Martinez-Conde, Macknik, Martinez, Alonso & Tse (Eds.)
Progress in Brain Research, Vol. 154
ISSN 0079-6123

CHAPTER 7

The sensitivity of primate STS neurons to walking sequences and to the degree of articulation in static images

Nick E. Barraclough[1,2], Dengke Xiao[1], Mike W. Oram[1] and David I. Perrett[1,*]

[1]*School of Psychology, St. Mary's College, University of St. Andrews, South Street, St. Andrews, Fife KY16 9JP, UK*
[2]*Department of Psychology, University of Hull, Hull HU6 7RX, UK*

Abstract: We readily use the form of human figures to determine if they are moving. Human figures that have arms and legs outstretched (articulated) appear to be moving more than figures where the arms and legs are near the body (standing). We tested whether neurons in the macaque monkey superior temporal sulcus (STS), a region known to be involved in processing social stimuli, were sensitive to the degree of articulation of a static human figure. Additionally, we tested sensitivity to the same stimuli within forward and backward walking sequences. We found that 57% of cells that responded to the static image of a human figure was also sensitive to the degree of articulation of the figure. Some cells displayed selective responses for articulated postures, while others (in equal numbers) displayed selective responses for standing postures. Cells selective for static images of articulated figures were more likely to respond to movies of walking forwards than walking backwards. Cells selective for static images of standing figures were more likely to respond to movies of walking backwards than forwards. An association between form sensitivity and walking sensitivity could be consistent with an interpretation that cell responses to articulated figures act as an implied motion signal.

Keywords: motion; implied motion; form; integration; temporal cortex; action

Introduction

Artists use many tricks to convey information about movement. One method commonly used is to illustrate a person with legs and arms outstretched or articulated as if the artist had captured a snapshot of the person mid-stride during walking or running. When we see such static images we commonly interpret the human as moving, walking or running forwards through the scene. Although no real movement occurs, the articulated human figure 'implies' movement forward by its configuration or form. There is considerable evolutionary advantage in this ability to infer information about movement from the posture; we can interpret movement direction and speed from a momentary glimpse of a figure.

Traditionally, form and motion information have been thought to be processed along anatomically separate pathways; relatively little effort has been spent investigating how the pathways interact and how motion and form are integrated. Recently, however, three fMRI studies have shown that the brain structure that processes motion, hMT+/V5 (Zeki et al., 1991; Watson et al., 1993; Tootell et al., 1995), is more active to images implying motion when compared to similar images

*Corresponding author. Tel.: +44-1334-463044;
Fax: +44-1334-463042; E-mail: dp@st-andrews.ac.uk

DOI: 10.1016/S0079-6123(06)54007-5

where motion in not implied (Kourtzi and Kanwisher, 2000; Senior et al., 2000; Krekelberg et al., 2005). In each study very different images were used to imply motion; Kourtzi and Kanwisher used images of athletes and animals in action, Senior et al. used images of moving objects and Krekelberg et al. used 'glass patterns', i.e., arrangements of dots suggesting a path of motion. These papers all argue that information regarding the form of static images is made available to hMT+/V5 for coding motion.

Neurons in the monkey homologue of human hMT+/V5, the medial temporal (MT) and medial superior temporal (MST) areas, also respond to glass patterns, where motion is implied (Krekelberg et al., 2003). Areas MT and MST contain neurons that respond to motion (Dubner and Zeki, 1971; Desimone and Ungerleider, 1986) and respond in correlation with the monkey's perception of motion (Newsome et al., 1986; Newsome and Pare, 1988). Neurons in MT/MST area respond maximally to movement in one direction; Krekelberg et al. (2003) showed that they respond preferentially to both real dot motion and implied motion in the preferred direction. Presentation of contradictory implied motion and real motion results in a compromised MT/MST neural response and compromises the monkey's perception of coherent movement.

The blood-oxygen level-dependent (BOLD) activity seen in human hMT+/V5 to complex images implying motion (Kourtzi and Kanwisher, 2000; Senior et al., 2000) could be explained by input from other regions of the cortex. Measurement of event-related potentials (ERP) responses from a dipole pair in the occipital lobe, consistent with localization to hMT+/V5, showed that the responses to the real motion of a random-dot field were 100 ms earlier than responses to static images containing human figures implying motion (Lorteije et al., 2006). The delay in the implied motion response indicates that this information arrives via a different and longer pathway. Kourtzi and Kanwisher (2000) concluded that since inferring information about still images depends upon categorization and knowledge, this must be analysed elsewhere. The activation of hMT+/V5 by implied motion of body images could be due to top-down influences. Senior et al. (2000) suggested that the activation they saw in hMT+/V5 is more likely due to processing of the form of the image in temporal cortex without the need for engagement of conceptual knowledge. At present, there is no evidence that cells in monkey MT are sensitive to articulated human figures implying motion despite active search (Jeanette Lorteije, personal communication).

Information about body posture and articulation in a human figure is likely to come from regions of the cortex that contain neurons sensitive to body form. The superior temporal sulcus (STS) in monkeys and the superior temporal gyrus (STG) and nearby cortex in humans is widely believed to be responsible for processing socially important information. Monkey STS contains neurons that respond to movement of human bodies (Bruce et al., 1981; Perrett et al., 1985), the form (view) of human bodies (Wachsmuth et al., 1994) and many appear to integrate motion and form to code walking direction (Oram and Perrett, 1996; Jellema et al., 2004). It is not known, however, if cells exist that are sensitive to the pattern of articulation that may differentiate postures associated with motion from those associated with standing still.

