself remains constant. [The head and legs are at opposite ends of the torso exemplifies an object-centred description of a human figure, and is valid for any viewpoint]. Theoretically only one object-centred description of an object would have to be coded for recognition to be possible from any view.

Characteristic views

Marr and Nishihara (1978) suggested that object-centred descriptions could be computed directly from low level

Offprint requests to: D.I. Perrett (address see above)

descriptions of surfaces relative to the viewer. Such computation is, however, likely to be complex, though progress has been made in this framework (see Lowe 1987). While viewpoint independent recognition may be an aim of visual processing, this could be achieved by combining several high level view-specific descriptions of an object.

A limited capacity to generalize across vantage point would allow recognition to be based on a small number of stored (viewer-centred) descriptions of an object from particular 'characteristic' views (e.g. Koenderink and van Doom 1976, 1979; Perrett et al. 1985a). Theoretical and computational models of visual recognition based on a limited number of views are becoming increasingly prevalent. Though different models suggest different numbers of view-specific templates need to be stored to allow view invariant recognition (Baron 1981; Ullman 1989; Poggio and Edelman 1990; Seibert and Waxman 1990). Thus the number of characteristic views necessary to represent an object and the manner in which the views can be defined are both controversial (Perrett and Harries 1989).

Physiological evidence for representations

Viewer-centred coding of heads. Cells have been found in various regions of the temporal cortex which are selectively activated by the sight of biologically important stimuli such as faces, hands and bodies (Gross et al. 1972). Studies of cells in this area can therefore shed light on the way such objects are represented in the nervous system.

The majority of cells responsive to the sight of the head are selective for particular perspective views. Subpopulations of cells in the superior temporal sulcus (STS) respond selectively to different views of the head, some respond most to the full face view, others to the profile view (Perrett et al. 1982, 1984, 1985a; Bruce et al. 1981; Desimone et al. 1984; Hasselmo et al. 1989a; Kendrick and Baldwin 1987). The cells show considerable generalization for the preferred view across changes in retinal position (Desimone et al. 1984; Bruce et al. 1981; Perrett et al. 1989a), size and distance (Perrett et al. 1982, 1984; Rolls and Baylis 1986), isomorphic orientation (upright or rotated horizontal, Perrett et al. 1982, 1984, 1985a, 1988) and lighting (Perrett et al. 1982, 1984). These findings indicate that the cells are not responding to simple visual features (local edges, texture etc.) since these change with image size, position and orientation. Instead the cells appear to represent high level descriptions of properties which are invariant across distance, orientation and size.

A cell tuned to one perspective view of the head can therefore be seen as providing a high level viewer-centred description of this object. Only a limited number of such high level descriptions need exist to cover all the possible ways in which a head can be seen. From the initial studies of view (Perrett et al. 1985a, 1987) it appeared that cells were selectively tuned for just 4 'characteristic' views in the horizontal plane (face, left and right profiles and the back of the head). Approximate estimates of tuning indicated that for most cells, 45–90° of rotation of the head reduced the magnitude of response to half that of the optimal view). With this width of tuning, a minimum of four populations of cells (each tuned to one of the four characteristic views) could cover all views in the horizontal plane, including the intermediate views such as the half profile.

More recent physiological studies have questioned the importance of the putative characteristic views of the head. Hasselmo et al. (1989a) found that more cells were responsive to front views of the head than to back views but found no other evidence that 4 views were selectively coded, and Perrett et al. (1989a) suggested that all views might be represented evenly.

Psychological studies also have disputed the importance of different views of the head. Harries et al. (1990) found that face and profile views were the most important for coding and recognition whereas other studies have stressed the importance of the half profile view, 45° from the full face (Thomas et al. in prep. Bruce et al. 1987; Logie et al. 1987).

While physiological and psychological evidence both demonstrate that perspective view is of central importance to the recognition of heads it is by no means clear whether particular views receive preferential coding.

Object-centred coding of heads. In the superior temporal sulcus populations of cells have also been found to respond to all views of an object that were tested. Perrett et al. (1985a) found that 25% of cells responding to the face were relatively insensitive to viewpoint, responding equivalently to different views of the head rotated in the horizontal plane. These cells appeared to exhibit object-centred coding. We have suggested elsewhere that the view-invariant coding of such cells could be established hierarchically by combining the outputs of cells selective for particular views (Perrett et al. 1984, 1985a, 1989a). In essence this scheme amounts to establishing object-centred descriptions by combining the outputs of several viewer-centred descriptions.

Initial studies suggested that cells responsive to heads responded similarly to different individuals (Perrett et al. 1982). More recent investigations, however, suggest that a fraction of the cells (10–50% depending on the study) discriminate between different species or between individuals of the same species (Perrett et al. 1984; Desimone et al. 1984; Rolls 1984, 1987; Leonard et al. 1985; Baylis et al. 1985; Kendrick et al. 1987; Yamane et al. 1988; Hasselmo et al. 1989a). These cells may be regarded as representing viewer or object-centred descriptions of familiar individuals depending on their generalization over perspective view (Perrett et al. 1984, 1987, 1989a; Hasselmo et al. 1989a).

Hasselmo et al. (1989a) compared cell responses to two different individuals in different views using 2-way analysis of variance (ANOVA). They found a significant main effect of identity for 18 cells (of 37 tested). Hasselmo et al., interpreted this result as evidence for object-centred coding (for identity). Fifteen of the cells which were sensitive to identity, however, also showed sensitivity to the viewing angle (evidenced by significant main effect of view).

Object-centred coding of body motion. Body movements provide an important means of analyzing the behaviour and intentions of other individuals. It is interesting therefore that neurons sensitive to body movements have also been found in the temporal cortex. These cells provide the strongest evidence of view-independent coding.

Hasselmo et al. (1989a) reported object-centred coding for neurons selective for head movements. For example, some cells responded to the head flexing up relative to the body, and continued to respond when the body was seen from the back or was inverted so that the retinal motion of the head was directed down. The directional selectivity can be understood as an object-centred description in which the head motion is referenced to the torso of the body.

Object-centred coding of limb and whole body movements has been described in several reports (e.g. cells selective for the sight of bringing the arm to the chest (Perrett et al. 1990a, b), walking backwards and walking forwards (Perrett et al. 1985b, 1989a, 1990a, b; Harries et al. in prep.). For a cell responding to walking 'forwards', the front view of the body is optimal when the body approaches the viewer, whereas the back view is optimal when the body retreats away from the observer (Perrett et al. 1985b). Here the view and directional selectivity is understandable as an object-centred description in which body motion is referenced to the direction in which the torso or face is oriented. (Walking forward equals following one's nose.)

Other cells with view-independent responses to body movements use 'goal-centred' coordinates where the direction of movement is related to the goal of the action (examples include: bringing food in the hand to the mouth, reaching for a target; walking toward an external door, Perrett et al. 1989a, 1990a, b). While view independent object- and goal-centred coding of body motion has been demonstrated for some cells in the temporal cortex, most cells responsive to body motion are selective for view.

Aims of the present study

The purpose of the present study was to apply a systematic and quantitative analysis to the tuning for perspective view amongst the population of cells selectively responsive to static views of the head. This analysis had three principle aims.

The first aim was to assess the extent to which responses of single neurons to static information about the head displayed viewer-centred or object-centred properties. In the context of this issue we analysed the effects of both view and identity for some cells, since Hasselmo et al. (1989a) argued that consistent effects of identity across different perspective views indicated object-centred coding.

For view-sensitive cells, the second aim was to determine the distribution of optimal views to examine the extent to which particular 'characteristic' views might be selectively or disproportionately represented.

The third aim was to characterize the tuning function of cell's responses to views deviating from the optimal view. Assessment of the width of tuning for non-optimal views allows evaluation of the number of different views that need to be represented to accommodate recognition from any view. These data are important in assessing the biological applicability of different computational models of object recognition.

Preliminary reports of some of the results have been presented elsewhere (Perrett et al. 1989a, b).

Methods

Subjects

Two female (wt 4 kg) and three male (wt 5–10 kg) rhesus macaque monkeys were used. The monkeys are referred to as F, J, B, D and H.

Fixation task

Before beginning recording the subjects were trained to discriminate between the red or green colour of an LED light. The LED was situated level with the monkey's line of sight on a blank white wall at a distance of 4 m. The LED and test visual stimuli were presented from behind a large aperture (6.5 cm diameter) electromechanical shutter (Compur) or an alternative (20 cm square) liquid crystal shutter (Screen Print Technology Ltd.) Both types of shutter had rise times of < 15 ms. On each trial the shutter was opened (after a 0.5 s signal tone) to reveal the stimulus and remained open for a period of 1 s.

The LED light became visible at the time of shutter opening (stimulus presentation) and was randomly red or green of different trials. The monkeys were trained to lick for fruit juice reward on trials with a green LED. On trials with a red LED they were trained to withhold response to avoid saline solution. Subjects were deprived of water for periods of up to twenty-four hours before training and recording sessions to motivate task performance.

The monkeys attended to the LED at the beginning of trials in order to lick several times for multiple juice rewards in the 1.0 s trial period. The 2D test stimuli were projected onto the wall on which the LED was located, 3D test stimuli were presented in front or to either side of the LED. In this way the monkey's attention was directed towards the experimental stimuli. The monkeys performed the task at a high level of accuracy and independent of simultaneously presented 2D test stimuli.

