Recognition of Objects and Their Component Parts: Responses of Single Units in the Temporal Cortex of the Macaque

E. Wachsmuth, M. W. Oram, and D. I. Perrett

School of Psychology, St Andrews University, Scotland, KY16 9JU, United Kingdom

We investigated the role that different component parts play in the neural encoding of the visual appearance of one complex object in the temporal cortex. Cells responsive to the sight of the entire human body (but no to control stimuli) were tested with two subregions (head alone with the body occluded from sight and the body alone with the head occluded). Forty-two percent (22 of 53) of cells responded to the whole body and to *one* of the two body regions tested separately: 72% (17 of 22) responding to the head and 28% (5 of 22) to the rest of the body. Forty-two percent (22 of 53) of cells responded independently to *both* regions of the body when tested in isolation. The remaining cells (17%, 9 of 53) were selective for the *entire* body and unresponsive to component parts. The majority of cells tested (90%, 35 of 39) were selective for perspective view (e.g., some cells respond optimally to the side view of the body, others to the back view). Comparable levels of view sensitivity were found for responses to the whole body and its parts. Results indicate (1) separate neuronal analysis of body parts and (2) extensive integration of information from different parts. Contrary to influential models of object recognition (Marr and Nishihara, 1978; Biederman, 1987), the results indicate view-specific processing both for the appearance of separate object components and for integration of information across components.

Visual object recognition is a fundamental part of our everyday activity. The brain is able to compare sensory information with internal representations of objects apparently independent of the object's orientation, distance, part occlusions, and lighting conditions. This raises the question of how the brain matches the infinite number of different retinal images of one object to the representation (or representations) of the same object stored in memory.

Two important issues have been raised in the context of object recognition. These concern (1) the role of the object's components and (2) the frame of reference used for specifying the spatial relationship between components of that object. Two types of frames of reference have been considered: object-centered descriptions relate an object's component parts to a framework based on the object itself, whereas viewer-centered descriptions relate an object's component parts to a framework based on the observer. This article presents a physiological study of the role of the component parts and perspective view in the coding of one type of complex object, the body.

The Role of Component Parts in Object Recognition

Several theories of object recognition have suggested that the processing of an object's parts plays an important role in the initial stages of recognition (Marr and Nishihara, 1978; Marr, 1982; Biederman, 1987). Recently, however, Baker Cave and Kosslyn (1993) have suggested that even though coding of parts is important for recognition, parts are processed only *after* the processing of the configuration of the whole object.

To isolate the component parts, the image of the object may be segmented at regions of sharp concavity of the external boundary (Marr and Nishihara, 1978; Hoffman and Richards, 1984; Biederman, 1987). Each of the resulting image regions can then be treated as if it corresponded to a volumetric primitive (i.e., a 3D component). Theories differ as to the type of volumetric primitives thought to be used in object recognition. Marr suggested that objects could be built up from a set of generalized cones (Binford, 1971; Marr and Nishihara, 1978). By contrast, Biederman postulates a more extensive set of 36 types of cone components called geons (Biederman, 1987).

Marr has suggested that the position, 3D orientation, and size of each cone component of an object

are described in relation to the object's principal axis (the longest axis of the object; Marr and Nishihara, 1978). Biederman (1987) relates geon components to other geons rather than to the object's principle axis. In Biederman's model the position and orientation of each geon are specified relative to other geons in qualitative terms (e.g., geon 1 to the side of geon 2 and joined at the expanded end). Both models suggest that the entire 3D object description can be accessed from the sight of key component parts. Thus, even when parts of an object are occluded from sight, recognition can occur on the basis of the remaining visible geons or cone components. From both Marr's and Biederman's theories, one might expect to find cellular units late in the visual pathway that respond selectively to one type of object and are activated by the sight of any major part of that object (a major part of an object can include several geons).

The processes of accessing the entire 3D description from the sight of one 3D part are similar to the "completion" property of parallel distributed processing (PDP) networks. Networks trained on a 2D image and tested with any large part of the trained image will settle into a pattern of activity within the network, which is equivalent to activity pattern generated by the complete training image.

View Specificity

Object-centered Representation

Both Marr's and Biederman's accounts of object recognition can be considered object centered (Marr and Nishihara, 1978; Biederman, 1987). In these models only one descriptive representation of the object is stored in long-term memory (see also Lowe, 1987; Porrill et al., 1988). This description should be accessible from all viewpoints provided that the principle axis is fully visible (Marr and Nishihara, 1978)

As mentioned above, under Biederman's account recognition can be based on a very small number of geons. Thus, as long as a sufficient number of geons are visible, recognition should not be affected by view. Under both object-centered schemes, perspective view should not affect the cellular mechanisms involved in the higher stages of object processing (beyond the limitations noted). One would expect, from both Marr's and Biederman's theories, to find cellular units late in the visual pathway that code an object in a way that is accessible from *all* views (i.e., cells that respond to all views of an object).

Recent psychological studies show that for brain-damaged and normal subjects, ease of recognition is influenced by the visibility of an object's salient features and cues to the 3D orientation of the object (Warrington and Taylor, 1973; Humphreys, 1984; Warrington and James, 1986; Quinlan, 1988; Humphrey, 1989).

Viewer-centered Representation

Representation of an object in a viewer-centered manner relies on descriptions of the object relative to the viewer. Such description includes a collection of the 2D visual characteristics of the object that are visible from a specific viewpoint. The number of characteristics present in any one viewer-centered description is therefore smaller than that of an object-centered description. Two main types of viewer-centered representation have been considered, in which (1) only one view of the object is stored or (2) multiple views are stored in long-term memory (Palmer et al., 1981; Jolicoeur, 1985; Tarr and Pinker, 1989; Ullman, 1989; Edelman and Bülthoff, 1990; McMullen and Farah, 1991; Cutzu and Edelman, 1992; Verfaillie, 1992). Note that under scheme 1, where a single view is stored, the orientation of the object's components would be specified with respect to the viewer. Hence, this single description is not object centered.

Palmer et al. (1981) have defined the "canonical views" as the single view that reveals the maximal information about an object's salient features. For most objects (8 of 12 objects of those studied by Palmer et al., 1981), the canonical view lies between the "front" and "side" views, with the object's principal axis being oriented 45° to the observer's line of sight. In addition, for heads, the 45° (or ¾) view is the most readily recognized in naming and matching tasks (Bruce et al., 1987). Models envisaging storage of a single viewer-centered description (such as the canonical view) rely on a transformation of the incoming image to match the stored description. The transformation may involve processes akin to mental rotation (Shepard and Cooper, 1982). If an object is represented by a single canonical viewer-centered description, then one might expect to find greater numbers of cells selectively tuned to this view of an object than to other views of the same object.

Most viewer-centered models of object recognition suggest that *several* views of an object are represented in memory (Koenderink and van Doorn, 1979; Tarr and Pinker, 1989; Ullman, 1989; Edelman and Bülthoff, 1990; Poggio and Edelman, 1990; Seibert and Waxman, 1991). The theories, however, differ as to the number and nature of the views stored. In these models either the incoming image is transformed to match to the nearest stored view, or alternatively the image is identified by interpolation from a minimum of three surrounding stored views. Multirepresentational viewer-centered models suggest the presence of cells coding different specific views of the same object.

Cellular Sensitivity to Objects and Their Parts

It has been proposed that the inferotemporal cortex (IT) plays a central role in visual pattern and object recognition (Gross, 1973, Ungerleider and Mishkin, 1982). Most physiological studies of this area have been concerned with the coding of geometrical features that may occur in several objects (such as bars, circular areas, Fourier descriptors; Gross et al., 1972; Schwartz et al., 1983; Desimone et al., 1984; Tanaka and Fujita, 1991; Fujita et al., 1992; Komatsu and Ideura, 1993).

Studies in IT and neighboring cortex within the superior temporal sulcus (STS) have also revealed populations of cells with greater visual selectivity that

respond preferentially to particular complex biologically important stimuli such as hands and faces (Gross et al., 1969, 1972; Perrett et al., 1982, 1989; Kendrick and Baldwin, 1987; Desimone et al., 1990). These cells offer an opportunity to determine the importance of an object's component parts in the processing of the entire form of the object. Hence, these cells can be used to investigate the neurobiological validity of psychological and computational models of object recognition.