Giese and Poggio (2003) extended models of object recognition (Riesenhuber and Poggio, 1999, 2002) to generate a plausible feed-forward model of biological motion recognition. A critical postulate of Giese and Poggio's model is the existence of 'snapshot' neurons, neurons tuned to differing degrees of articulation of bodies. Giese and Poggio suggest that these neurons should be found in inferotemporal (IT) or STS cortex, and would feed-forward to neurons coding specific motion patterns, e.g., walking (Oram and Perrett, 1996; Jellema et al., 2004).

In this study we set out to investigate if neurons in temporal cortex can code the degree of articulation of a human figure. Video taping a person walking or running produces a series of stills capturing discrete moments in time. Some of these stills show the person in an articulated pose, others in less-articulated poses akin to standing still. We made use of such video footage in order to compare the responses of STS neurons to

a human figure articulated and standing. Neurons in STS sensitive to non-walking articulated postures are also sensitive to actions leading to such postures (Jellema and Perrett, 2003). It is possible, however, to arrive at a posture from two different directions, by walking forwards, or by walking backwards, both movement directions are consistent with the same static form. We therefore used the video footage played forwards and backwards to investigate how form sensitivity was related to walking.

Following Giese and Poggio (2003) we hypothesized that STS neurons would discriminate articulated postures from standing postures. We also hypothesized that the ability to differentiate posture in static images would relate to sensitivity to motion type for the same neurons. To this end we explore the cells' sensitivity to images of static figures taken from video and movies containing the same images, played forward and in reverse. We also investigate the sensitivity to body view since cells sensitive to static and moving bodies exhibit viewpoint sensitivity (Perrett et al., 1991; Oram and Perrett, 1996).

Methods

Physiological subjects, recording and reconstruction techniques

One rhesus macaque, aged 9 years, was trained to sit in a primate chair with head restraint. Using standard techniques (Perrett et al., 1985), recording chambers were implanted over both hemispheres to enable electrode penetrations to reach the STS. Cells were recorded using tungsten microelectrodes inserted through the dura mater. The subject's eye position ($\pm 1^{\circ}$) was monitored (IView, SMI, Germany). A Pentium IV PC with a Cambridge electronics CED 1401 interface running Spike 2 recorded eye position, spike arrival and stimulus on/offset times.

After each electrode penetration, X-ray photographs were taken coronally and para-sagitally. The positions of the tip of each electrode and its trajectory were measured with respect to the intra-aural plane and the skull's midline. Using the distance of each recorded neuron along the penetration, a three-dimensional map of the position of the recorded cells was calculated. Coronal sections were taken at 1 mm intervals over the anterior–posterior extent of the recorded neurons. Alignment of sections with the X-ray co-ordinates of the recording sites was achieved using the location of microlesions and injection markers on the sections.

Stimuli and presentation

Stimuli consisted of four (16 bit colour) movies of a human walking and four images of the human in different poses. One movie (4326 ms duration) was made by filming (Panasonic, NV-DX110, 3CCD digital video camera) a human walking to the right across a room (walk right). Each individual frame of the movie was flipped horizontally to create a second movie of the human walking to the left (walk left). The frames of both of these movies were arranged in the reverse order to create two movies, one of the human walking to the right backwards (walk right backwards) and the second to a human walking to the left backwards (walk left backwards). There were thus two movies of compatible or forward walking (walk right, walk left) and two movies of incompatible or backward walking (walk right backwards, walk left backwards); two of these movies contained movement in the rightwards direction (walk right, walk right backwards) and two contained movement in the leftwards direction (walk left, walk left backwards).

Two frames from the walk right movie were selected, one when the human was in an articulated pose with legs and arms away from the body (articulated right) and one when the human appeared to be standing with legs and arms arranged vertically (standing right). In both frames the human was in the centre of the room and the time between the two poses was not more than 210 ms. Both frames were flipped horizontally to create two more images (articulated left and standing left). There were thus two images of an articulated human pose (articulated left, articulated right) and two images of a standing pose (standing left,

standing right); two images contained a view of a human facing right (articulated right, standing right) and two images contained a view of a human facing left (articulated right, standing right).

Stimuli were stored on an Indigo2 Silicon Graphics workstation hard disk and presented centrally subtending $25^\circ \times 20.5^\circ$ on a black monitor screen (Sony GDM-20D11, resolution 25.7 pixels/deg, refresh rate 72 Hz), 57 cm from the subject. Movies were presented by rendering each frame of the movie on the screen in sequence, where each frame was presented for 42 ms. Occasionally, movies were presented in a shortened form (duration 1092 ms), where the earlier and later frames were removed from the sequence to show the human walking only across the centre of the room.

Testing procedure

Responses were isolated using standard techniques, and visualized using oscilloscopes. Responses were defined as arising from either single units or multiple units. Both are referred to hereafter as 'cells', 44% was multiple units. Pre-testing was performed with a search set of (on average 55) static images and movies of different objects, bodies and body parts previously shown to activate neurons in the STS (Foldiak et al., 2003; Barraclough et al., 2005). Within this search set were the four different movies of a human walking and four different static images of human forms. Initially, this screening set was used to test each cell with the images and movies presented in a pseudorandom sequence with a 500 ms inter-stimulus interval, where no stimulus was presented for the $n+1$ time until all had been presented n times. Presentation commenced when the subject fixated within $\pm 3^\circ$ of a yellow dot presented centrally on the screen for 500 ms. To allow for blinking, deviations outside the fixation window lasting <100 ms were ignored. Fixation was rewarded with the delivery of fruit juice. Spikes were recorded during the period of fixation, if the subject looked away for longer than 100 ms, spike recording and presentation of stimuli stopped until the subject resumed fixation for >500 ms. Responses to each stimulus in the screening set were displayed as online rastergrams and post-stimulus time histograms (PSTHs) aligned to stimulus onset. If after 4–6 trials the cell gave a substantial response to one of the four walking stimuli or four static human images as determined by observing the online PSTHs, the additional images and movies were removed and testing resumed. From this point, cell responses were saved to a hard disk for offline analysis.