Recording procedures

Each monkey was sedated with a weight-dependent dose of intramuscular ketamine and anaesthetised with intravenous barbiturate (Sagatal). Full sterile precautions were then employed to implant 2 stainless steel recording wells (16 mm internal diameter, ID) 10 mm anterior to the interaural plane and 12 mm to the left and right of midline. Plastic tubes (5 mm ID) were fixed horizontally with dental acrylic in front and behind the wells. Metal rods could be passed through these tubes to restrain the monkey's head during recording sessions.

For each recording session topical anaesthetic, lignocaine hydrochloride (Xylocaine 40 mg/ml) was applied to the dura and a David Kopf micro-positioner fixed to the recording well. A transdural guide tube was inserted 3–5 mm through the dura and a tungsten in glass microelectrode (Merrill and Ainsworth 1972) advanced with a hydraulic micro-drive to the temporal cortex. The target area for recording was the anterior part of the upper bank of the STS (areas TPO, PGa, TAa of Seltzer and Pandya 1978).

Localization of recording

Following the last recording session, a sedating dose of ketamine was administered followed by a lethal dose of barbiturate anaesthetic. The monkey was then perfused transcardially with phosphate buffered saline and 4% gluteraldehyde/paraformaldehyde fixative. The brain was removed and sunk in successively higher concentrations (10, 20 and 30%) of sucrose solution or 2% Dimethylsulphoxide (DMSO) and 20% glycerol (Rosene et al. 1986).

Frontal and lateral X-radiographs were taken of the position of microelectrodes at the end of each recording session. Reconstruction of electrode possition was achieved by reference to the positions of micro-lesions (10 microamp DC for 30 s) made at the end of some electrode tracks which were subsequently identified using standard histological techniques. In 3 monkeys additional markers used in calibration of electrode position were provided by microinjection of anatomical tracers (horseradish peroxidase and fluorescent dyes true blue and diamadino yellow) at the site of cell recording on 3 recording tracks. For these markers the position of injection, recorded in X-radiographs, could be compared to the anatomical location of injection revealed through normal or fluorescence microscopy.

Recording methods

Subjects were restrained in a primate chair for periods of 2–4 h. Various types of visual stimuli were presented while the monkeys performed the fixation task (see below). Neuronal firing rates were measured using standard techniques in a period of 250 ms beginning 100 ms after stimulus presentation. [A 500 ms sample period was occasionally used for cells with small or late responses.] These data were analysed on-line by a microcomputer Cromemco System 3 or AT compatible PC (Hyundai, Dell).

Horizontal and vertical eye movements were monitored using an infra-red corneal reflection system (ACS, modified to allow recording of both signals from one eye) to determine whether any response differences reflected differential patterns of fixation.

Visual stimuli

Responses were measured to both real 3D heads (the experimenters') and 2D heads (video disk images and slides of the heads of humans and macaque monkeys). Four or eight different views of stimuli were tested. The views included four hypothetical characteristic views, namely the face (0°), left profile (90°), back of head (180°) and right profile (270°); plus the four intermediate views: 45°, 135°, 225° and 315°.

Responses to heads were compared to responses to a variety of control stimuli. These included a collection of real (3D) objects of differing size, shape and texture and a large collection of 2D stimuli (slides and video disk images of single objects or complex scenes) and simple geometrical images (bars, spots, gratings etc., generated on-line using a Fairlight Computer Video Instrument).

Specific controls such as a hand or photographs of monkey paws, wigs, and pieces of artificial fur were used to test whether cells responding to the face or head responded to simple features such as hair/fur texture, or skin/fur colour. For example a cell apparently responding to all views of a head might respond because of the presence of hair, a feature visible in any view.

Testing procedure

Each cell recorded was first subjected to exploratory testing involving the presentation of a variety of static and moving control objects. Potential response to heads was first examined 'clinically' with a real head presented in 8 different views. Testing associated with other experiments involved presenting tactile, auditory stimuli and 4 views of the whole body walking (forwards and backwards).

Cells which showed any tendency to discriminate one or more views of a head from control objects were then tested with 5 trials of four or eight views of head and various controls presented in a computer controlled and randomized order. Testing was performed in one mode using either real 3D, projected 2D slides or video disk stimuli. Computer-controlled testing protocols enabled data to be subjected to ANOVA and regression analysis on-line. Cells showing significant tuning for view were subjected to further study using different modes of presentation 2D/3D, identities or species of head, and for effects (to be reported elsewhere) of motion, gaze direction, lighting, and vertical head posture.

Data analysis

ANOVA. Cell responses to 4 or 8 views, controls and spontaneous activity were compared on line using 1-way ANOVA and post-hoc tests (protected least significant difference (PLSD), Snedecor and Cochran 1980). If more than one analysis was performed on a cell's responses (e.g. for different heads or time periods) the most statistically significant results were used to classify the cell.

Regression. For cells tested with eight views multiple linear regression analysis was used to estimate the best relationship between response and 2nd order cardioid function of angle of view of the head. In effect this calculates the values of the coefficients β_{1-5} of the Eq. (1) below which produce the highest correlation between response and the angle of view.

$$R = \beta_1 + \beta_2 \cos(\theta) + \beta_3 \sin(\theta) + \beta_4 \cos(2\theta) + \beta_5 \sin(2\theta)$$

where R is the response, θ is the angle of head view and β_{1-5} are coefficients.

This equation was chosen because it makes very few assumptions about the nature of view tuning. At the outset of the investigation we were aware of only two types of view tuning; cells with a single preferred view and cells with two preferred views approximately 180° apart (e.g. left and right profiles). For a cell with a single preferred view from the 360 degree range the sin θ and cos θ terms specify the angle of best view and describe a monotonic decay of response with angular deviation from optimal view. The second two terms (sin 2θ) and cos (2θ) allow the description of variation in response with view to have two peaks and determine their relative amplitude, separation and sharpness. Thus the full 4 term equation was anticipated to provide a good approximation of the view tuning previously characterised.

Cell responses giving a significant regression analysis were further assessed by statistical comparison (Chi-squared) of the observed response rates with the response rates predicted by equation (1). Chi-squared overestimates discrepancies when predicted responses are small but presented a useful guide to 7 cases where the cardioid function was an inappropriate description of view tuning. These cases were dropped from further analysis.

Where the regression analysis produced a significant ($p < 0.05$) relation between predicted and observed values, the regression equation was used to define: (a) the optimal angle of view (θmax), (b) the maximum response at this view (Rmax), (c) the sharpness of tuning (average angle of rotation required to reduce the response to half Rmax) and (d) the angle and magnitude of any second peak in the view tuning.

Results

Cell classification

119 cells were classified as *'head-selective'* when statistical analysis revealed their responses to one or more views of the head were significantly greater than responses to

controls and spontaneous activity. These were subdivided as follows:

Viewer-centred cells. 110 cells were classified as *'viewer-centred'* when their responses to one or more (but not all) views were significantly greater than responses to controls and spontaneous activity. Regression analysis revealed a significant relation between response and cardioid function of view for 69 of those cells tested with 8 views. The responses of 99 of the viewer-centred cells followed a *unimodal* pattern, with one view evoking the optimal response and a monotonic decline in response as the head was rotated from that view (e.g. Fig. 1).

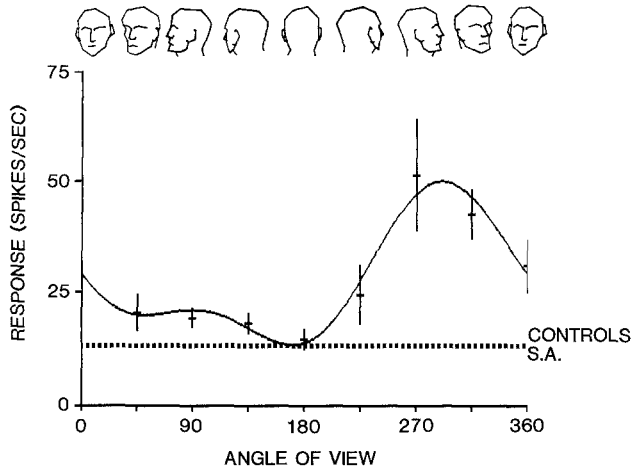

Fig. 1. Responses of a viewer-centred cell with unimodal tuning to perspective view. The mean responses (±1SE) are illustrated for one cell (J012 25.99) to 8 views of the head. View, expressed as the angle of rotation from face, is illustrated schematically at the top. The curve is the best fit cardioid function, relating response to view ($R^2=0.47$; $F(4,35)=7.9$, $p<0.0005$). *Dashed lines* are the mean responses to control stimuli and spontaneous activity (S.A.). Responses to the views close to the left profile (270, 315) were not significantly different (Protected LSD tests $p>0.2$) but were significantly greater than response to all other views, controls and spontaneous activity ($p<0.03$ each comparison). ANOVA: $F(9,40)=5.4$, $p<0.0005$

11 viewer-centred cells were classified as *bimodal* because their responses to two non-adjacent views were both significantly higher than intervening views (either side of the bimodal peaks), controls and spontaneous activity. For 8 cells the two views evoking high responses were approximately 180° apart. In 5 cases these were the profile views (e.g. Fig. 2). Three cells gave a major response to the full face view and a subsidiary response to the back view. The face and back of head have little in common visually but are equivalent in outline and differ from other views in having symmetry. Three cells exhibited bimodal responses for the two half profile views (45 and 315°) only 90° apart. Thus for bimodal cells, 8 were selective for views that were mirror images (5 for profile and 3 for 1/2 profile).

The criteria for classification as bimodal used here was fairly stringent and a further 13 cells showed a degree of bimodal view tuning, in that their response to a second or minor view was greater than half the response to the optimal view. (The optimal and minor views being separated by views evoking responses less than half the maximal response).