Early studies focused on the importance of facial components for cell responses to the whole face. Different cells were found to be selective for different regions of the face (some tuned to the eyes, others to the mouth; Perrett et al., 1982). Though not systematically studied, it was also noted that most cells responded to several regions tested in isolation (Perrett et al., 1982; Desimone et al., 1984). Facial characteristics, particularly when seen from the front, are defined by differences in pigmentation and surface structure. The parts cannot be segmented from points of maximum concavity in the external contour (silhouette). Hence, models such as those of Marr and Biederman that embody segmentation at concave points in the external boundary may be less applicable to the processing of the internal structure of the face.

In the present study we extended the investigation of the role of components in the processing of whole objects by comparing the response of cells to the entire body and to two major parts: the head with the body occluded from sight and the body with the head occluded from sight.

Cellular Sensitivity to Object View

Physiological studies have investigated the view sensitivity of cells responsive to faces in the temporal cortex (Desimone et al., 1984; Perrett et al., 1984, 1985, 1989, 1991, 1992; Kendrick and Baldwin, 1987; Hasselmo et al., 1989a,b). All studies reveal that the majority of cells are sensitive to change in perspective view. That is, most cells code in a viewer-centered fashion. Different cells are tuned to different optimal views (some to the face, some to the back of the head, etc.). It is interesting to note that in humans evoked potential studies also indicate viewer-centered processing of the face (Bötzel and Grüsser, 1989; Jeffreys and Turkmachi, 1992).

A small population of cells in the macaque STS has been found to respond equally to all views of the head (Perrett et al., 1984, 1985, 1989, 1991, 1993; Hasselmo et al., 1989b). The insensitivity of these cells to changes in view is in accordance with the definition of object-centered coding. Object-centered cells could arise from a combination of the outputs of different viewer-centered cells (Perrett and Oram, 1993). This suggestion is supported by the finding that latencies of object-centered cells are longer than those of viewer-centered cells (Perrett et al., 1992).

An additional aim of the present study was to compare view tuning for the whole body with view tuning for component body parts. This aspect of the study could provide potential insight in the role of view-selective processes in the integration of information about separate object parts.

Preliminary results of have been reported previously (Perrett et al., 1993; Wachsmuth et al., 1993).

Materials and Methods

Recordings of responses of single cells are from five macaque monkeys (*Macaca mulatta*; two females, 4-8 kg; three males, 5-8 kg). The techniques applied have been previously described (Perrett et al., 1991).

Training and Fixation Task

Presurgical training involved the following. Situated in a primate chair, the monkeys were trained to fixate on an LED light presented onto a white wall at eye level at a distance of 4 m. The monkey's task was to discriminate the color 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 a pseudorandom order under computer control.

Surgery

Under full sterile conditions two stainless steel rings (16 mm diameter, 10 mm deep) were surgically implanted bilaterally under sodium pentothal (Sagital) anesthesia with their centers at predetermined stereotaxic coordinates 12-14 mm midline and anterior to the interaural plane. Two plastic tubes (5 mm diameter) were attached with dental acrylic horizontally in front and behind the wells. By passing a metal bar through each tube, the monkey's head could be restrained during recording sessions. The monkey was allowed to recover from surgery and retrained until presurgical performance in the color discrimination task was reached.

Recording Techniques

Before each recording session a topical anesthetic (xylocaine, 40 mg/ml) was applied to the dura. A David Kopf micropositioner was fixed to the recording well, allowing a tungsten in glass microelectrode to be inserted through a transdural guide tube into the temporal cortex, aiming for the anterior part of the STS (areas TPO, PGa, and TAa of Seltzer and Pandya, 1978).

The LED light and the visual stimuli were presented from behind a shutter with rise time of less than 15 msec. A large-aperture (6.5 cm diameter) electromechanical shutter (Compur) or an alternative (20 cm square) liquid crystal shutter (Screen Print Technology Ltd.) was used. The shutter opened for 1.0 sec after a 0.5 sec warning tone that allowed the monkey to prepare fixation of the LED position before the shutter opened. This enabled the monkey to lick several times for multiple juice rewards during the trial period. Visual stimuli (2D) were projected onto the white wall in front of the monkey on which the LED light was positioned, and 3D stimuli were presented in front or to either side of the LED. The test stimuli were interspersed with control stimuli and a no-stimulus

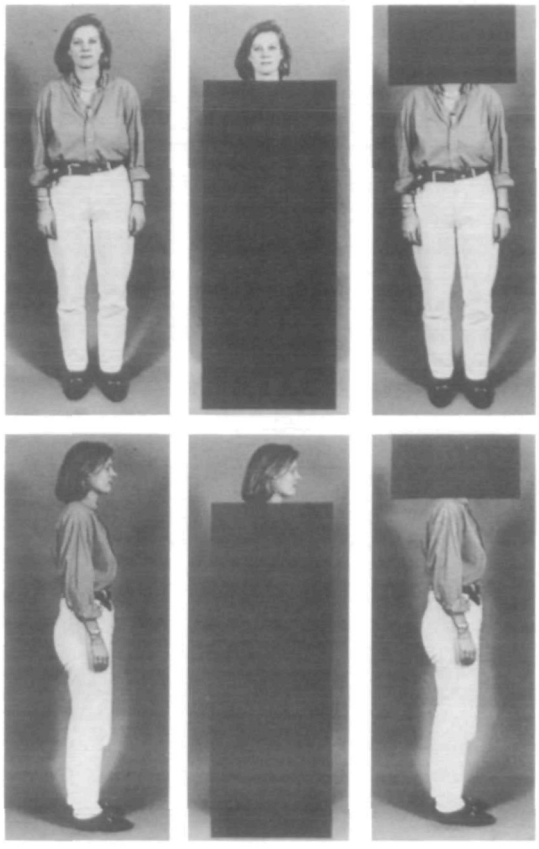

Figure 1. Examples of stimuli used for testing: whole-body, head-only, and body-only stimuli in different views.

condition (the LED alone, to measure spontaneous activity).

The neuronal activity recorded was amplified, filtered, and displayed on an oscilloscope using standard techniques and equipment. The spike activity was also converted to digital signals and stored on a PC-compatible computer using CED1401 (Cambridge Electronic Design) in 5 msec time bins. Eye movements were recorded with each trial using an infrared reflection system (ACS, modified to allow recording of both horizontal and vertical signals from one eye), and stored with 8-bit accuracy.

Visual Stimuli

Neuronal responses were recorded for both real 3D objects and 2D objects (video disk images and slides). The visual stimuli tested were the (1) whole human body (head and body), (2) head alone, and (3) body alone (defined here as torso, arms and legs; see Fig. 1). The posture in all cases was bipedal. The human body stimuli were photographed onto a 200 ASA ectochrome slide film, with the person standing against a light gray background. Slides with occluded body parts were produced by first taking the whole-body slide and then occluding the unwanted parts with black tape.

Alternatively the stimuli were filmed with a video camera (JVC BY-110E), recorded on ¾ inch U-Matic videotape, edited on a JVC editing suite (control unit RM-88U), and transferred on a laser video disc (RLV Mk II, Optical Disc Corp.). The video stimuli were then replayed with a video disk player (Philips VP406 LaserVision Disc Drive) and projected onto the display screen (using a Sony color video projector VPH-1041QM).

Where 2D stimuli did not activate cells, 3D stimuli were used. 3D head-only and body-only stimuli were created by occluding the unwanted parts with a large sheet of black cardboard or curtain material.

For most cells multiple views of the stimuli were tested: front view (0°), left profile (90°), back of head (180°), and right profile (270°). In addition to these, four intermediate views (45°, 135°, 225°, and 315°) were tested for some cells. For each cell, a number of different control stimuli were tested (2D and 3D). These included complex 3D objects of different sizes, shapes, and textures (lab coats, chairs, etc.), simple 2D geometrical shapes (bars, spots, and gratings) and simple 3D forms (cylinders, balls, 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. Cells found to be responsive to static views of the whole body were then further investigated for sensitivity to body parts. After identifying the optimal stimulus with a 3D whole body, further testing was carried out with 2D stimuli. Stimulus order was presented in a computer-controlled pseudorandom order. The cell was then tested with at least five trials of each of the whole body, its two parts (head only and body only), and different control objects. Testing was performed, where possible, with stimuli of the (cell's) preferred view, the opposite view (180° rotation was usually the least effective view), control stimuli, and no stimulus. The views used for testing components were the same as used for testing the whole body.