Cell response analysis

Offline isolation of cells was performed using a template-matching procedure and principal components analysis (Spike2, CED, Cambridge, UK). Each cell's response to a stimulus in the experimental test set was calculated by aligning segments (duration > stimulus duration) in the continuous recording, on each occurrence of that particular stimulus (trials).

For each stimulus a PSTH was generated and a spike density function (SDF) calculated by summing across trials (bin size = 1 ms) and smoothing (Gaussian, $\sigma = 10$ ms). Background spontaneous activity (SA) was measured in the 250 ms period prior to stimulus onset. Response latencies to each stimulus were measured as the first 1 ms time bin, where the SDF exceeded 3 SD above the spontaneous activity for over 25 ms in the period following stimulus onset (Oram and Perrett, 1992; Edwards et al., 2003).

The response to each static image was measured within a 250 ms window starting at the stimulus response latency. The response to each walking movie was measured within a 500 ms window starting at the stimulus response latency. Subsequent analysis was performed if the cell's response to one of the stimuli was significantly (3 SD) above the spontaneous background activity.

For each cell showing a significant visual response, the responses to the static images were entered into a 2-way ANOVA [articulation (articulated, standing) by view (left, right) with trials as replicates]. Cells that showed a significant main effect of articulation ($p<0.05$) or a significant interaction between articulation and view (PLSD post-hoc test, $p<0.05$) were classified as sensitive

to articulation. Cells that showed a significant main effect of view ($p<0.05$) or a significant interaction between articulation and view (PLSD post-hoc test, $p<0.05$) were classified as sensitive to view.

The responses to the walking stimuli were entered into a separate 2-way ANOVA [compatibility (forwards, backwards) by direction (left, right) with trials as replicates]. Cells that showed a significant main effect of compatibility ($p<0.05$) or a significant interaction between compatibility and direction (PLSD post-hoc test, $p<0.05$) were classified as sensitive to compatibility. Cells that showed a significant main effect of direction ($p<0.05$) or a significant interaction between compatibility and direction (PLSD post-hoc test, $p<0.05$) were classified as sensitive to direction.

We were also interested in the responses to the articulated human form as it occurred within the walking sequences; this was achieved by measuring responses within a 500 ms window centred around the point in time the articulated form frame occurred within each walking movie. The responses to the walking stimuli, measured in this fashion, were entered into a separate 2-way ANOVA [compatibility (forwards, backwards) by direction (left, right) with trials as replicates]. Cell responses were analysed and cells classified in an analogous fashion.

Results

We tested 55 cells that responded significantly (see methods) to either static images of a human figure or movies of a human walking.

Form-sensitive cells

Thirty-five of the 55 tested cells (64%) showed a significant response to at least one image of a human figure. The sensitivity to different images of a human figure was tested for each cell using ANOVA [articulation: (articulated or standing) view: (left or right)]. Twenty (57%) of the 35 cells responding to images of humans were sensitive to the degree of articulation of the human figure (see methods). Ten out of the twenty cells (50%) responded significantly more to articulated human figures than standing human figures, while the remaining ten (50%) responded significantly more to standing human figures than articulated human figures. The mean response latency to the most effective stimulus for cells responding more to articulated figures was 83 ms (SEM 4.9 ms), significantly (t-test, $t_{[18]} = 2.11$, and $p = 0.049$) less than the mean response latency for cells responding more to standing figures, 111 ms (SEM 12.3 ms). For all other cells that did not differentiate between articulated and standing figures the mean response latency was 93 ms (SEM 13.7 ms). Figure 1 shows the responses of a multiple unit and a single unit sensitive to the degree of articulation of a human figure.

The middle row in Fig. 1 shows the responses of a multiple unit that responds significantly more to the human figure in an articulated pose than a standing pose. The mean response of the multiple unit to the articulated human facing right was 113.6 spikes/s, this decreased to 56.8 spikes/s when the human was standing facing right. The bottom row in Fig. 1 shows the response of a single unit that is also sensitive to the degree of articulation of a human figure, but prefers standing poses. The mean response to the articulated human facing right was 80.4 spikes/s, this increased to 107.6 spikes/s when the human was in a standing pose.

Of the 35 cells responding to a human form, 14 (40%) were sensitive to the direction the human was facing (view) (see methods). Seven out of 14 (50%) of these view-sensitive cells were also sensitive to the degree of articulation in the human figure. Figure 2 shows the responses of an example of a single unit that prefers articulated figures facing to the right (30.3 spikes/s). The responses to the articulated human facing left (19.8 spikes/s), standing human facing right (7.8 spikes/s) or standing human facing left (6.8 spikes/s) are significantly smaller.