Object-centred cells. Four cells were classified as *'object-centred'* because analysis revealed (a) their responses to all views of the head were significantly greater than response to control stimuli and spontaneous activity and (b) their response did not discriminate between any of the head views. The responses of one such cell is illustrated in Fig. 3.

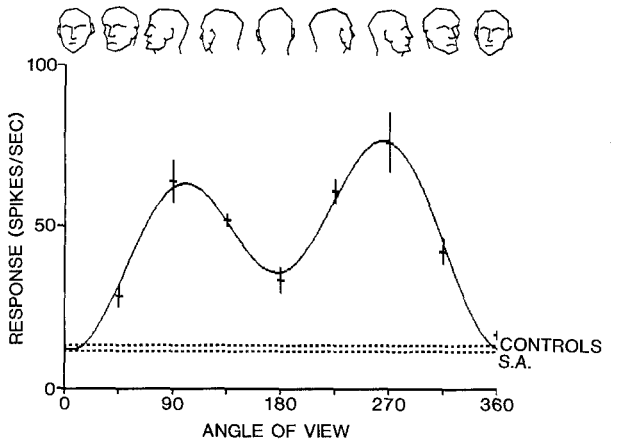

Fig. 2. The responses of a bimodal viewer-centred cell. Responses (mean ±1SE) of cell D023 28.90 to 8 views of the head illustrated schematically at the top. The curve is the best fit cardioid function, relating response to view ($R^2=0.77$, $F(4,36)=29,5$ $p<0.0005$). *Dashed lines* are the mean responses to control stimuli and spontaneous activity (S.A.). Responses to the two profile views (90 and 270°) were not significantly different ($p=0.064$) but were both significantly greater than front (0) and back (180) views, controls and spontaneous activity ($p<0.0005$ each comparison). ANOVA: $F(9,40)=26.2$, $p<0.0005$

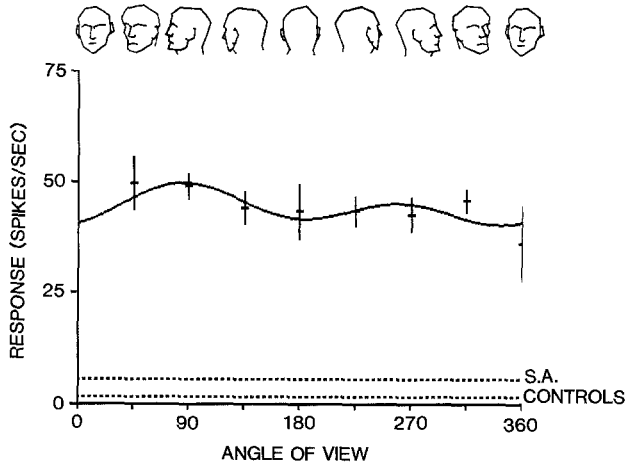

Fig. 3. The responses of a cell displaying object-centred properties. Responses of cell H005 28.16 to all views were higher than to control stimuli and spontaneous ($p<0.0005$), indicating object-centred coding. The cell responses, however, showed little selectivity between views. Protected LSD tests indicated a difference between 0 and 45 degree views ($p=0.042$) but this is likely to reflect a Type I error of statistical interpretation, given overlapping standard errors. There were no other differences in response to different views ($p>0.05$). ANOVA: $F(9,40)=14.6$, $p<0.0005$. Regression analysis: $R^2=0.07$, $F(4,35)=0.6$, $p=0.65$)

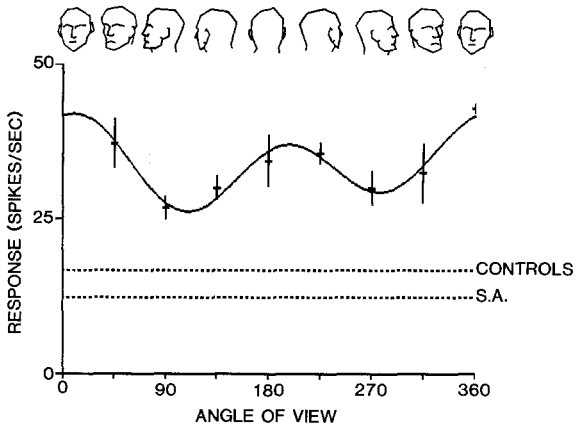

Fig. 4. The responses of a cell displaying mixed object-centred and viewer-sensitive properties. Responses of cell D119 31.17 to all views were higher than to control stimuli and spontaneous activity ($p<0.05$), indicating object-centred coding. The cell responses, however showed selectivity for view with a significantly greater response to the face (0°) view than to views at 315, 135, 270 and 90 degrees ($p<0.05$ each comparison). ANOVA: $F(9,40)=8.8$, $p<0.001$. Regression analysis: $R^2=0.35$, $F(4,35)=4.7$, $p=0.004$

Cells displaying mixed coding. Five cells were found to display mixed properties exhibiting both object-centred and viewer-centred properties in that analysis revealed (a) their responses to all views of the head were significantly greater than responses to control stimuli and spontaneous activity and (b) their responses discriminated between different views of the head. Cells in this group displayed either unimodal view preference or a tendency to bimodal view preferences (e.g. Fig. 4).

Anomalies. One cell discriminated some (but not all) views from controls and spontaneous activity, yet did not discriminate statistically between views. For three other cells regression analysis indicated a significant effect of view on response but more conservative ANOVA failed to confirm a significant effect of view. These cells were view-sensitive but the difference in response between best and worst view was small. They are not considered further in the analysis.

Discrimination between stimuli

Control stimuli. For the majority of head-selective cells, response to control stimuli was not significantly different to spontaneous activity. The average spontaneous activity for the 119 head selective cells was 9.0 ± 0.67 spikes per second (mean $\pm$ 1SE) whereas the average response to control stimuli was 14.9 ± 1.28 spikes per second. The average response to the most effective head view for each cell was 46.6 ± 2.15 ($n=119$).

To contrast cell responses to control stimuli with the response to heads an index of discrimination was computed for each cell using the formula (Rctrl-SA)/(Rmax-SA); where Rmax = mean response to the optimal view, Rctrl = mean response to controls and SA = mean spontaneous activity. The distribution of the relative response to control stimuli is illustrated for cells

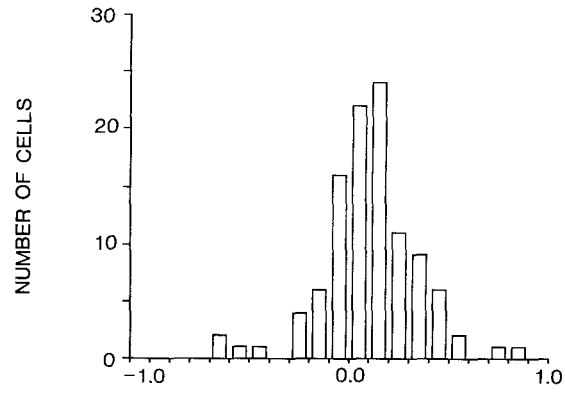

Fig. 5. Discrimination between optimal view and control stimuli. Responses to control stimuli (Rctrl) are expressed as a fraction [(Rctrl-SA)/(Rmax-SA)] of responses to the optimal view of the head (Rmax). SA = spontaneous activity

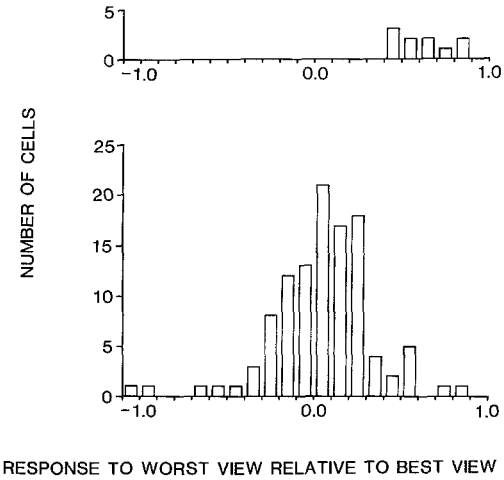

Fig. 6. Discrimination between views. The responses to the least effective view of the head (Rmin) are expressed as a fraction (Rmin-SA)/(Rmax-SA) of the responses to the most effective view (Rmax). SA = spontaneous activity. Upper: Relative response of least effective view for cells displaying object-centred properties. Lower: Relative response of least effective view for cells displaying viewer-centred properties

selective for heads excluding anomalies (see above) in Fig. 5. Negative values here are notable because they indicate that control stimuli inhibited activity relative to the spontaneous firing rate. 72% of cells selectively responsive to the head are five times as responsive to the optimal head view as to control stimuli.

Discrimination between views. For the 110 viewer-centred cells the average response to the least effective view was 13.0 ± 1.31 spikes per second (mean $\pm$1SE, $n=110$). Whereas the response to the optimal view of the head was 46.0 ± 2.29 spikes per second. To provide an index of discrimination between different views of the head, cell responses to the least effective view were compared to the responses to the most effective view (in a similar manner to that described for control stimuli). Discrimination

between views was computed for each cell using the formula, (Rmin-SA)/(Rmax-SA), where Rmin = mean response to the least effective view. Fig. 6 displays the distribution of responses to the least effective view, defined in this relative way. Negative values again indicate that the least effective view reduced activity below the spontaneous firing rate. For the majority of viewer-centred cells response to the least effective view was less than 0.5 of the response of the most effective view.

Not surprisingly cells with object-centred properties (i.e. responding to all views of the head more than to controls and spontaneous) showed less discrimination between views of the head than cells displaying viewer-centred responses. Consequently for object-centred cells the response to the least effective view ranged between 0.4–0.9 of the response to the most effective view.