Data Analysis

Since most cells in the STS respond with a latency of 100 msec (±30 msec; Oram and Perrett, 1992), the magnitude of cell activity on individual trials was assessed over the 0.25 sec time period occurring 100-350 msec after stimulus onset. For some cells [with late response onset (>200 msec) or inhibitory responses] a 0.5 sec time period (100-600 msec post-stimulus) was used to assess cell activity.

Cell responses to the whole body and two components (head only, body only), controls, and spontaneous activity were compared on line by using one-way ANOVA and post hoc tests [protected least significant difference (PLSD); Snedecor and Cochran, 1980]. The influence of view on cell responses to the whole and component parts was analyzed off line with two-way ANOVA with view and body parts tested as main factors.

Recording Sites

After each recording session, frontal and lateral x-ray photographs were taken to localize the electrode. Mi-

crolesions (10 μA DC for 30 sec) made at the end of some electrode tracks, subsequently identified using standard histological techniques, allowed reconstruction of the electrode position within the brain. In addition, reference markers were made by injection of HRP and fluorescent dyes true blue and diamidino 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 anesthetic. After transcarinal 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% dimethyl sulfoxide and 20% glycerol (Rosene et al., 1986; see Harries and Perrett, 1991, for detail).

Results

For five subjects, a total of 7287 cells were screened for neuronal responses. Of these, 23% (1691 of 7287) were found to be visually responsive, including cells responsive to moving or static visual stimuli (see, e.g., Bruce et al., 1981; Oram et al., 1993) and visual general cells (where no apparent specific visual stimulus drives the cell). Of the visually responsive cells in the cortex of the STS, a total of 64 cells that were found to be responsive to the whole body [i.e., have significantly greater response to the whole human body than to control objects and spontaneous activity (S/A)] were tested for selectivity to component parts of the body. All the cells included in this study did not selectively respond to a wide variety of different control objects tested. For 53 of these cells responses were measured for two body parts [head only; body only (torso, arms, and legs)] and the whole body (head and body). The remaining 11 cells were tested for the whole body and only one part, and therefore are not included in the analysis of coding component parts, but are included in the view discrimination study. Cells were categorized on the basis of their response to the two parts: they could either respond to none, one, or two parts.

View sensitivity was investigated for a total of 39 (of 64) cells. For 28 of these cells, view sensitivity was tested for the whole and both parts, whereas for 11 cells view sensitivity was measured for the whole body and only one part.

The optimal stimuli for a given cell was defined as the stimuli causing maximal change in the firing rate relative to baseline activity (S/A). For the majority (60 of 64) of cells, optimal stimuli produced excitatory responses. Four cells gave inhibitory responses to optimal stimuli. For clarity of explanation, "greater" response is defined as greatest change from S/A (whether excitatory or inhibitory).

Coding of Parts

Twenty-two cells responded to *only one of the two* component body parts when tested in isolation.

Cells Only Responsive to the Head

For 17 cells (32% of the 53 cells tested) the response to the sight of the head alone (with the rest of the body occluded from sight) was significantly greater than that to controls and to S/A. Additionally, for these cells the body alone (with the head occluded from sight) did not produce a response that was significantly different from controls or S/A. This pattern of response is shown in Figure 2*A*. For this cell the presence of the head was both necessary and sufficient to account for the response to the whole body.

Figure 2*B–D* also displays responses of the same cell recorded on the five individual trials with the whole-body, head-alone, and body-alone stimuli. Responses to the whole body (Fig. 2*B*) occurred at approximately 120 msec after stimulus onset, with an initial transient burst (lasting ~350 msec) following a response decline, though remaining substantially greater than the prestimulus activity typical for the cells found in STS (Oram and Perrett, 1992). Figure 2*C* shows responses of similar latency and time course of activity during trials when the head alone was presented. When the body was presented in isolation (Fig. 2*D*), however, there was no change in cell activity in comparison to the prestimulus period. Thus, the body presented without the head was an ineffective stimulus for this particular cell.

In addition, Figure 2*B–D* indicated that for all stimulus types the position of the eyes was held constant (within ±5°) during the time of neuronal response analysis (100–350 msec). Thus, differences in cell responsiveness for the different types of visual stimuli (effective or ineffective) were not due to different patterns of fixation.

Cells Only Responsive to the Body (Torso and Limbs)

Nine percent (5 of 53) of the cells showed a response to the body tested in isolation, which was significantly different from the response to control objects and S/A (e.g., Fig. 3). For these cells the responses to head alone did not differ from S/A or the response to control objects.

The 22 cells described in this section were responsive to one of the two body parts tested in isolation. This classification could include cells that show hidden sensitivity to multiple body parts. Thus, even though a cell responds only to one component when tested in isolation, the other component may influence the response to the whole object. For 12 of these cells, there was no significant difference between the response to the effective component part and response to the whole body. For these cells the response to one part was necessary and sufficient to account for the response to the whole body. The remaining 10 cells showed a significant difference between the responses to the effective component-part and the whole-body stimuli (see, e.g., Fig. 3).

Coding the Entire Body

Cells Only Responsive to Whole Body

Seventeen percent (9 of 53) of the cells showed a response depending on the visuality of the whole

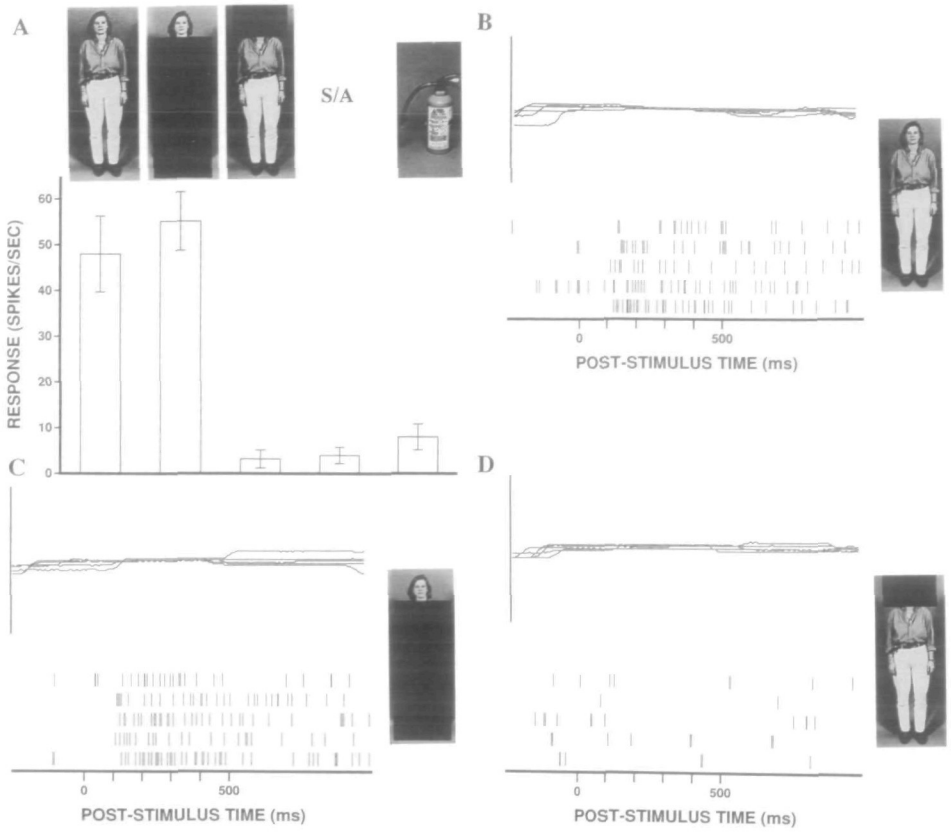

Figure 2. Neuronal responses of a cell only responsive to the head. *A:Top*, Photographic representation of stimuli used for testing. *Bottom*, Mean responses (±1 SE) to the whole body, its components (view 0°), spontaneous activity (*S/A*), and control stimuli are illustrated for one cell (E29_33.56). The responses to the whole-body and the head-only stimuli were not significantly different ($p > 0.05$), whereas both these stimuli produced a significantly higher neuronal response than activity rates produced by the body-alone stimulus, S/A, or control objects [$p < 0.005$ for each comparison; ANOVA: $F(7,39) = 26.1$, $p < 0.0005$]. *B–D*, Rastergram displays of responses of one cell (E29_33.56) on five trials for each of the stimulus conditions. Each trial (originally in pseudorandom order) is represented by a *single row of ticks marks*; each *tick mark* indicates one action potential. Prestimulus time is given at the figure base. Horizontal eye movements are shown for each trial in the *upper sections*. The responses to the whole body (*A, far left column*) and the head tested in isolation (*second column from left*) were not significantly different ($p > 0.05$), whereas both these stimuli produced a significantly higher neuronal response than activity rates produced by body alone (*center column*) or control objects (*far right column*) ($p < 0.005$ for each comparison) ANOVA: $F(7,39) = 26.1$, $p < 0.0005$. Vertical calibration: ±100° for horizontal (eye movements were recorded over a range of positions ±20° from straight ahead).

body. These cells responded to *neither* component part when tested in isolation. For these cells the whole-body stimulus produced a response significantly different from that to either of the body components (head alone and body alone), control objects, and S/A. The responses to the components individually did not show any statistical difference from responses to controls or S/A (e.g., Fig. 4).