Relationship between sensitivity to articulation and motion during walking

To test how the sensitivity to static images of human figures was related to sensitivity to movies of

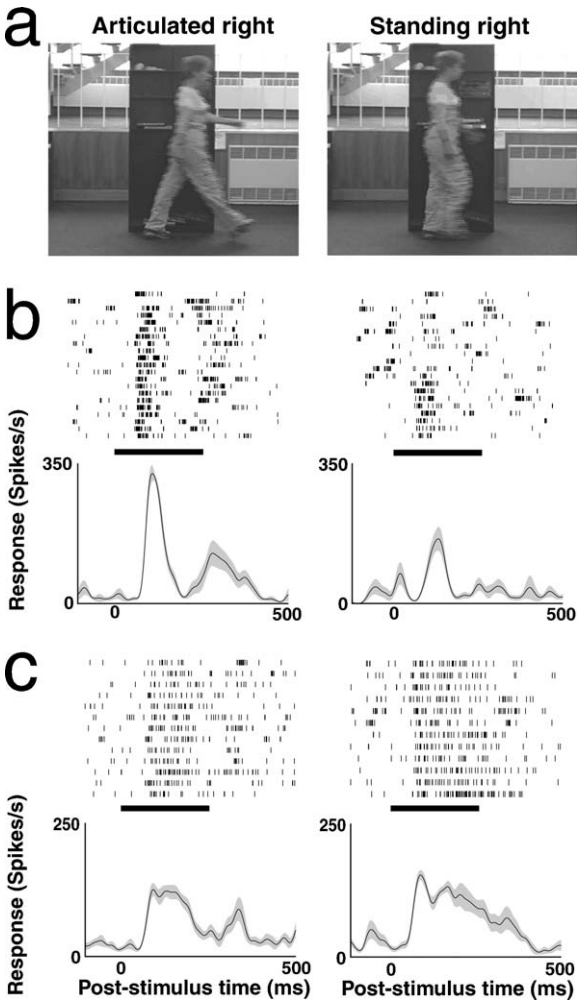

Fig. 1. Responses to human figures. (a) Grey-scale representations of the static visual images used to test responses. (b, c) Plots of responses of a multiple unit and a single unit to the images illustrated. The upper section of each plot shows individual trial responses as rastergrams, the lower section the spike density functions (SDFs) calculated from all trials (grey-SEM) and the black bar in between indicates the onset and duration of the stimulus. (b) Responses of a multiple unit to the image of an articulated body (trials = 21) and a standing body (trials = 20). The responses to articulated human forms were greater than the responses to standing human forms (ANOVA: articulation $F_{[1,77]} = 21.79$, $p < 0.0001$). (c) Responses of a single unit to the image of an articulated body (trials = 13) and a standing body (trials = 12). The response to the standing human form was greater than the response to the articulated human form (ANOVA: interaction articulation × view $F_{[1,45]} = 7.629$, $p = 0.0083$ and PLSD post-hoc test, $p < 0.05$).

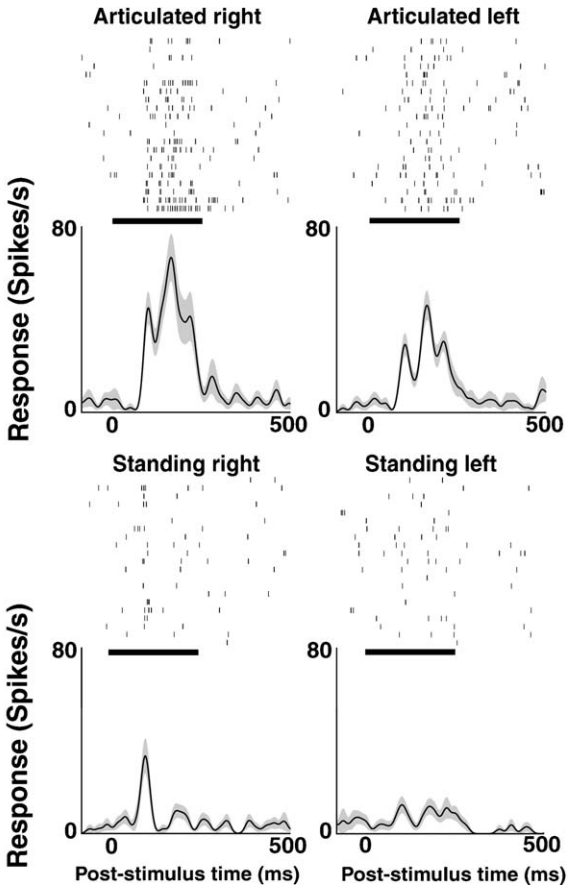

Fig. 2. Single cell responses to human figures articulated and standing, facing right and left. Rastergrams and SDFs plotted as in Fig. 1. The top row shows the responses to the articulated human figure facing right (trials = 21) and facing left (trials = 21). The bottom row shows the responses to the standing human figure facing right (trials = 21) and facing left (trials = 21). The responses to the articulated figures are greater than the responses to the standing figures, and the responses to the figure facing right are greater than the responses to the figure facing left (ANOVA: articulation $F_{[1,80]} = 40.214$, $p < 0.0001$, direction $F_{[1,80]} = 4.18$ and $p = 0.04$).

walking, the sensitivity to different walking movies was calculated for all 35 cells showing a response to a static image of a human figure. We were interested in the responses to the key frames with articulated and non-articulated human figures as they occurred within a walking sequence; this was achieved by centring the 500 ms response measurement window around the point in time the articulated form frame occurred within each

Table 1. Test of sensitivity of cells to different movies of a human walking using ANOVA

	Preferred walking			
Preferred static figure	Compatible	Incompatible	Non-discriminative	Total
Articulated	5	1	4	10
Standing	0	6	4	10
Non-discriminative	2	3	10	15
Total	7	10	18	35