View tuning

Width of tuning of viewer-centred cells. Width of tuning was calculated as the average angle required to reduce firing rate to half of the difference between response to the most and least effective views [(Rmax-Rmin)/2 or ½ width at ½ height measure]. Width of tuning estimated with this measure for viewer-centred cells tested with 8 views is illustrated in Fig. 7. Half width at half height ranged from 50° to 125°.

The distribution of tuning in Fig. 7 illustrates two important points, first the majority of cells have tuning less than 60° (½ width) and second, the distribution appears bimodal with a distinct population of cells exhibiting tuning greater than 90° (½ width). Since the distribution of width of tuning indicated two types of cell, the population was split for further analysis. Using 90 degree ½ width tuning as a cut off point 54 cells were defined here as having relatively 'narrow' tuning and 19 cells as having 'broad' or asymmetric tuning (with one or both ½ width measures were greater than 90°). [For each cell there being two ½ width at ½ height measures.]

The distribution of width of tuning is skew positive with no cells having ½ width $<45°$. This is in part an artifact of regression analysis since the cardioid equation used cannot follow changes in response from maximum to minimum in less than 90°. Therefore for narrowly tuned cells, the width of tuning is artificially broad (by an estimated 5–15 degrees). This error affected tuning estimates for a minority of cells (10 out of 110 cells with tuning between 20 and 50 degrees ½ width at ½ height) and therefore does not affect estimates of the width of tuning of the cell population unduly.

Average shape of view tuning. To make a visual comparison across different tuning curves for those cells where regression analysis produced a good fit to observed data, the following analysis was performed: For each cell the regression analysis was used to estimate the maximal response (Rmax) and the optimal angle of view (θmax) at which Rmax occurred. The raw data for each cell was then scaled so that Rmax was equal to 1.0 and the mean spontaneous activity was equal to 0.0 spikes/second. For each cell views used to measure responses were re-expressed as angles of rotation from θmax. Regression analysis was then performed on the transformed data for each cell to produce 'normalized' tuning curves. This procedure aligns the peak of each tuning curve (e.g. Fig. 1) along the horizontal axis and stretches the vertical (response) axis so that the height of the peaks above spontaneous baseline is the same. Figure 8 displays the individual tuning curves for a sample of 43 unimodal viewer-centred cells with narrow tuning.

To obtain the average tuning curve for different cells the coefficients of the regression analysis (equation 1, above) of normalized data were averaged. Figure 9 displays the average tuning curves for different categories of view tuning. The average tuning curves for all classes except broadly tuned cells exhibit a dip in response to views 90° away from optimal view. This dip may well arise from inhibition from cells tuned to these orthogonal views. Evidence for inhibition comes from the observation that for 46 cells response to non-optimal views was less than spontaneous activity. (For such cells the relative

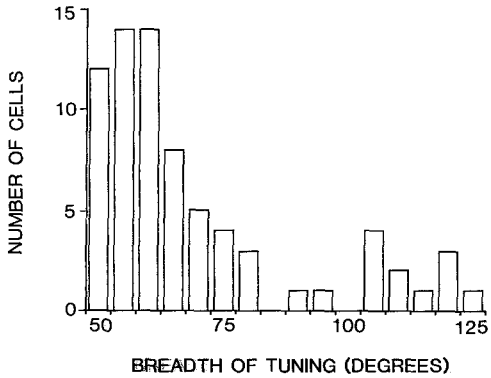

Fig. 7. Width of tuning. The average angle of rotation required to reduce response by half of difference between response to the most and least effective views {Rmax-Rmin)/2} is plotted for 73 view-selective cells.

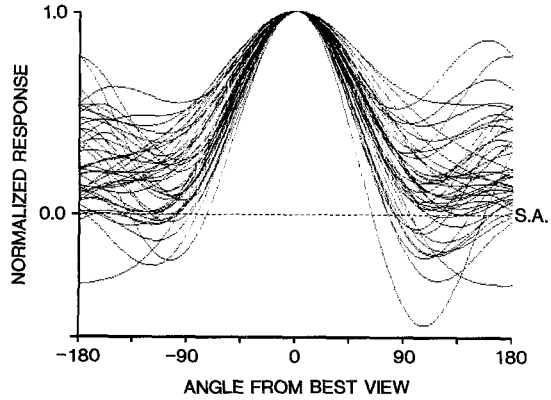

Fig. 8. Tuning curves of viewer-centred cells displaying unimodal narrow tuning. The tuning curves (estimated from best fit cardioid function relating response to angle of view) for 43 unimodal view-selective cells. Each tuning curve is normalized so that maximum response = 1.0 and spontaneous activity (S.A.) = 0. Angle of view is expressed as an angle of rotation from optimal view (θmax)

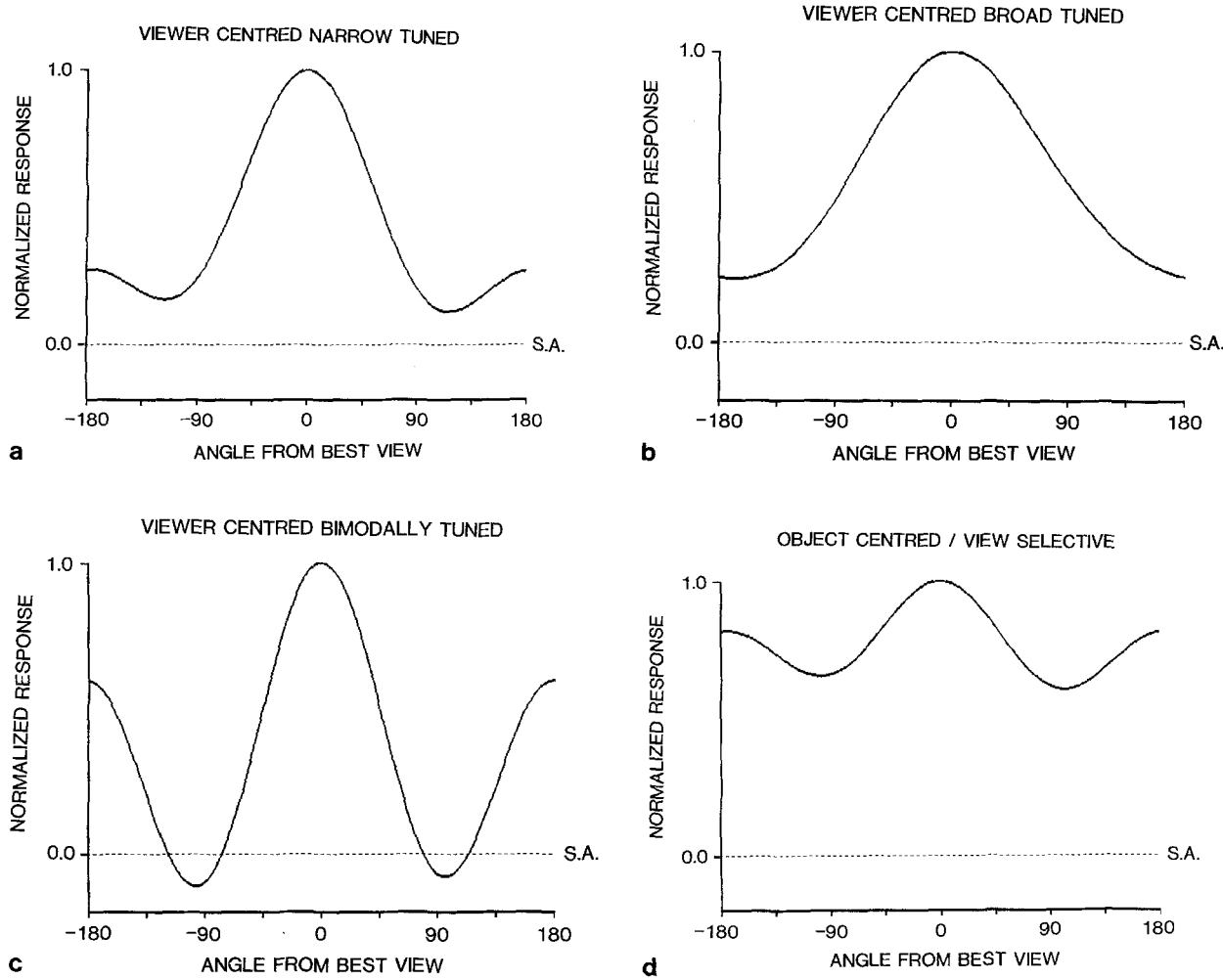

Fig. 9. Average tuning curves for different classes of cell. **a** Narrow band viewer-centred cells ($n=43$). **b** Broad band viewer-centred cell ($n=11$). **c** Bimodal viewer-centred cells ($n=8$). **d** Mixed object and viewer-centred cells ($n=3$)

response to the worst view in Fig. 6b assumes negative values.)

Distribution of optimal response angles. The optimal response angles were analysed for cells which (a) were tested with 8 views of the head, (b) displayed a significant ($p<0.05$) relation between response and a cardioid function of viewing angle (equation 1) and (c) for which Chi-Squared comparisons between predicted and observed response indicated a good fit. Thus data were considered for only those cells for which regression analyses produced appropriate optimal response angles.