Cells Responsive to Multiple Body Parts

The remaining 42% (22 of 53) of the cells studied showed responses to *both* components of the body when tested in isolation. The responses to components and the whole body were significantly different from the responses to control objects and S/A Typically the response to the whole body was greater than the response to the individual components (e.g., Fig. 5).

View Sensitivity

View sensitivity was examined for the sight of the whole body for 39 cells (of the 53 cells described above).

Cells with Viewer-centered Properties

Ninety percent (35 of 39) of the cells studied displayed sensitivity to view. These cells showed significantly different responses between (a minimum of) two views of the whole body. As should be evident from the Figures 2–5, different cells were found to be tuned to different views (some responded maximally to the front view of the body, others to the back, others to the left profile, others to the right profile, etc.). The view preference exhibited by cell responses to the whole body generalized to the parts (e.g., Fig. 6).

The view discrimination for the whole and parts was found to be compatible for all cells (see below). If a cell exhibited view discrimination between two views of an isolated body part, then the cell also showed the same direction of view discrimination for the whole body. Two cells, however, responded to one part tested in isolation but only displayed view discrimination for the whole body.

Figure 6*B–D* shows poststimulus time histograms (PSTHs) for a viewer-centered cell that discriminated view for the whole body and its component parts. The cell responded with a latency of approximately 100–

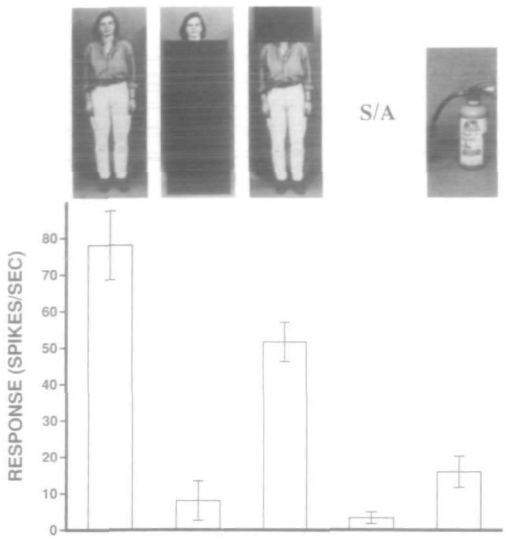

Figure 3. Neuronal responses of a cell only responsive to the body. *Top*, Photographic representation of stimuli used for testing. *Bottom*, Mean responses (±1 SE) to the whole body, its components (view 0°), spontaneous activity (*S/A*), and control stimuli are illustrated for one cell (B77_21.54). The whole-body and the body-alone stimuli produced no significant difference in the cell's activity rate ($p > 0.05$), whereas both these stimuli produced significantly higher neuronal activity than activity rates produced by head-alone stimulus, S/A, or control object stimuli ($p < 0.005$ for each comparison). ANOVA: $F(4,39) = 11.4$, $p < 0.0005$.

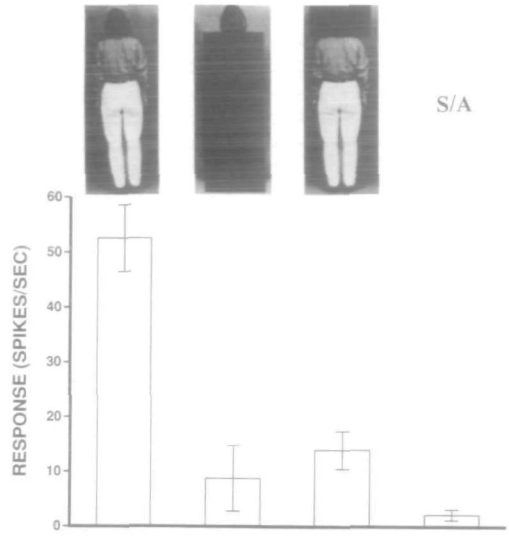

Figure 4. Neuronal responses of a cell only responsive to the whole body. *Top*, Photographic representation of stimuli used for testing. *Bottom*, Mean responses (±1 SE) to the whole-body, head-alone, and body-alone stimuli (view 180°) and spontaneous activity (*S/A*) are illustrated for one cell (D107_35.41). The whole-body stimuli gave a significantly higher response than the other stimuli tested ($p < 0.0005$ for each comparison). The head tested in isolation and the body tested in isolation did not show any significant difference from the S/A ($p > 0.05$ each comparison). ANOVA: $F(3,23) = 35.3$, $p < 0.0005$.

120 msec. After an initial transient burst (lasting ~300 msec) the response declined, though remained greater than the prestimulus activity. This pattern of cellular activity can be seen most easily for the whole-body stimuli (*C*), but is also present to a lesser extent for the head-alone (*B*) and body-alone (*D*) stimuli. Note, that the activity profile for each of these stimuli presented in the preferred view is enhanced, compared to the activity profile for the nonpreferred view.

Cells with Object-centered Properties

Ten percent (4 of 39) of the cells showed no preferred view for either the whole body or its components (Fig. 7).

View Discrimination Indices

View Discrimination: Whole Body versus Head Alone

The responses of 18 viewer-centered cells were used in a population analysis to compare the efficiency of view discrimination for the whole body and the head presented alone (Fig. 8*A*). Cells were only included in this analysis if they responded to the whole body ***and*** to the head when tested in isolation.

A view discrimination index was computed for each cell from the view producing the greatest change in activity from S/A (best angle) and smallest change in activity (worst angle) to the whole-body stimulus. The cell's activity was then measured to the component stimulus (head alone) in the same two views.

The formula used to calculate these indices was the same as used previously (Oram and Perrett, 1992): I_p = [(response to best angle − S/A) − (response to worst angle − S/A)]/(response to best angle − S/A).

The distribution of I_p values is shown in Figure 8*A* for the whole body and for the head alone. For an index value of 1 0 response to the worst view was the same as S/A. Index values >1.0 arise when the cell response to the best view was greater than S/A, and the response to the worst view was less than S/A. I_p can have a negative value if a neuronal response is

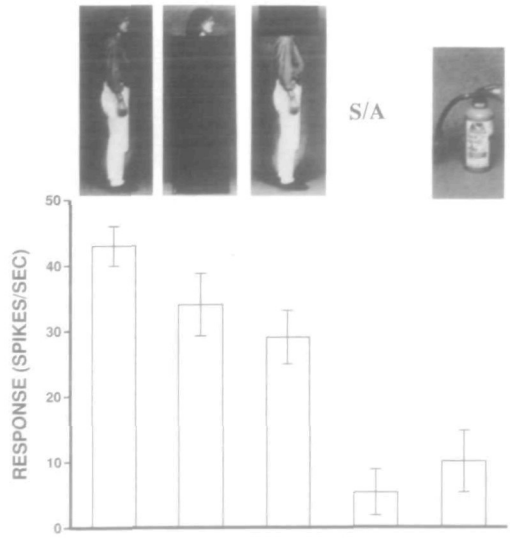

Figure 5. Neuronal responses of a cell responsive to multiple parts. *Top*, Photographic representation of stimuli used for testing. *Bottom*, Mean responses (±1 SE) to the whole body, its components (view 270°), control stimuli, and spontaneous activity (*S/A*) are illustrated for one cell (D30_27.72) The cell responded more to the whole-body stimuli ($p < 0.05$ each comparison) than to any other stimuli. The cell also responded significantly more to either of the body regions tested in isolation (head-alone, body-alone stimuli) than to S/A or control objects ($p < 0.005$). ANOVA: $F(4,18) = 14.1$, $p < 0.0005$.