Association between cell sensitivity to body articulation in static images and sensitivity to type of walking movement. Compatible = body facing the direction of walking; incompatible = body facing away from the direction of walking. Cell tuning for the degree of articulation of static body images significantly predicted cell tuning for compatibility of walking.

movie. The sensitivity to different movies of a human walking [compatibility: (forwards or backwards) by direction: (left or right)] was tested for each cell using ANOVA (Table 1). Of the 20/35 cells sensitive to the degree of articulation in the static human figure, 12 (60%) were also sensitive to the compatibility of movement in the movies of the human walking (see methods). For the 10 cells that responded more to articulated human figures in static images, five responded significantly more to forward walking, one responded significantly more to backward walking and four were not sensitive to the compatibility of walking. For the 10 cells that responded more to standing human figures in static images, none responded significantly more to forward walking, six responded significantly more to backward walking and four were not sensitive to the compatibility of walking. For the 15/35 cells that were not sensitive to the degree of articulation in the human figures, two responded significantly more to forward walking, three responded significantly more to backward walking and 10 were not sensitive to the compatibility of walking. In the cells showing sensitivity to the degree of articulation in static images there was an association between the preferred degree of articulation for static images and the preferred compatibility of a walking human (Pearson $\chi^2 = 8.571$, $df = 1$, $p = 0.003$ and Fisher's exact test, $p = 0.015$).

Figure 3 illustrates the responses of a single unit to static images and walking movies. The cell responds significantly more to a static image of an articulated human figure facing to the right (25.6 spikes/s) than to a static image of a standing human figure facing right (16.6 spikes/s). The cell also responds significantly more to a movie of walking forwards to the right (26.8 spikes/s) than walking backwards to the right (14.0 spikes/s). The frame containing the articulated human form occurs within the walking forwards movie at 1092 ms and the walking backwards movie at 1218 ms. An increase in the response to the walking forwards movie can be seen both at the start of the movie and at the time of the occurrence of the articulated frame. No similar increase in the response to the walking backward movie can be seen at the start or at the time when the articulated frame occurs.

Figure 4 illustrates an equivalent relationship between responses to standing human figures and walking backwards movies. The cell responds more to a static image of standing human figure facing to the left (23.7 spikes/s) than a static image of an articulated human figure facing left (11.5 spikes/s). The cell also responds more to a movie of walking backwards to the right (22.2 spikes/s) than to a movie of walking forwards to the left (15.7 spikes/s). The frame containing the articulated human form (in this case, to the left) occurs within both the walking movies (walking to the left after 630 ms and walking backwards to the right after 714 ms).

Illustrated in Fig. 4b are the responses of the cell to the static images and walking movies in the other directions. The left-hand column shows the responses to a standing human form facing right (2.4 spikes/s) and to an articulated form facing right (5.0 spikes/s), the right column shows the responses to walking backwards to the left (5.6 spikes/s) and walking forwards to the right

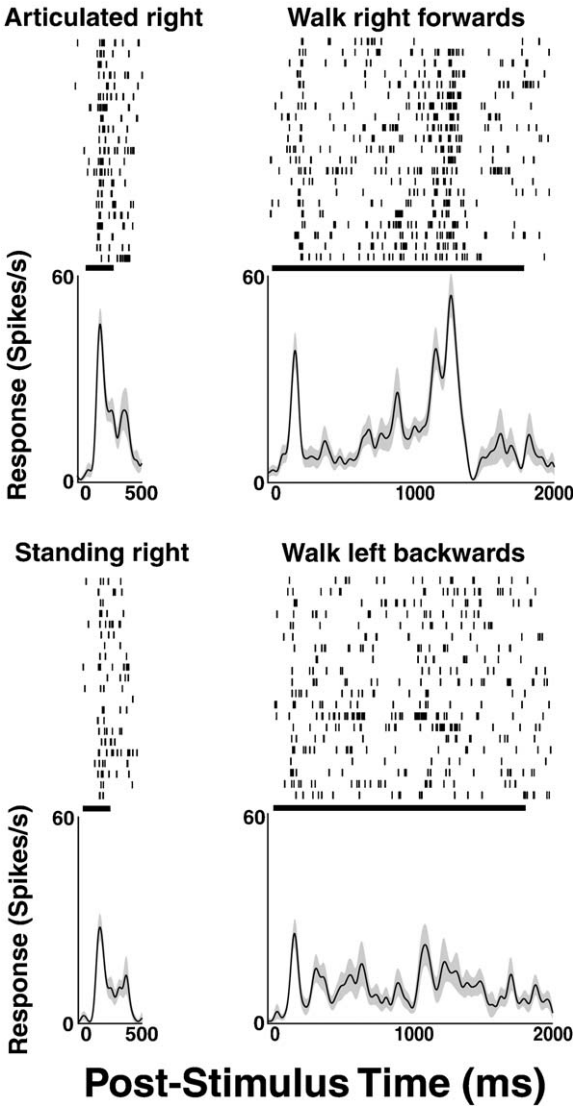

Fig. 3. Responses of a single STS cell that prefers articulated human figures and walking forwards. Rastergrams and SDFs are plotted as in Fig. 1. The left column shows example of responses to images. Top: responses to an articulated human facing right (trials = 21) and bottom: responses to a standing human facing right (trials = 21). The cell responds more to articulated human figures than to standing human figures (ANOVA: articulation, $F_{[1,80]} = 7.472$ and $p = 0.0077$). The right column shows example of responses to movies of walking. Top: responses to walking right forwards (trials = 21) and bottom: responses to walking left backwards (trials = 20). The cell prefers compatible walking to incompatible walking (ANOVA: compatibility, $F_{[1,78]} = 4.365$, $p = 0.039$, interaction compatibility × direction, $F_{[1,78]} = 20.982$, $p < 0.0001$, PLSD post-hoc test, $p < 0.05$).