Further checks were made for the accuracy of the estimated optimal angle of view for cells with broad tuning. These cells were defined objectively as having an angle of rotation required to reduce response to ½ maximal that was greater than 90°. For nine such cells with broad view tuning, response was large and even to all views except one. The optimal view for these 9 cells was calculated as the view producing minimum response +180°. This method of estimating optimal view was introduced because the cardioid function correctly fits a single narrow peak (or trough) but introduces 1 to 2

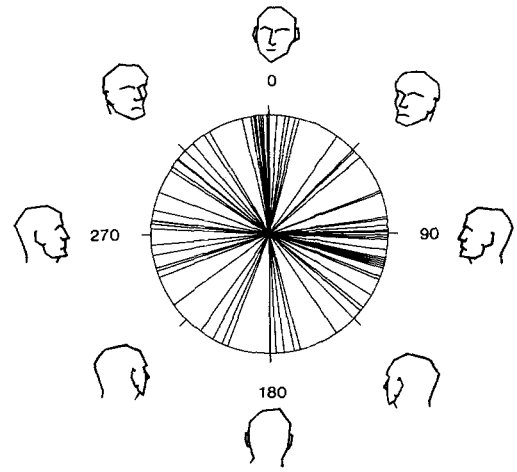

Fig. 10. The distribution of view tuning across the population of viewer-centred cells. Each line represents the view estimated from regression analysis to evoke maximal response for one cell. Significantly more cells exhibit a preference for views within 22.5° of putative characteristic views (Face, left and right profiles and the back of head) than for intermediate views (*Binomial/Test*, $p=0.0002$)

Table 1. Number of cells tuned to characteristic views of the head. The distribution of view tuning for cells in the STS is given separately for four experimental subjects. On axis tuning refers to cells with an optimal view within 22.5° of one of the four putative characteristic views (face, left profile, right profile and back views of the head or 0°, 90°, 270°, 180°). Off axis refers to cells tuned to other views

Subject	On	Off	Number tested
J	5	1	6
H	6	2	8
D	34	8	42
B	7	10	17
Totals	52	21	73

small ripples with 'false' maxima in approximating very broad/flat peaks.

Figure 10 shows the distribution of the optimal response angles of 73 cells from 4 monkeys B, D, H and J. Each cell is represented by a single line. To assess potential clustering around characteristic views, cells were divided into two categories: those with angles of optimal response 'on axis' within 22.5° either side of one of the characteristic views and those with optimal angles 'off axis' outside this range. For 73 cells tuned for view 52 were found to be 'on axis'. This fraction is significantly greater than that expected by chance (Binomial Test $p=0.0002$). Of the 73 cells 54 had narrow tuning for perspective view (of which 38 were on axis) and 19 had broad tuning (of which 14 were on axis).

Previous assessments of the distribution of view tuning across cells (Perrett et al. 1989b, 1990d) were made using smaller numbers of cells and smaller numbers of views for some of the testing (4 rather than 8 used here for all cells illustrated).

The number of cells recorded in different subjects with on or off axis responses is given in Table 1. There is an indication of individual differences or sampling of different cell populations across subjects. The majority of cells were on axis for 3 subjects (D, H and J) but not for the 4th subject B. The fraction of cells exhibiting on axis tuning in subjects D and B is significantly different (Chi-Squared = 7.25, df = 1, $p<0.05$).

Sensitivity to identity

Sensitivity to identity was not the main focus but such sensitivity is relevant to the issue of cell classification and was studied additionally when cells were found by serendipity to respond differently to different individuals. 56 cells were tested with more than one head. Using the same mode of presentation (e.g. all stimuli 3D or all stimuli 2D projected images), 6 of these cells showed significant differences in response magnitude to different individuals as assessed by 1- or 2-way ANOVA. Cells with sensitivity to identity were found in categories defined here as both viewer-centred and object-centred.

Viewer-centred coding of identity. Two cells were studied where view selectivity was found for one individual in the

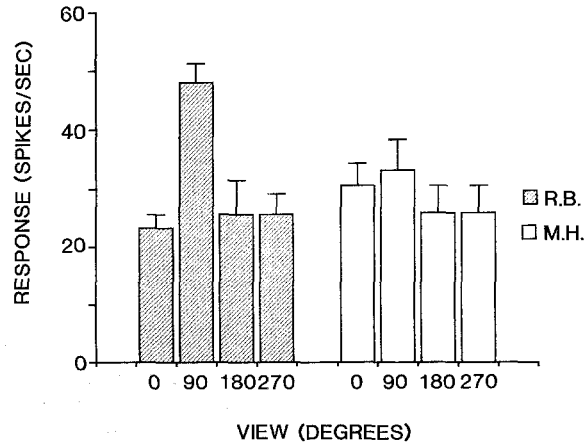

Fig. 11. Combined selectivity for view and identity. Cell D043 29.44 gave significantly larger responses ($p<0.05$ each comparison) to the left profile (90°) of RB than to other views of RB, MH and controls (23 spikes/s). No view of MH evoked higher responses than controls ($p>0.05$ each comparison). 1-way ANOVA: $F(8,34)=3.2$, $p<0.01$. A 2-way ANOVA showed a significant effect of view [$F(3,32)=5.2$, $p=0.005$]. The main effect of identity, however, did not reach significance [$F(1,32)=0.22$, $p=0.65$] but a significant interaction between view and identity ($F=2.93$, DF = 3,32, $p=0.049$) confirmed that the pattern of response to the views was different for the two individuals

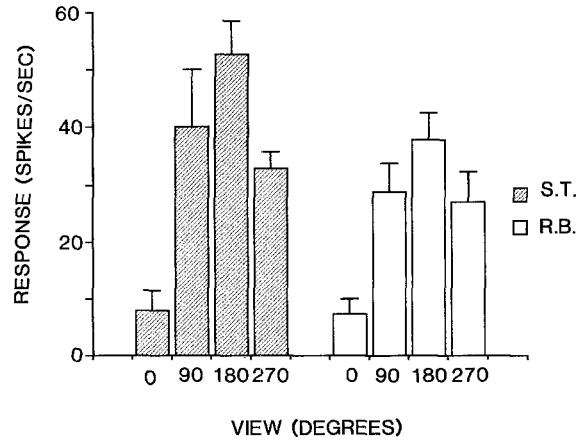

Fig. 12. Responses of a broadly tuned viewer-centred cell sensitive to identity. For cell D107 35.41 the back (180) views of the heads of two experimenters (ST and RB) evoked higher responses than the front (0) views ($p<0.001$). Responses to the back of ST's head were higher than responses to the back view of RB, controls (1.6 spikes/s) and spontaneous activity (6.4 spikes/s) ($p<0.05$ each comparison). 1-way ANOVA: $F(9,55)=14.2$, $p<0.001$. A 2-way ANOVA showed a significant effects of view [$F(3,43)=15.9$, $p=0.001$) and identity [$F(1,43)=6.6$, $p=0.037$]. There was no significant interaction [$F(3,43)=0.0$, $p=1.0$], showing that the pattern of view selectivity was the same for both individuals

absence of any response (or view tuning) to a second individual. An example is illustrated in Fig. 11. For the cell illustrated responses to one experimenter (RB) typified a viewer-centred cell with narrow tuning for perspective view. 1-way ANOVA (for F values see Fig. legends) indicated that just one of the four views of RB was significantly ($p=0.05$) different from control stimuli and spontaneous activity. The cell, however, failed to distin-

guish any view of the head of a second experimenter (MH) from control objects and spontaneous activity.

A 2-way ANOVA (comparing 4 views and 2 identities) showed a significant effect of view ($p=0.005$). The main effect of identity, however, did not reach significance ($p=0.65$), though there was a significant interaction between view and identity ($p=0.049$) which confirmed that the pattern of response to the views was different for the two individuals.

Figure 12 illustrates the responses of a cell with broad tuning for perspective view. While the cell showed the same pattern of view selectivity for two experimenters (ST and RB), in the optimal view (back of the head or 180°) responses to ST were significantly higher than those to RB ($p<0.05$).

2-way ANOVA (with view and identity as main factors) showed a significant effect of view ($p=0.001$), identity ($p=0.037$), but no interaction between view and identity ($p=1.0$). The pattern of view selectivity was thus the same for both individuals.

Object-centred coding of identity. A dramatic example of a cell sensitive to identity independent of view is shown in Fig. 13. 1-way ANOVA revealed that the cell displayed object-centred properties. Firstly, responses to four views of one experimenter (JH) were all significantly greater than response to controls and spontaneous activity and secondly responses to the different views of JH were not significantly different. The cell was, however, unresponsive to a second experimenter (DP) with no view evoking responses higher than controls or spontaneous activity. 2-way ANOVA (view versus identity) showed significant effect of identity ($p<0.001$) but no effect of view ($p=0.9$) and no interaction ($p=0.5$).

The visual basis of this discrimination was not determined but appeared to rely on cues from the head and the upper torso, since no responses were present to the head of JH when the body was obscured from sight (or to the body when the head was covered). This indicates that the sensitivity for identity was unlikely to have arisen from any simple visual cue such as hair style.

View preference independent of identity. Cells insensitive to identity were important to the present study since they allowed assessment to be made of the consistency of view tuning across individuals. For 12 cells regression analyses gave significant relationships between response and view for the heads of two to five individuals (each tested separately and with 8 views). This testing allowed 22 comparisons to be made between the estimated optimal angles of view for responses of one cell to the head of two different individuals. The optimal angles of view for any two assessments with the same cell were highly correlated (correlation coefficient $= +0.973$, $df=20$, $p<0.00005$). A matched paired t test between the optimal views revealed no systematic difference between test and re-test (mean signed difference between test and re-test $=2.3\pm3.9°$ (mean ±1 SE); $t=0.59$, $df=21$, $p=0.57$). The mean absolute difference between any two assessments of optimal view for the same cell was 11.5° (±3.1). The results indicated a high degree of consistency of the cells' view preferences over time and confirmed the test re-test reliability of the method of analysis.