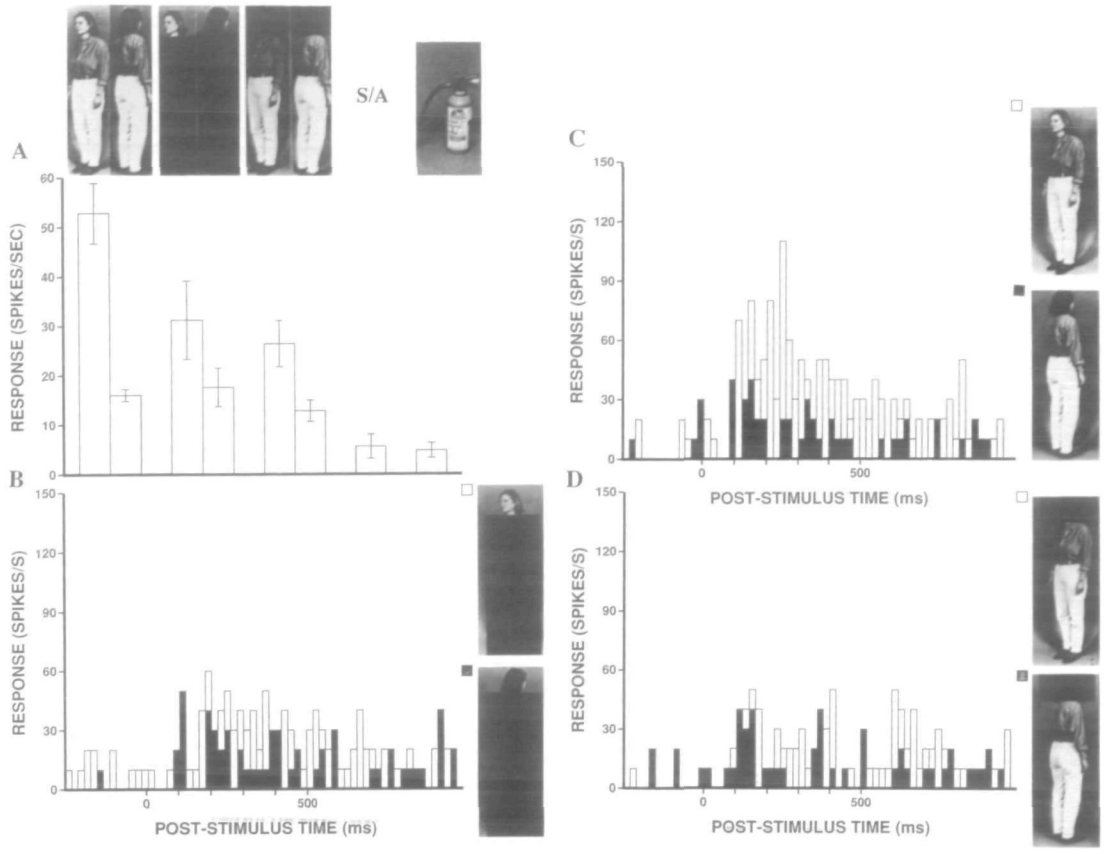

Figure 6. Neuronal responses of a cell with viewer-centered properties to the body and its components. *A*, Histogram of response (spikes/sec) to different stimuli. *Top*, Photographic representation of stimuli used for testing. *Bottom*, Mean responses (±1 SE) of one cell (J126.33.66) tested to view 45° and view 225° of the whole body and its components. Two-way ANOVA revealed a significant main effect of view [$F(1,24) = 23.28$, $p < 0.0005$], body part tested [$F(2,24) = 3.34$, $p < 0.05$], and interaction between these factors [$F(2,24) = 2.49$, $p < 0.5$]. Protected least significant difference tests (PLSD), post hoc tests, indicated significant response discrimination between the different views of the entire body ($p < 0.0005$) and of the body parts tested in isolation: head only ($p < 0.05$) and body only ($p < 0.05$). *B–D*, Poststimulus time histograms (PSTHs) show averaged responses from five trials (bin width, 20 msec). The cell responded strongly to the sight of the whole body and the component parts presented in view 45° (*white bars*), but failed to respond to the same stimuli in an opposite view (*black bars*, view 225°; stimulus onset at time 0). The *ordinate* of the PSTHs denotes the cell responsivity for 100 spikes/sec.

numerically greater for the "worst" view than the "best" view. This can only occur when the index is computed for the component body parts, since the best and worst views were defined on the basis of responses to the whole body.

For cells responsive to the head, the distribution of I_v for the entire body was not significantly different from the distribution of I_v when the head was tested alone (matched pairs $t = 0.20$, df $= 17$, $p = 0.85$).

View Discrimination: Whole Body versus Body Alone

The efficiency of view discrimination for the whole body and the body alone was computed in a similar way for 10 cells (Fig. 8*B*). These cells were responsive to both the whole body *and* to the body without the head visible. The distribution of I_v values obtained with the whole body visible was not significantly different from that obtained from the body alone ($t = -0.76$, df $= 9$, $p = 0.47$).

This analysis was repeated for seven cells responsive to multiple parts and again it was shown that the distribution of I_v for the entire body did not significantly differ from the distribution of I_v for head alone (matched pairs $t = -0.74$, df $= 6$, $p = 0.49$) or body alone (matched pairs $t = -1.5$, df $= 6$, $p = 0.19$).

Thus, as a population of cells there was no significant difference of quality for view discrimination between viewing the entire body or its isolated parts.

Histological Localization

Reconstruction of cell position in four monkeys revealed that cells responsive to the head and body were located in the upper bank (and to a certain extent in the lower bank) of the anterior STS. Figure 9 illustrates the position of the different cell types recorded in one monkey. The different cell types were intermixed within the cortical areas sampled. For one monkey histology was unavailable but x-ray analysis revealed the recording site to be in the same brain region.

Discussion

Cell Sensitivity to the Body

Previous studies of form processing of biological stimuli in the temporal cortex have focused on cell re-

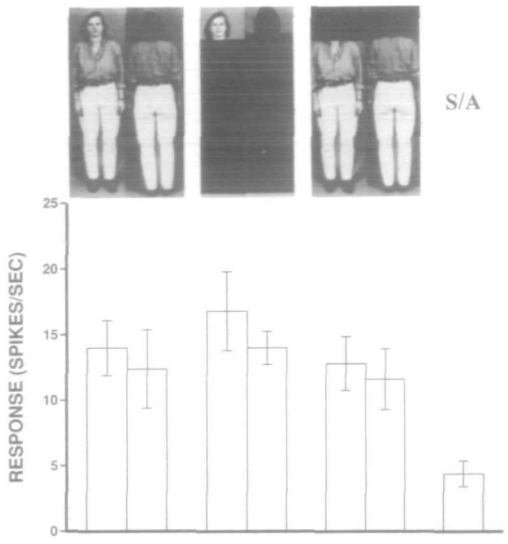

Figure 7. Neuronal responses of a cell with object-centered properties to the body and its components. *Top,* Photographic representation of stimuli used for testing. *Bottom,* Mean responses (± 1 SE) of one cell (J22_27.78) to front (view 0°) and back (view 180°) views of the whole body and its components are shown. Two-way ANOVA showed a significant main effect of the body part tested [$F(2,24) = 10.3$, $p < 0.005$] but no effect of view [$F(1,24) = 0.9$, $p > 0.3$], and no interaction between view and part tested [$F(2,24) = 0.2$, $p > 0.5$].