(3.2 spikes/s). The cell responds significantly more to the static image of a human standing facing left than any other static image and responds significantly more to the movie of a human walking backwards to the right than any other movie. The illustrated cell is thus additionally sensitive to the view of the human figure in the static images and the walking direction in the walking movies.

Cells sensitive to walking

Having established that a cell's sensitivity to the degree of articulation in a static image of a human form can predict the cell's sensitivity to the compatibility of walking stimuli, we wanted to know if the inverse was true. Could knowing the sensitivity to the compatibility of a walking human be used to predict sensitivity to static images of human figures? We measured the responses to the four walking stimuli in a 500 ms window starting at the stimulus response latencies.

Fifty-two of 55 tested cells (95%) showed a significant response to at least one movie of a human walking. The sensitivity to different walking directions [compatibility: (forwards or backwards), direction: (left or right)] was tested in each cell using ANOVA (Table 2). For 19 (36%) of the 52 cells the compatibility of walking was a significant factor (see methods). Seven out of the 19 cells (37%) responded more to walking forwards and 12 out of 19 cells (63%) responded more to walking backwards.

In the 19 cells sensitive to the compatibility of walking, 12 (63%) were also sensitive to the degree of articulation in the static image of the human figure. For the seven cells that responded more to walking forwards, four responded significantly more to articulated figures, none responded significantly more to standing figures in static images and three were insensitive to the degree of articulation of the static human figure. For the 12 cells that responded more to walking backwards, one responded significantly more to articulated figures, four responded significantly more to standing figures in static images and seven were insensitive to the degree of articulation of the