Figure 14 illustrates the consistency of view tuning across individuals and across presentation media. Responses to views of one experimenter (MH) were assessed with real 3D images while responses to views of a second experimenter (MO) were assessed with 2D projected slides. Figure 14 illustrates the best fit regression equation relating view to response for the two heads. For both

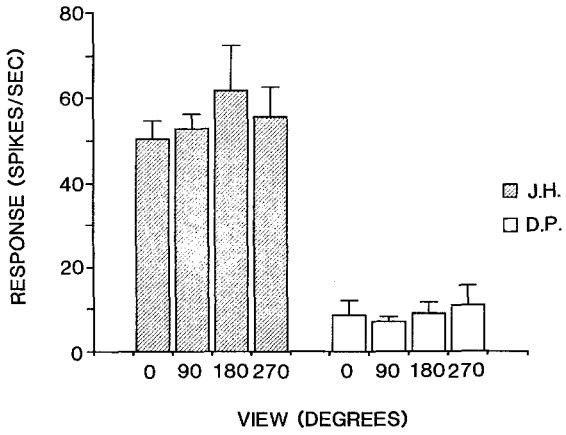

Fig. 13. Responses of an object-centred cell sensitive to identity. Responses of cell D105 29.74 to the 4 views of JH were not significantly different ($p>0.05$ each comparison) but responses to each of the views of JH was significantly greater than responses to the views of DP, control stimuli (16.0 spikes/s) and spontaneous activity (13.6 spikes/s) ($p<0.0005$ each comparison). [1-way ANOVA, $F(9,57)=17.4$, $p<0.0005$]. A 2-way ANOVA (with view and identity as main factors) showed a significant effect of identity [$F(1,47)=112.7$, $p<0.001$) but no effect of view [$F(3,47)=0.12$, $p=0.95$] and no interaction [$F(3,47)=0.75$, $p=0.53$]

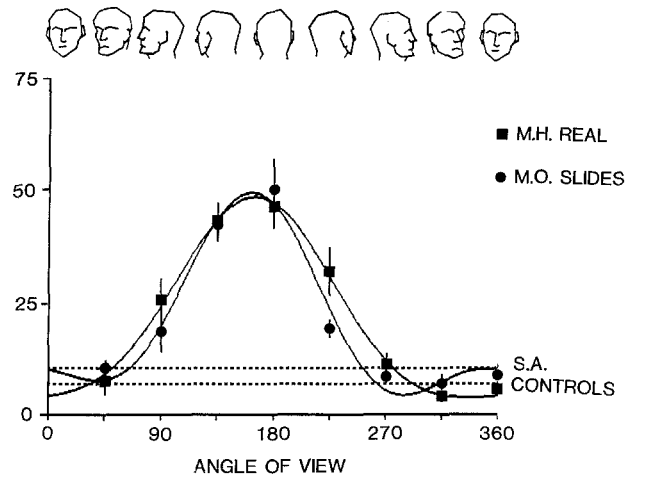

Fig. 14. Consistency of view independent of identity and presentation media. Responses of one cell (D114 29.83) to 2D slides of one individual (MO) and 3D real views of a second individual (MH). For both individuals the cell showed a preference for the back view (regression analysis MH: $R^2=0.83$, $F(4,35)=9.3$, $p<0.0005$; MO: $R^2=0.80$, $F(4,35)=35.2$, $p<0.0005$). A 2-way ANOVA showed a significant effect of view [$F(7,64)=42.3$, $p<0.0005$], but no effect of identity or media (2 or 3D) of presentation [$F(1,7)=0.6$, $p=0.46$] or interaction between view and identity [$F(7,64)=1.3$, $p=0.25$]

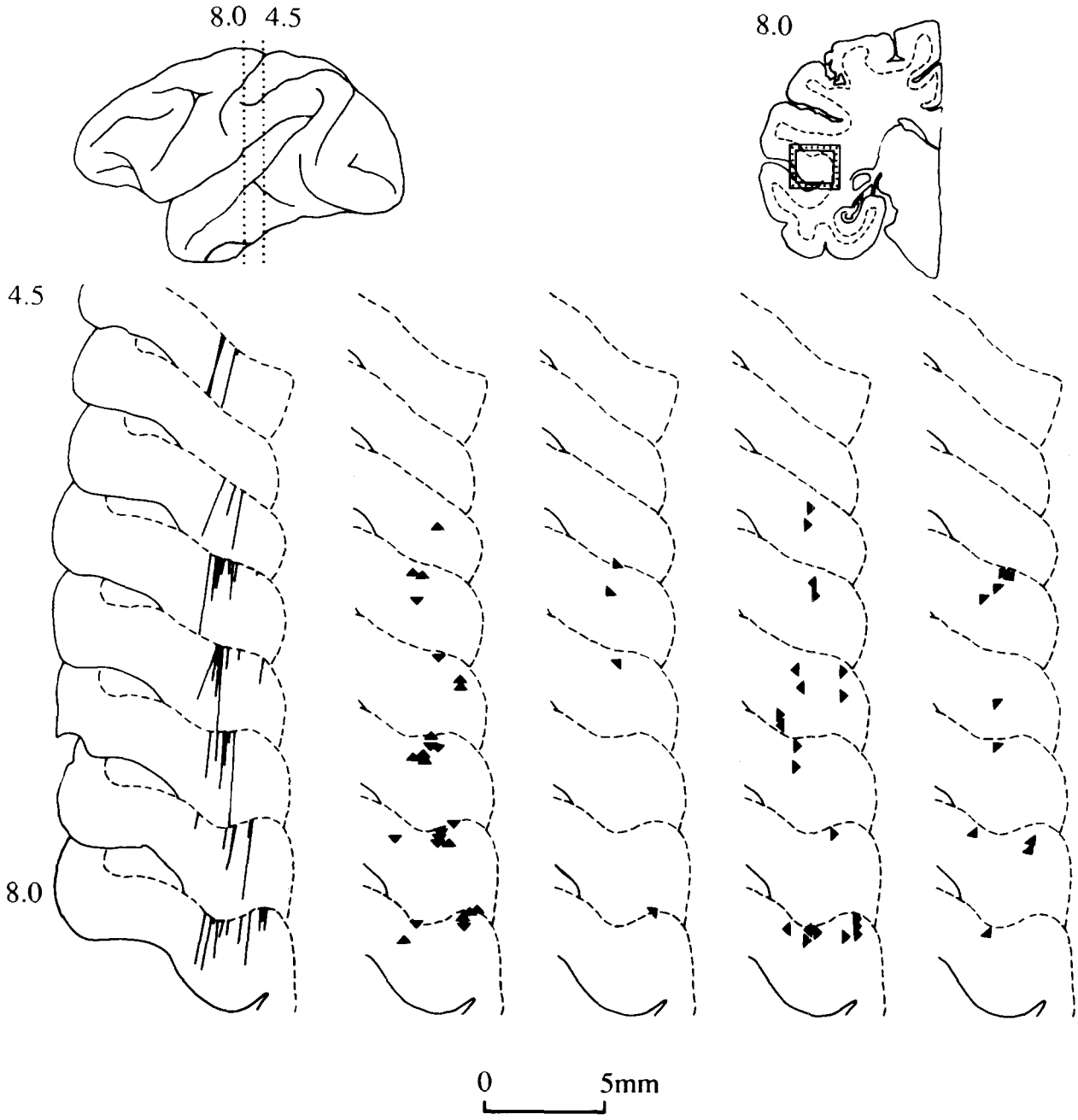

Fig. 15. Histological reconstruction of position of cells selectively responsive to views of the head. *Upper:* Side view and coronal section at 8.0 mm anterior to the inter aural plane showing the position of the superior temporal sulcus (STS). *Lower:* serial sections of the upper bank STS from 4.5 mm to 8.0 mm anterior to the interaural plane from one monkey (B). Left column indicates the position of all recording tracks. Other columns indicate the position of cells responsive to different views. The preferred angle of view for each cell (to the nearest 23°) is indicated by the direction of the 8 arrow types. up = face (0°), down = back of head (180°), right = right profile (90°), left = left profile (270°) diagonal arrows = intermediate views (45°, 135°, 225°, 315°)

heads the cell displays a preference for the back view. A 2-way ANOVA of the responses to the 8 views of the 2 heads showed a significant effect of view ($p < 0.0005$) but no effect of identity or media of presentation ($p = 0.46$). There was no significant interaction ($p = 0.25$), showing that the pattern of view selectivity was the same for both individuals. The lack of difference across the two heads also indicates that the cell responded equivalently to two and three dimensional images.

Location of cells

Histological reconstruction of the positions of cells recorded in monkeys F, B, D indicated that the majority of cells responsive to faces and other views of the head were located in the cortex of the upper bank of the superior temporal sulcus (areas TPO and PGa of Seltzer and Pandya 1978). The proportions of cells found responsive to the head varied from subject to subject (65 out of 524

cells recorded in the STS upper bank in the right hemisphere of B, 45/914 for D right hemisphere, 9/483 D left hemisphere and 15/1553 F right hemisphere). [N.B. these figures include cells responsive to the head that were not investigated for view sensitivity.] Measurements of the position of recording electrodes (from X-radiographs) indicated cells responsive to faces in monkeys J and H were recorded in the temporal cortex (mainly in the upper bank of the STS but also in the lower bank and inferior temporal cortex).

Figure 15 displays cells selectively responsive to head views that were recorded in the upper bank of the STS in the right hemisphere of one monkey. With the resolution of reconstruction present (± 1.0 mm) there was no obvious anatomical organization of view coding within the cortex of this monkey. All views in the horizontal plane appeared to be coded in the same patch of cortex.