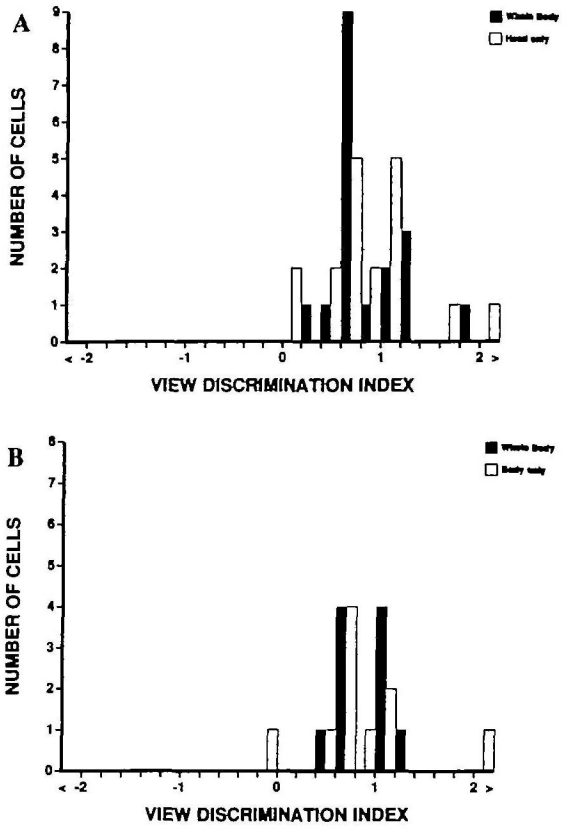

Figure 8. View discrimination indices. *A,* Whole body versus head alone. An index of view discrimination was computed (see text) for 18 viewer-centered cells responsive to the whole body and the head alone. The *black bars* display the ability of the cell population to discriminate between views of the whole-body stimuli. The *gray bars* display the ability of the same cells to discriminate between views of the head alone. The distributions of index values were not significantly different ($t = 0.20$, df = 17, $p = 0.85$). *B,* Whole body versus body alone. A similar comparison was made for 10 viewer-centered cells responding to the whole body and the body alone. The *black bars* display the population's ability to discriminate between views of the whole-body stimuli. The *gray bars* display the population's ability to discriminate between views of the body-only stimuli. The distributions did not significantly differ ($t = -0.75$, df = 9, $p = 0.47$).

sponses to the sight of the face and other views of the head (Perrett et al., 1982, 1984, 1985, 1989, 1991, 1992; Kendrick and Baldwin, 1987; Hasselmo et al., 1989b). In the present study it was not surprising to find cells responding selectively to the sight of the head alone, since the face carries much social information.

The surprising result of the present study was the extent to which information arising from parts of the entire body, other than the head, influence cell responses. The responses of the majority of cells (83%, 44 of 53) carried information about regions of the body other than the head. There were three types of cell for which body information was found to be important. For the five cells responsive to the body alone, information about the head in isolation was insufficient to drive responses. For 22 cells independent responses could be measured both to the head and to the rest of the body. Finally, for nine cells, information from the head and the body was critical before any response could be stimulated.

The high proportion of cells found in the present study sensitive to body information suggests that other studies overlooked the importance of the body. It is unlikely that we have uncovered four new classes of cell; rather, we may have revealed additional information processing abilities of cells that have already been described. In previous studies cells have been described as "face responsive" or even dubbed "face cells." The present analysis suggests such labeling may be inappropriate, as it can underestimate or bias interpretation of cell selectivity and functions. The function of cells may require an integration of information from multiple body parts including the eyes, face, hands, body posture, and so on (Perrett et al., 1992).

Implications for Models

Cells Selective for One Body Part

Forty-two percent (22 of 53) of the cells studied responded to only one of the two body parts tested. Hence, the component parts of an object appear to be coded separately within higher visual association cortex. One might suppose that such findings fit models suggesting an initial encoding of objects in terms of their component 3D volumetric parts (Marr and Nishihara, 1978; Biederman, 1987). Detailed consideration (below) indicates that the present findings do not fit these models.

The recordings indicate response sensitivity to body components that are more complex than simple 3D volumetric shapes. First, the cells are unresponsive to a wide variety of simple and complex control stimuli. If a cell was selective for a particular geon (e.g., a cylinder shape), then one would expect responses both to the shape of a human body that would include several cylindrical components and to a great variety

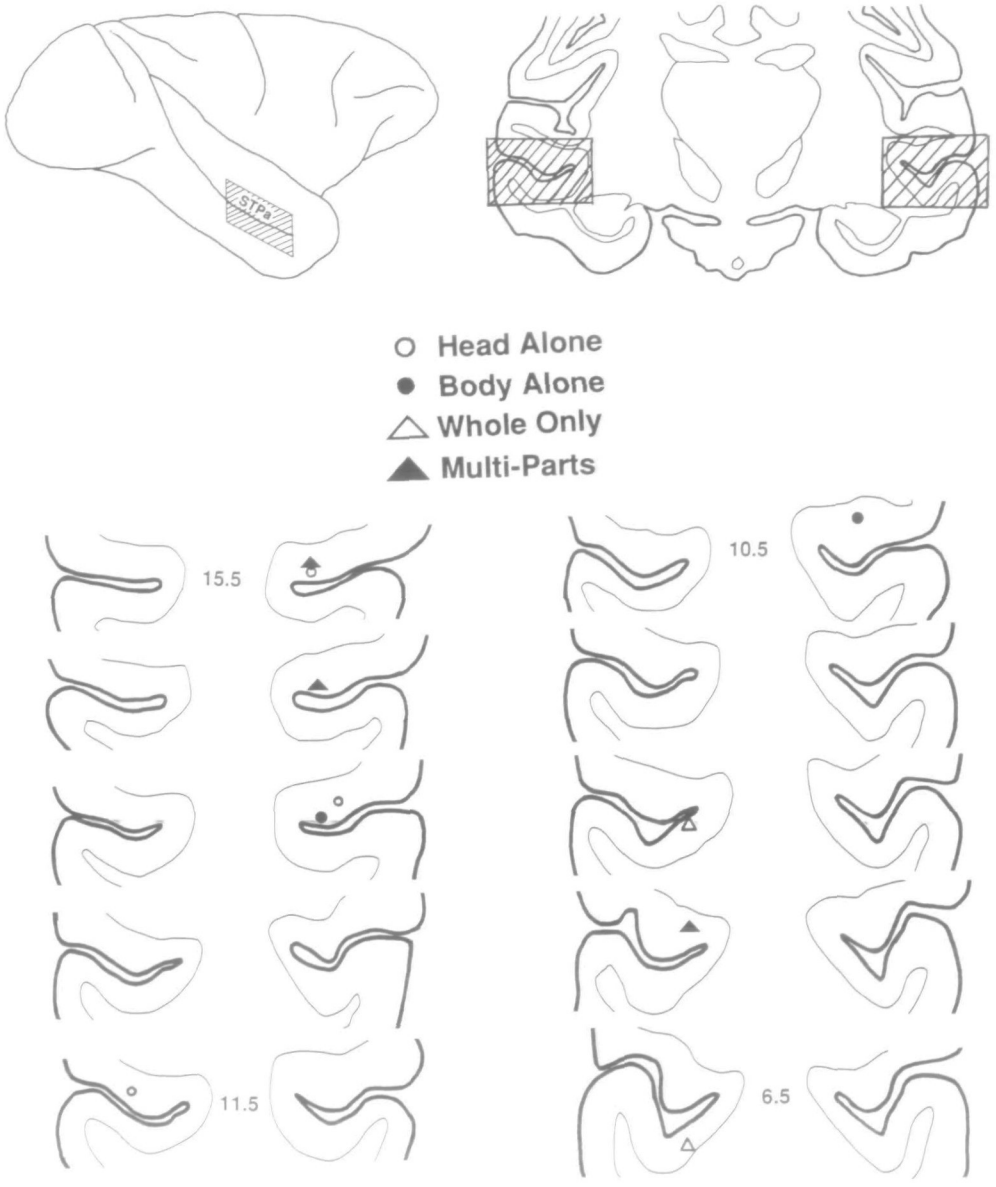

Figure 9. *Top left,* Schematic drawing of the right hemisphere of the macaque brain. *Hatched region* indicates the location area STPa. *Top right,* Frontal section of the brain of one monkey (J) 9.5 mm anterior to the interaural plane. The STS of the left and right hemispheres is indicated by the *hatched areas. Bottom,* Series of frontal sections every 1 mm (from 15.5 to 6.5 mm anterior to the interaural plane) showing the location of cells responsive to the head alone (*open circles*), body alone (*solid circles*), whole body only (*open triangles*), and responsive to all parts (*solid triangles*). The *thick lines* indicate the brain surface; the *thin lines* show the boundary between white and gray matter. Cells were located in both the upper bank and fundus of the STS.

of control objects (e.g., a mug, a broom, etc.) that also have cylindrical components. Second, most cells showed sensitivity to body view. Indeed, some cells discriminate between mirror-symmetrical views (e.g., left profile but not right profile) that contain identical geometric shapes. If a cell response to one body part is to be explained by the presence of a particular 3D volumetric shape (geon or cylinder), then the cell should be equally responsive to all views where this shape remains visible. Though object components are encoded independently during visual processing, the complexity of the component parts receiving separate analyses within the STPa cortex is greater than suggested by the models of Marr and Biederman.