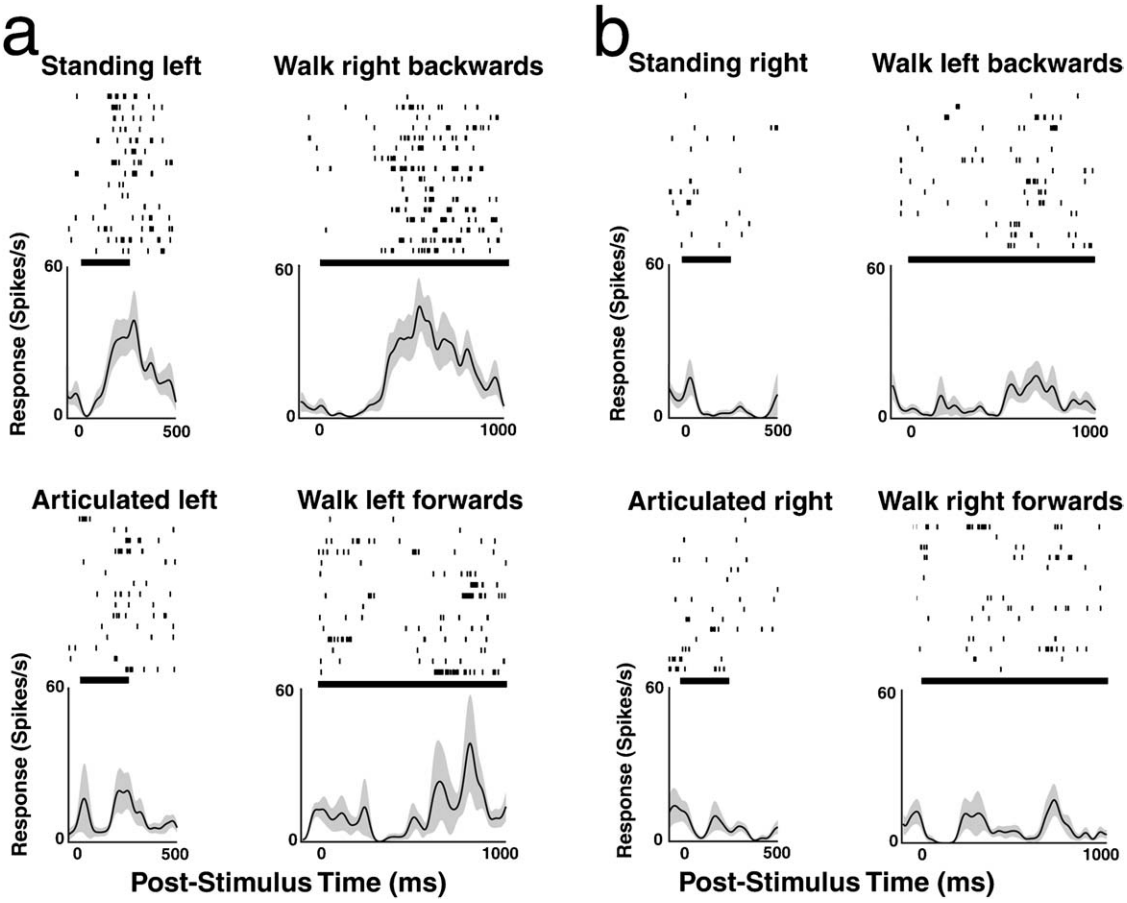

Fig. 4. Responses of a single STS cell to static images and movies of a human walking. Rastergrams and SDFs plotted as in Fig. 1. (a) The left column shows responses to static images, top: responses to a standing human facing left (trials = 15) and bottom: responses to articulated human facing left (trials = 15). The cell responds more to standing figures than articulated figures (ANOVA: articulation, $F_{[1,57]} = 5.030$, $p = 0.0288$, interaction articulation × view, $F_{[1,57]} = 10.950$, $p = 0.0016$, PLSD post-hoc test, $p < 0.05$). The right column shows the responses to movies of walking, top: responses to walking right backwards (trials = 16) and bottom: responses to walking left forwards (trials = 15). The cell responds more to incompatible walking to the right more than compatible walking to the left (ANOVA: interaction compatibility × direction, $F_{[1,57]} = 19.772$, $p < 0.00001$, PLSD post-hoc test, $p < 0.05$). (b) The left column shows responses to static images, top: responses to a standing human facing right (trials = 15) and bottom: responses to an articulated human facing right (trials = 16). The cell responds less to figures facing right than to the left (ANOVA: view, $F_{[1,57]} = 39.620$, $p < 0.001$), and responds more to a standing figure facing left than any other static image (ANOVA: interaction articulation × view, $F_{[1,57]} = 10.950$, $p = 0.0016$, PLSD post-hoc test, $p < 0.05$ each comparison). The right column shows the responses to movies of walking, top: responses to walking left backwards (trials = 15) and bottom: responses to walking right forwards (trials = 15). The cell responds more to incompatible walking to the right than any other walking stimulus (ANOVA: interaction compatibility × direction, $F_{[1,57]} = 19.772$, $p < 0.0001$, PLSD post-hoc test, $p < 0.05$ for each comparison).

static human figure. For the 33/52 cells that were not sensitive to the compatibility of the walking stimuli, 4 responded significantly more to articulated human figures, 5 responded significantly more to standing human figures and 24 were insensitive to the degree of articulation of the static human figure. In the cells showing sensitivity to the compatibility of walking, there was an association between the preferred compatibility and the preferred degree of articulation of a human figure (Pearson $\chi^2 = 5.76$, df = 1 and $p = 0.016$) (Fisher's exact test, $p = 0.048$).

Table 2. Test of sensitivity of cells to different walking directions using ANOVA

	Preferred static figure			
Preferred walking stimulus	Articulated	Standing	Non-discriminative	Total
Compatible	4	0	3	7
Incompatible	1	4	7	12
Non-discriminative	4	5	24	33
Total	9	9	34	52

Association between cell sensitivity to type of walking movement and degree of body articulation in static images. Compatible = body facing the direction of walking; incompatible = body facing away from the direction of walking. Cell tuning for compatibility of walking significantly predicted cell tuning in the degree of articulation of static body images.

Histological localization

Cells showing sensitivity to static images of human forms and movies of a human walking were found in the target area of the upper bank, lower bank and fundus of rostral STS and inferotemporal cortex. As defined in previous studies (Desimone and Gross, 1979; Bruce et al., 1981, 1986; Baylis et al., 1987; Hikosaka et al., 1988; Distler et al., 1993; Seltzer and Pandya, 1994; Saleem et al., 2000), rostral STS is the region of cortex in the upper bank (TAa, TPO), lower bank (TEa, TEm) and fundus (PGa, IPa) of the STS that lies rostral to the fundus of the superior temporal area (FST). The anterior–posterior extent of the recorded cells was from 6.9 to 10.5 mm posterior of the anterior commissure. We saw no apparent concentration of cells showing sensitivity to one figure or walking type within sub-regions of the STS. Figure 5 shows the position of all neurons that responded to at least one of the stimuli tested.

Discussion

The results of this study show two main findings: (1) Fifty-seven per cent of STS neurons that respond to static images of a human figure are sensitive to the degree of articulation of the figure itself. (2) There is an association between STS neuronal response sensitivity to the degree of articulation of a human figure and sensitivity to the compatibility between the direction of locomotion and view of the body of a human walking. For the cells that were sensitive to both the degree of articulation of a human figure in a static image and compatibility of walking: cells that 'preferred' articulated human figures 'preferred' compatible walking (walking forwards), cells that 'preferred' standing human figures 'preferred' incompatible walking (walking backwards).

STS neurons have been known to be sensitive to the form of faces (Bruce et al., 1981; Perrett et al., 1982), body parts (Perrett et al., 1989) and whole bodies (Wachsmuth et al., 1994) for a significant time. The cells sensitive to the articulation of a human figure described here are a novel subset of STS neurons that code the form of whole bodies. A previous study of STS cell responses to whole bodies described cells that were sensitive to the left or right view of the body (Wachsmuth et al., 1994). In this study many cells selectively responded to one articulation type irrespective of the view of the figure, others, however, were additionally sensitive to the view of the body.

The model of Giese and Poggio (2003) describing a feed-forward model of motion recognition was reliant upon the existence of a subset of neurons in IT or STS responding selectively to different 'snapshots' of a human walking. The neurons we describe here might potentially represent snapshots. Giese and Poggio's model, however, used 21 types of snapshot neuron, tuned to 21 different degrees of articulation of the human figure. We demonstrate only two types of snapshots, articulated and standing, we did not investigate intermediate poses. Further fine-grained analysis of the number of prototypical poses would be needed to reveal full supporting physiological evidence for Giese and Poggio's model. The number of