Discussion

Viewer-centred coding

Previous work with cells responsive to the head focused mainly on coding of the face or frontal views. The current study evaluated coding of the head throughout the horizontal plane. One of the clearest findings of the present study was the prevalence of view specific or viewer-centred coding. Cells responsive to the head in the STS were most frequently sensitive to perspective view. 96% (129/134) of the cells in the present study exhibited sensitivity to view in the horizontal plane.

Previous work also emphasises the prevalence of view sensitive coding. Perrett et al. (1985a) found 67% of cells responsive to the face were view selective. In the study of Hasselmo et al. (1989a), 16/19 of cells insensitive to identity and 15/18 of those sensitive to identity, were selective for view. View sensitivity was also evident in earlier reports (Bruce et al. 1981; Desimone et al. 1984; Perrett et al. 1982; Kendrick and Baldwin 1987).

The data from different studies of the processing of static information about the head are thus comparable and indicate that coding is critically dependent on the view-point of the observing subject.

Object-centred coding

The present study provided evidence for some cells in the temporal cortex coding in an object-centred manner. The most important property of object-centred coding is the ability to generalize response across different perspective views. This capacity was demonstrated here at the level of object class in that some cells responded to all views of heads (in the horizontal plane) but not to other objects (e.g. Figs 3, 4). The capacity was also found in the coding of identity in that cells were found selective to all views of one individual's head but were unresponsive to all views of a different individual (e.g. Fig. 13). Such object-centred coding of identity has been reported in other studies (e.g. Hasselmo et al. 1989a; Perrett et al. 1984, 1987, 1989a).

Despite the undisputed existence of object-centred or view invariant coding, we found such coding rare in the STS. Very few cells ($n=9$) responded to all views of the head and even less ($n=4$) responded equally to all views. We found that half of the cells responding to all views of the head in the horizontal plane showed significant differences in their response to different views of the same head.

We therefore diverge from Hasselmo et al. (1989a) in the stress placed on the relative importance of viewer and object-centred coding for STS cells responsive to static heads. We emphasize that since the majority of cells are selective for perspective view they should be considered examples of viewer-centred coding.

It could be argued that the degree of generalization across perspective view depends on the supposed function of the cells. If one considers the STS cells as part of a *face* recognition system rather than part of a system analyzing the *head* (and other parts of the body), then one might not expect complete generalization to the back views of the head. A cell might be considered object-centred in its coding of the face if it responded equally to the front and side views of the head. This would lead to the prediction of broad tuning for cells responsive to the frontal views. Overall the width of tuning for perspective view, however, was *not* found to vary with optimal view ($r=0.034$, $df=64$, $p=0.784$). Indeed the only exception to this was the observation of a very few ($n=4$) cells with the sharpest tuning (½width at ½height $<30°$). The tuning of these cells was associated with and was symmetrically centred on the full face view. Thus most cells responding well to the face view, responded significantly less to the profile (or vice versa). Therefore coding lacked generalization across perspective view even when analysis is restricted to views in which the facial features were visible.

As amplified in the introduction there have been demonstrations of some cells exhibiting object-centred coding for body movements (Perrett et al. 1985b, 1990a; Hasselmo et al. 1989a). Again we would stress that the responses of the majority (98%, Harries et al. in prep; see also Perrett et al. 1985b) of STS cells coding body movements are dependent on the perspective view as shown in the present study for the coding of static heads and bodies.

The results of the present study thus emphasise the importance of viewer-centred coding of heads. Of course the present results do not deny that object-centred coding may be more prevalent in other regions of the brain (perhaps at a processing stage subsequent to the STS).

Range of views preferred by cells

The work of Hasselmo et al. (1989a) revealed cells tuned to a range of views. They noted, however, that the face or front of the head was disproportionately represented. In their Fig. 11, it appears that 29/49 (or 59%) of the displayed lines marking the preferred viewing angles (at which different neurons responded optimally), occurred to front views of the head (within 45° of the frontal view).

[Hasselmo et al. studied 31 cells in total but presumably displayed 2 preferred viewing angles for each of the 18 cells responding differently to the 2 heads tested, $31 + 18 = 49$]. Hasselmo et al. note that the bias for front views could reflect the fact that the front view contains most information and is most important socially.

The present study does not show such a bias towards the full face view. We found that cells were clumped around the face and the profile views. The differences in the two studies could represent the results of sampling differences. Perrett et al. (1984), showed that cells responsive to faces are clumped in distribution. Most frequently cells processing different views of the head are found in the same clump or patch of cortex (Perrett et al. 1984, 1985a, and here). Clumps of cells can occasionally be found, however, where processing of one view predominates. For example Perrett et al. (1987, 1988) reported separate clumps processing face and profile views. Studies sampling from clumps containing multiple views will produce evidence of even distribution of view tuning while studies sampling from clumps processing the face will emphasise the importance of the front view of the head.

A further potential source of bias concerns methods of estimating optimal tuning angle. The method used by Hasselmo et al. (1989a) is not appropriate for bimodal cells. To illustrate this, consider a bimodal cell with a large but equal response to the left and right profile views, a small response to the face and no response to the back of the head (e.g. Fig. 10b, Hasselmo et al. 1989a). The method of determining the optimal view used by Hasselmo et al. (1989a) would wrongly ascribe the orientation preference to the front view. This is because the influence of the two profiles effectively cancel out and the front/back difference becomes the major determinant of the estimated orientation specificity. By contrast the cardioid equation used here would indicate 2 modal angles close to the profile views. Thus methods of analysis can influence the estimation of preferred views particularly in bimodal cells.

Results of previous studies may also be biased towards finding front view preferences because more front views were tested than back views. Ideally testing a larger number of views than the 8 used here would present a more accurate specification of view tuning. Indeed views changing the vertical elevation of the head have been found to effect some cell responses yet were not studied here (Perrett et al. 1985).

Characteristic views

In the present study, the optimal angles of view for the entire population were not evenly distributed through 360° of rotation in the horizontal plane. The overall results showed statistically significant clustering around putative characteristic views (the face, left and right profile and back). These data provide empirical support for the notion that particular views of an object are differentially represented in the nervous system. It is relevant that the physiological evidence for the importance of the face and profile views presented here, parallels the importance of the same views in psychological studies of preferential inspection (Harries et al. 1990). Indeed both studies indicate preferential coding of the face and profile views and only a small degree of coding for the back view.

Physiological data from different experimental subjects suggested individual differences. While data for three subjects were consistent with the notion of characteristic views, data from a fourth subject did not. These individual differences most likely reflect limitations in sampling. It is also possible that subregions of cortex involved in the analysis of heads, may be specialized for processing particular views (as suggested by previous studies, Perrett et al. 1987). Neurophysiological studies can only explore a minority of cells and regions of the cortex. Thus some differences in the results from different laboratories and experimental subjects may be inevitable.

Function in coding attentional direction

The present study emphasises the importance of the visual appearance of the entire head. Cells responsive to the front view of the head have been found to be sensitive to socially important information about facial expression (Perrett et al. 1984; Hasselmo et al. 1989b; Perrett and Mistlin 1990). Expression is not visible from the back of the head yet, as argued below, even this view may have significance in a social context.

Given the width of tuning for perspective view of cells studied here (range 50–125° ½ width at ½ height), recognition of an object as a head or as an individual could be accomplished by cell populations coding just 4 evenly spaced views (e.g. the 4 proposed characteristic views). It is clear, however, that there are many cells analysing views in-between the putative characteristic views. More views receive coding at the single cell level than are theoretically necessary for recognition independent of view.

One clue as to the function of the 'supernumerary' coding of view comes from the observation that most cells are not affected by the identity of the face or its species (monkey/human). It is therefore unlikely that these cells are involved in the recognition of identity or species. By contrast the majority of cells responsive to the head are selective for the direction in which the head points. It is more reasonable to speculate that cells selective for head view have some function in coding *the direction in which other individuals point*. This may itself provide an index of where the attention of others lies.

Realizing the direction of others' attention is vital for many aspects of social life. It is not only important to know that X is threatening, it is also important to know whether the threat is directed at oneself, one's ally, one's kin, etc (Chance 1967).

Human infants, exhibit a capacity to react to where in the environment another individual is attending using the cue of gaze direction (Butterworth and Cochran 1980). This capacity appears early in life and is perhaps a basis for the development of more complex under-

standing of the attention and even the 'minds' of others, both for humans and monkeys (Byrne and Whiten 1987; Baron-Cohen 1989).

If the view coding has a role in the analysis of where other individuals are attending, then one would expect gaze direction to influence responses, since gaze direction is a more accurate guide to the direction of attention. This suggestion is borne out. Perrett et al. (1985a) found that 64% of cells responsive to the face or profile views of the head were also selective for the direction of gaze. Optimal gaze direction was found to be compatible with optimal head direction; cells responsive to the head directed towards the observer (face view) were more responsive to eye contact than laterally averted gaze; cells responding to the profile responded more to gaze directed laterally. For several cells gaze direction was found to be more important than head view. (For more detailed consideration see Perrett et al. 1990c).

Deriving view-invariant descriptions

Stressing the importance of viewer-centred coding in its own right does not deny the possibility that such coding could also be used as in intermediate stage in establishing view independent recognition (with object- or goal-centred frames of reference). Cells responsive to many views of an object or action could be established by combining the outputs of several view sensitive cells tuned to different views of the same object or action. Indeed such a scheme of processing has been suggested many times on the basis of physiological evidence (Perrett et al. 1984, 1985a, b, 1987, 1989a, 1990a, b; Hasselmo et al. 1989a). An analysis of the time course of responses supports this contention, since view-selective cells respond at a slightly earlier latency compared to cells with view invariant responses (Oram and Perrett, unpublished studies).