The results are also incompatible with viewer-centered models of recognition suggesting that only the global shape of particular object view is represented (without independent part representation; Koenderink and van Doorn, 1979; Edelman and Bülthoff, 1990; Seibert and Waxman, 1991).

Cells Selective for Multiple Body Parts

One population of cells (42%, 22 of 53) studied was responsive to the whole body and to multiple parts of the body when tested in isolation. This population of cells is in accordance with the models of Biederman (1987) and Marr and Nishihara (1978). These models predict one global description of an object stored in memory that is independently accessible from the sight of any major object component. How-

ever, the view tuning of these cells suggests that even these cells do not support these models (see below).

We note that categorization used may have underestimated multicomponent coding. Sensitivity to information from both the head and the body alone may have been present even for some of the cells categorized as responsive to only one part. First, for some of these cells (10 of 22) responses to the entire body were different from the response to the most effective part tested in isolation (e.g., Fig. 3), indicating that the "noneffective" part could influence response when viewed together with the effective part. Second, for four cells view tuning was present for the entire body but was not evident for the effective part tested in isolation.

Cells Selective for the Whole Body

The models of Marr and Nishihara (1978), Biederman (1987), and Lowe (1987) do not predict the population of cells (17%, 9 of 53) that were only responsive to the entire body. For these cells there was no activity provoked by isolated parts. Thus, the cell activity provides a description of the appearance of the entire body but this description is not accessible from the isolated parts. The cell responses provide information about the overall appearance of the object but they do not provide information about independent parts of the object.

The responses of such cells may fit suggestions of Baker Cave and Kosslyn (1993). These authors propose a model of visual processing in which the overall appearance of the entire object is processed before the parts. If the cells selective for the entire body represent a description of the overall configuration that is processed first, then one would not expect them to be activated by independent parts. The overall configuration of an object could perhaps be revealed from an analysis of coarse image attributes (e.g., low spatial frequencies). This course of analysis could be independent of the detailed form of individual object parts. It is noted, however, that the majority of cells selective for the whole body (as all other cell classes) showed viewer-centered properties in coding. View discrimination would not be predicted from a very coarse analysis of the body form.

With larger numbers of cells defined in each population, an analysis of the latency of the different cell types would be possible and may clarify the order of processing. Such analysis could reveal whether parts of an object are processed prior to the configuration of the whole object that would be consistent with "bottom-up" processing models. Alternatively latency analysis could reveal the opposite order of processing as suggested by some models employing "top-down" processing (e.g., Baker Cave and Kosslyn, 1993).

Sensitivity to View

Object-centered Coding

The analysis of view sensitivity revealed that representation of the body and its parts within area STPa occurs mostly in a viewer-centered fashion; 90% of cells studied were selective for view. This result is in marked contrast to models that employ object-centered representations since these representations should be accessible from all views (e.g., Marr and Nishihara, 1978; Biederman, 1987; Lowe, 1987; Porrill et al., 1988; Hasselmo et al., 1989b).

The remaining 10% of cells studied were found to respond to all views of the body that were tested and can therefore be considered as coding in an object-centered manner. Multiple inputs from cells with viewer-centered properties may account for the responses of object-centered cells. Latency estimates for cells sensitive to the head are consistent with this scheme (Perrett et al., 1992). It is possible that object-centered coding will be more frequently encountered in areas subsequent to the STPa.

Viewer-centered Coding

It emerged that different cells were selective for different views of the whole body. This parallels the observation that different cells in the STPa are tuned to a range of head views (Desimone et al., 1984; Perrett et al., 1985, 1991; Hasselmo et al., 1989b). Thus, no single canonical view of the body accounts for the entire range of view tuning (Palmer et al., 1981). The results are therefore consistent with models suggesting multiple viewer-centered representations stored in memory (Koenderink and van Doorn, 1979; Tarr and Pinker, 1989; Ullman, 1989; Edelman and Bülthoff, 1990; Poggio and Edelman, 1990; Seibert and Waxman, 1991; Cutzu and Edelman, 1992). As noted above, the results are consistent with a view-sensitive representation of parts, rather than a view-sensitive representation of the global form that is common to most viewer-centered models.

The present study did not set out to define the distribution of view tuning among cells responsive to the body. Nevertheless, view tuning for the whole body appears to have similar properties to view tuning for the component parts. For a minority of cells (4 of 22 cells) view tuning was apparent for the whole body but was not present for component parts that provoked responses when tested in isolation. As a population of cells, analysis showed view discrimination (as measured by our index) exhibited to the whole body that was equivalent to the view discrimination manifest to the head alone or to the body alone. There was no improvement in quality of view discrimination for the entire body, compared to a situation in which only one body part was visible.

Coding of Isolated Object Parts

One question arising from the present study is why parts of the body are coded independently of each other. Under natural viewing conditions individual parts of an body, or indeed any object, are often not fully visible. Such situations arise when a body is viewed from behind an intervening object or when one part of the body occludes the sight of another part. It would be impossible to recognize objects in such circumstances if the cortex only contained cells selective for the intact or entire object (such as that

depicted in Fig. 4). The separate coding of object parts allows recognition under conditions of partial object occlusion.

If each of the major (articulating) parts of an object is coded separately from a small number of views, then recognition of the object is possible despite changes in the configuration of these parts. An alternative processing scheme could rely on cells selective for the entire object's configuration, with different cells selective for different configurations. While this scheme is possible, it is also inefficient and would require large numbers of "templates" or cells, each selective for one of the huge range of possible configurations. Thus, an advantage of separate coding for distinct body parts is that such coding is compact and accommodates the great variety in configurations of those body parts. We note that single-cell coding for specific configurations may exist for some of the more meaningful body postures (e.g., Perrett et al., 1984).

Learning the Association between Parts

One issue that is raised by the present study concerns the mechanism by which the nervous system integrates information about different parts of the same object. Several unsupervised learning mechanisms (rules) have been proposed whereby output units in artificial or real neural networks "learn" the pairing of independent input patterns when these inputs are associated over time (e.g., Rumelhart and Zipser, 1985; Foldiak, 1990, 1991). After learning the association, the output units of the network are able to respond to any of the input paired patterns.

Studies of Miyashita and colleagues (Miyashita, 1988, 1990; Miyashita and Chang, 1988) have demonstrated that single cells in the anterior and ventral temporal cortex do register the temporal association of abstract patterns that are presented sequentially with an interval of 15 sec. A recent report by Sobotka and Ringo (1993), however, failed to find evidence that association between complex arbitrary patterns (presented simultaneously in pairs) is in cells of the inferior temporal (IT) cortex. After an extensive behavioral training period, some IT cells exhibited selective responses for a particular pattern but these cells were *not* more likely to be additionally selective for the paired pattern (Sobotka and Ringo, 1993). The results of the above studies suggest that associations are learned when the stimuli are presented sequentially (Miyashita, 1988, 1990; Miyashita and Chang, 1988) but not when presented simultaneously (Sobotka and Ringo, 1993).

Twenty-two of the cells described here were selectively responsive to either of the two components of the body tested in isolation. The body parts for which the cells' responses were conjointly sensitive are visually distinct, but nonetheless these two parts are physically related in the entire object. This physical (spatial) association means that the head and the body (in the same view) are frequently encountered together. We speculate, therefore, that the sensitivity of cells responsive to multiple parts is established through a learned association of simultaneously presented inputs from cells selective to only one body part.

The scheme described above implicitly assumes that cells responsive to individual body parts should be activated by the visual input before cells responsive to multiple parts. Further study of the latencies of the different cell types could define the order of processing and be used to determine the validity of the proposed processing scheme.

It was apparent in the present study that cells tuned for multiple parts exhibited compatible view tuning for these parts. If such a cell responded more to the front than the back view of the head, then the cell was also more responsive to the front than the back view of the rest of the body. This compatible view tuning supports the speculation that sensitivity to multiple body parts arises through experience of the association between the parts. For example, the face is seen in association with the front view of the body. If cells learn the association of parts, then it follows that the cells will tend to be selective for the same view of these parts (assuming that the parts are normally coaligned).

Notes

This research was funded by project grants from the United Kingdom MRC, SERC, and the United States ONR. E.W was awarded an SERC studentship. We acknowledge the contribution of L. K. Harrison, J. K. Hietanen, M. H. Harries, and P. J. Benson, who participated in some of the experiments. We are grateful for the support given by University technical and photographic staff.

Correspondence should be addressed to E. Wachsmuth, School of Psychology, St Andrews University, Scotland, KY16 9JU UK

References

Baker Cave C, Kosslyn M (1993) The role of parts and spatial relations in object identification. Perception 22:229-248.

Biederman I (1987) Recognition by components: a theory of human image understanding. Psychol Rev 94:115-145.

Binford TO (1971) Visual perception by computer Paper presented at the IEEE Conference on Systems and Control, Miami, FL, December.

Bötzel K, Grüsser O-J (1989) Electric brain potentials evoked by pictures of faces and non-faces: a search for "face specific" EEG potentials. Exp Brain Res 77:349-360.

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, Baddely A (1987) The basis of ¾ view advantage in face recognition. Appl Cognit Psychol 1: 109-120.

Cutzu F, Edelman S (1992) Viewpoint-dependence of response time in object recognition. Rehovot: Weizmann Institute of Science.

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051-2062.

Desimone R, Wessinger M, Thomas L, Schneider W (1990) Attentional control of visual perception: cortical and subcortical mechanisms. Cold Spring Harbor Symp Quant Biol 55:963-971.

Edelman S, Bülthoff H (1990) Viewpoint-specific representations in three-dimensional object recognition. Cambridge, MA: MIT Press.

Foldiak P (1990) Forming sparse representation by local anti-Hebbian learning. Biol Cybern 64:165-170.

Foldiak P (1991) Learning invariance from transformation sequences. Neural Comput 3:194–200.
Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. Nature 36:343–346.
Gross CG (1973) Visual functions of inferotemporal cortex. In: Handbook of sensory physiology, VII/3 (Jung V, ed). Berlin: Springer.
Gross CG, Bender DB, Rocha-Miranda CE (1969) Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science 166:1303–1306.
Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. J Neurophysiol 35:96–111.
Harries MH, Perrett DI (1991) Visual processing of faces in the temporal cortex: physiological evidence for a modular organization and possible anatomical correlates. J Cognit Neurosci 3:9–24
Hasselmo ME, Rolls ET, Baylis GC (1989a) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey Behav Brain Res 32:203–218.
Hasselmo ME, Rolls ET, Baylis GC, Nalwa V (1989b) Object centred encoding by face-selective neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res 75:417–429.
Hoffman DD, Richards WA (1984) Parts of recognition. Cognition 18:65–96.
Humphrey GK (1989) Visual object identification. some effects of image foreshortening and monocular depth cues. In: Computational processes in human vision a interdisciplinary perspective (Pylyshyn Z, ed), pp 429–442.
Humphreys GW (1984) Shape constancy: the effects of changing shape orientation and the effects of changing focal features. Percept Psychophys 36:50–64.
Jeffreys DA, Turkmachi ESA (1992) The vertex-positive scalp potential evoked by faces and objects. Exp Brain Res 91: 340–350.
Jolicoeur P (1985) The time to name disorientated natural objects. Mem Cognit 13:289–303.
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 (1979) The internal representation of solid shape with respect to vision. Biol Cybern 32:211–216.
Komatsu H, Ideura Y (1993) Relationships between color, shape, and pattern selectivities of neurons in the inferior temporal cortex of the monkey. J Neurophysiol 70:677–694.
Lowe DG (1987) Three-dimensional object recognition form single two-dimensional images. Artif Intell 31:355–395.
Marr D (1982) Vision. San Francisco: Freeman.
Marr D, Nishihara HK (1978) Representation and recognition of the spatial organization of three dimensional shapes. Proc R Soc Lond [Biol] 200:269–294.
McMullen PA, Farah MJ (1991) Viewer-centred and object-centred representations in the recognition of naturalistic line drawings. Am Psychol Soc 2:275–277.
Miyashita Y (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature 335:817–820.
Miyashita Y (1990) Associative representation of visual long term memory in the neurons of the primate temporal cortex. In: Vision, memory and the temporal lobe (Iwai E, Mishkin M, eds), pp 75–87. New York: Elsevier.
Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331:68–70.
Oram MW, Perrett DI (1992) Time course of neural responses discriminating different views of the face and head. J Neurophysiol 68:70–84.
Oram MW, Perrett DI, Hietanen JK (1993) Directional tuning of motion-selective cells in the anterior superior temporal polysensory area of the macaque. Exp Brain Res 97:274–294.
Palmer SE, Rosch E, Chase P (1981) Canonical perspective and the perception of objects. In: Attention and performance IX (Long J, Baddely AD, eds). Hillsdale, NJ: Erlbaum.
Perrett DI, Oram MW (1993) Neurophysiology of shape processing. Image Vision Comput 11:317–333.
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 (1985) Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc R Soc Lond [Biol] 223:293–317
Perrett DI, Harries MH, Bevan R, Thomas S, Benson PJ, Mistlin AJ, Chitty AJ, Hietanen J, Ortega JE (1989) Frameworks of analysis for the neural representation of animate objects and actions. J Exp Biol 146:87–114.
Perrett DI, Oram MW, Harries MH, Bevan R, Hietanen JK, Benson PJ, Thomas S (1991) Viewer-centred and object-centred coding of heads in the macaque temporal cortex Exp Brain Res 86:159–173
Perrett DI, Hietanen JK, Oram MW, Benson PJ (1992) Organization and functions of cells responsive to faces in the temporal cortex. Philos Trans R Soc Lond [Biol] 335:23–30
Perrett DI, Oram MW, Wachsmuth E (1993) Understanding minds and expression from facial signals: studies at the brain cell level. IEEE Conf Proc [Suppl] 22:138–139.
Poggio T, Edelman S (1990) A network that learns to recognise three-dimensional objects. Nature 343:263–266.
Porrill J, Pollard SB, Pridmore TP, Bowen JB, Mayhew JEW, Frisby JP (1988) TINA: a 3D vision system for pick and place. Image Vision Comput 6:91–99.
Quinlan P (1988) Evidence for the use of perceptual reference frames in two-dimensional shape recognition. PhD thesis, University of London.
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.
Rumelhart DE, Zipser D (1985) Feature discovery by competitive learning. Cognit Sci 9:75–112.
Schwartz EL, Desimone R, Albright TD, Gross CG (1983) Shape recognition and inferior temporal neurons. Proc Natl Acad Sci USA 80:5776–5778.
Seibert M, Waxman AM (1991) Learning aspect graph representations from view sequences. In: Advances in neural network information processing systems (Touretzky DS, ed), pp 258–265. San Mateo, CA: Kaufman.
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.
Shepard RN, Cooper LA (1982) Mental images and their transformations. Cambridge, MA: MIT Press.
Snedecor GW, Cochran WG (1980) Statistical methods, 7th ed. Ames: Iowa State University.
Sobotka S, Ringo JL (1993) Investigation of long term recognition and association memory in unit responses from inferotemporal cortex. Exp Brain Res 96:28–38.
Tanaka K, Fujita M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol 66:170–189.
Tarr MJ, Pinker S (1989) Mental rotation and orientation dependence in shape recognition. Cognit Psychol 5:233–282.
Ullman S (1989) Aligning pictorial descriptions: an approach to object recognition. Cognition 32:193–254.

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: Analysis of visual behaviour (Ingle DJ, Goodale MA, Mansfield RJW, eds), pp 549–585, Cambridge, MA: MIT Press

Verfaillie K (1992) Variant points of view on viewpoint invariance. Can J Psychol 46:215–236.

Wachsmuth E, Oram MW, Perrett DI (1993) Response of single units in the anterior superior temporal polysensory area (STPa) of the macaque to objects and their components. Perception [Suppl] 22:138–139.

Warrington EK, James M (1986) Visual object recognition in patients with right hemisphere lesions: axes or features? Perception 15:355–366.

Warrington EK, Taylor AM (1973) The contribution of the right parietal lobe to object recognition. Cortex 9:152–164.