Width of tuning

Receptive fields of cells in the early stages of the visual system are characterised by a central excitatory region flanked by an inhibitory surround. The response profile thus has a familiar 'Mexican hat' like structure. The inhibitory side bands arise from antagonistic influences from cells with adjacent receptive fields. It is interesting to note the 'Mexican hat' shape of view tuning amongst the majority of cells responsive to the sight of the head. The dip in response, for perspective views 90° to the cells' preferred views, may represent inhibitory interactions between cells tuned to different views. Indeed inhibition relative to spontaneous activity was observed for many cells to *non-optimal* views.

In keeping with the relatively low rate of spontaneous activity, *optimal* views appeared to be coded by excitation rather than inhibition. A cell could theoretically code the presence of one head view by a selective reduction of response rate below spontaneous activity for one view, with no change in activity for other views. No such cells were found.

Identity sensitivity

Examples of selectivity for identity were found here for cells displaying viewer-centred and object-centred coding. The responses of the cell illustrated in Fig. 11 indicates that sensitivity to the difference between individuals can be restricted to one view. Such view specific sensitivity to identity has been noted in other reports (Perrett et al. 1984, 1987, 1989a).

Hasselmo et al. (1989a) emphasised that cellular sensitivity to identity was largely independent of view and was consistent with object-centred coding. They based this claim on the fact that 18 cells exhibited a significant main effect of identity in 2-way ANOVA (with identity and view as main factors). It is important to note that reliance on main effects of identity in 2-way ANOVA overlooks the possibility of selectivity for identity occurring at specific views. Main effects also overlook situations where responses to two individuals may be different at some views but not at other views. The interaction term of 2-way ANOVAs may confirm such combined sensitivity to view and identity (see Fig. 11).

If it is proposed that cells exhibiting sensitivity to identity but generalizing across many views are built by combining view specific descriptions, then it is not surprising that sensitivity to identity should occur amongst view specific cells.

Acknowledgements. This research was funded by grants from the S.E.R.C. Image Interpretation Initiative (GR/E 43881) and linked earmarked PhD studentship (used to support R.B.), the Japanese New Energy and Industrial Technology Development Organization (N.E.D.O.) and the E.S.R.C. (XC15250005). D.I.P was supported by a Royal Society University Research Fellowship and S.T. by a Wellcome Trust Prize Studentship. We are grateful to Dr P Jupp of the Department of Mathematics University of St Andrews for devising the method of regression analysis. We thank Dr W Dittrich, H Carus and V Walsh for their helpful comments on drafts of the manuscript.

References

Baron RJ (1981) Mechanisms of human facial recognition. Int J Man-Machine Studies 15:137–178

Baron-Cohen S (1989) Perceptual role taking and protodeclarative pointing in autism. Br J Devel Psychol 7:113–127

Baylis GC, Rolls ET, Leonard CM (1985) Selectivity between faces in the responses of a population of neurons in the cortex of the superior temporal sulcus of the macaque monkey. Brain Res 342:91–102

Bruce CJ, Desimone R, Gross CG (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46:369–384

Bruce V, Valentine T, Baddeley A (1987) The basis of the ¾ view advantage in face recognition. Appl Cogn Psychol 1:109–120

Butterworth GE, Cochran EC (1980) Towards a mechanism of joint attention in human infancy. Int J Behav Devel 3:253–272

Byrne RW, Whiten A (1987) A thinking primate's guide to deception. New Scientist 116:54–57

Chance MAR (1967) Attention structure as the basis of primate rank orders. Man 2:503–518

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 8:2051–2062

Feldman JA (1989) Four frames suffice: a provisional model of vision and space. Beh Brain Sci 8:265–289

Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the monkey. J Neurophysiol 35:96–111

Harries MH, Perrett DI, Lavender A (1990) Visual inspection during encoding and recognition of 3D heads. Perception (in press)

Hasselmo ME, Rolls ET, Baylis GC, Nalwa V (1989a) Object-centred encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. Exp Brain Res 75:417–429

Hasselmo ME, Rolls ET, Baylis GC (1989b) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Beh Brain Res 32:203–218

Hinton D, Parsons LM (1988) Scene-based an viewer-centred representations for comparing shapes. Cognition 30:1–35

Kendrick KM, Baldwin BA (1987) Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces. Science 236:448–450

Koenderink JJ, van Doorn AJ (1976) The singularities of the visual mapping. Biol Cybem 24:51–59

Koenderink JJ, van Doorn AJ (1979) The internal representation of solid shape with respect to vision. Biol Cybern 32:211–216

Leonard CM, Rolls ET, Wilson FAW, Baylis GC (1985) Neurons in the amygdala of the monkey with responses selective for faces. Behav Brain Res 15:159–176

Logie RH, Baddeley AD, Woodhead MM (1987) Face recognition: Pose and ecological validity. Appl Cogn Psychol 1:53–69

Lowe D (1987) Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence 31:355–395

Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. W H Freeman, San Francisco

Marr D, Nishihara HK (1978) Representation and recognition of the spatial organization of 3-dimensional shapes. Proc R Soc Lond B 200:269–294

Merrill EG, Ainsworth A (1972) Glass-coated platinum-plated tungsten microelectrodes. Med Biol Eng 10:662–672

Perrett DI, Harries MH (1989) Characteristic views and the visual inspection of simple faceted and smooth objects: "tetrahedra and potatoes". Perception 17:703–720

Perrett DI, Rolls ET, Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329–342

Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1984) Neurons responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. Hum Neurobiol 3:197–208

Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1985a) Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc R Soc Lond B 223:293–317

Perrett DI, Smith PAJ, Mistlin AJ, Chitty AJ, Head AS, Potter DD, Broennimann R, Milner AD, Jeeves MA (1985b) Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: a preliminary report. Beh Brain Res 16:153–170

Perrett DI, Mistlin AJ, Chitty AJ (1987) Visual cells responsive to faces. Trends Neurosci 10:358–364

Perrett DI, Mistlin AJ, Chitty AJ, Smith PAJ, Potter DD, Broennimann R, Harries MH (1988) Specialized face processing and hemispheric asymmetry in man and monkey: evidence from single unit and reaction time studies. Beh Brain Res 29:245–258

Perrett DI, Harries MH, Bevan R, Thomas S, Benson PJ, Mistlin AJ, Chitty AJ, Hietanen JK, Ortega JE (1989a) Frameworks of analysis for the neural representation of animate objects and actions. J Exp Biol 146:87–114

Perrett DI, Mistlin AJ, Harries MH (1989b) Seeing faces: The representation of facial information in temporal cortex. In: JJ Kulikowski, CM Dickinson, IJ Murray (eds) Seeing contour and colour. Permagon Press, Oxford, pp 770–754

Perrett DI, Harries MH, Benson PJ, Chitty AJ, Mistlin AJ (1990a) Retrieval of structure from rigid and biological motion; an analysis of the visual response of neurons in the macaque temporal cortex. In: T Troscianko, A Blake (eds) AI and the eye. J Wiley, Chichester, pp 181–201

Perrett DI, Harries MH, Chitty AJ, Mistlin AJ (1990b) Three stages in the classification of body movements by visual neurons. In: HB Barlow, C Blakemore, M Weston-Smith (eds) Images and understanding. Cambridge University Press, Cambridge, pp 94–107

Perrett DI, Harries MH, Mistlin AJ, Hietanen JK, Benson PJ, Bevan R, Thomas S, Ortega J, Oram MW, Brierly K (1990c) Social signals analysed at the single cell level: someone's looking at me, something touched me, something moved. Int J Comp Psychol (in press)

Perrett DI, Harries MH, Mistlin AJ, Chitty AJ (1990d) Recognizing objects and actions: frameworks for neuronal computation and perceptual experience. In: DM Guthrie (ed) Higher order sensory processing. Studies in neuroscience series. Manchester University Press, Manchester, pp 155–173

Perrett DI, Mistlin AJ (1990) Perception of facial characteristics by monkeys. In: WC Stebbins and MA Berkley (eds), Comparative perception, Vol II. Complex signals. John Wiley, New York pp 187–215

Poggio T, Edelman S (1990) A network that learns to recognize three-dimensional objects. Nature 343:263–266

Rock I, di Vita J (1987) A case of viewer-centred object perception. Cogn Psychol 19:280–293

Rolls ET (1984) 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

Rolls ET (1987) Information representation, processing, and storage in the brain: analysis at the single neuron level. In: JP Changeux, M Konishi (eds), The neural and molecular base of learning, pp 503–540

Rolls ET, Baylis CG (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 macaque monkey. Exp Brain Res 65:38–48

Rosene DL, Roy NJ, Davis BJ (1986) A cryoprotection method that facilitates cutting frozen sections of whole monkey brains for histological and histochemical processing without freezing artifact. J Histochem Cytochem 34:1301–1315

Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res 149:1–24

Seibert M, Waxman AM (1990) Learning aspect graph representations from view sequences. In: Touretzky DS (ed) Advances in neural network information processing systems, Vol 2. Morgan Kaufman, San Mateo CA (in press)

Snedecor GW, Cochran WG (1980) Statistical methods (7th edn). Iowa State University Press, Ames Iowa, pp 215–237

Yamane S, Kaji S, Kawano K (1988) What facial features activate face neurons in the inferotemporal cortex. Exp Brain Res 73:209–214

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