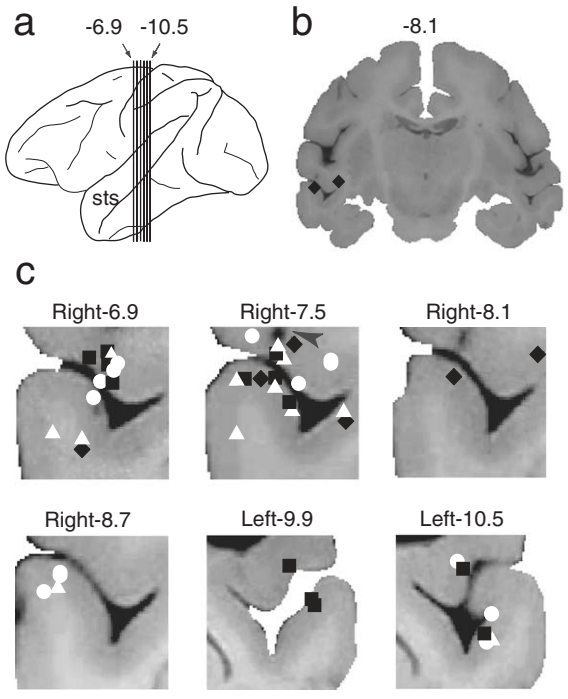

Fig. 5. Location of cells responding to walking movies or images of human poses. (a) Positions along the STS illustrated on a schematic representation of the brain of the six sections shown in (c). (b) Photograph of the section at −8.1 mm (posterior from the anterior commissure — ac). (c) Photographs of six sections (−6.9, −7.5, −8.1, −8.7, −9.9, −10.5) cropped and enlarged to illustrate the right and left STSs where cells were recorded. In order to illustrate the grey matter–white matter boundary, the contrasts of the photographs have been enhanced using Adobe Photoshop [The 8 bit contrast range from 91 to 177 was increased to the range from 0 to 255]. All 55 cells responding to at least one visual stimulus are plotted, where: cells sensitive to the articulation of a human figure in static images and compatibility during walking (white circles), cells sensitive to the articulation of a human figure (white triangles), cells sensitive to compatibility during walking (black squares), cells responsive to stimuli but not sensitive to articulation of a human figure or compatibility during walking (black diamonds). The black arrow indicates the position of a marker lesion used during the reconstruction.

prototypical poses could be <21, as acknowledged by Giese and Poggio, and depending upon the tuning to the degree of articulation of the human figure. It is worth noting that the STS neurons code all head views in the horizontal plane, but show a biased distribution for prototypical head views (namely face, left and right profile, and back), and broad tuning — 60% bandwidth (Perrett et al., 1991). Several computational models (Ullman, 1989) use interpolation between key templates to 'recognize' intermediate views. A similar mechanism may interpolate between a small number of key postures.

Neurons sensitive to the degree of articulation of a human figure might be used in coding walking movements of other agents when visual information is degraded. 'Biological motion' stimuli, consisting of points of light attached to a human in the dark, result in a vivid perception of a walking figure where no form information is available to the visual system (Johansson, 1973). The conventional interpretation is that multiple local motion vectors for each point of light are first calculated within a motion processing system before integration to generate a form signal and the perception of a particular type of action. In a recent study, the position of each light dot on the human walking figure was moved to a different point on the figure between each frame of the biological motion movie (Beintema and Lappe, 2002). Perception of the form of the walking human was similar to the conventional biological motion stimulus even though local motion vectors could not be used to generate a global biological motion signal. Beintema and Lappe suggest that templates of different articulated human figures might be used to interpret these biological motion stimuli. Interestingly, some cells recorded in the STS that are sensitive to biological motion stimuli also respond entirely to static body views (unpublished observations). It remains to be seen if the cells recorded here are sensitive to biological motion stimuli and thus represent the templates proposed by Beintema and Lappe.

Similar to the finding of Jellema and Perrett (2003) we found an association between sensitivity to static forms and sensitivity to articulating actions. Jellema and Perrett's study, however, used actors moving and posing, thereby creating 3-D articulation between the head and trunk and between the upper and lower body. The association between sensitivity to static forms and articulated actions in this study were shown with video stimuli, in 2-D, and articulation was between the limbs and body in the cyclic action of walking.

This association shows how these cells code walking from images of human figures in the absence of motion information. Forward walking and backward walking, coded by STS neurons (Oram and Perrett, 1994, 1996), are two different meaningful actions. It is important for survival and social interaction to be able to interpret if a predator/prey/friend is approaching or backing away. You would gain a considerable advantage from detecting a predator approaching despite being afforded only a brief glimpse. During walking forwards and walking backwards, however, a figure cycles through the same repertoire of postures, only in a different sequence (the kinematics will be slightly different). It is possible that a glimpse of an articulated posture could arise from a figure walking either forwards or backwards. We however interpret articulated figures as walking forwards (Pavlova et al., 2002). Coding of forward walking in cells that respond significantly to articulated figures is consistent with this bias in our perception.

It is slightly more difficult to interpret the association between the sensitivity for standing postures and walking backwards. One approach is to regard the mutually exclusive postures and behaviours as opposite to each other. A standing figure cannot also be adopting an articulated posture, and we consistently interpret standing figures as not moving (unlike articulated figures). An agent backing away cannot also be approaching, since approaching and retreating behaviours involve opposite directions of motion. Increased responses to backward walking for cells that are selective for standing postures is consistent with this association and might represent an opponent population of cells to those that code forward walking and articulated postures.

Human cortex contains hMT+/V5 and posterior STS/superior temporal gyrus (STG), which are homologous regions to monkey MT and STS, respectively. Assuming the presence of similar cells in both species, the response sensitivity described here could explain activity to images implying motion in human hMT+/V5 (Kourtzi and Kanwisher, 2000), but less activity in STS. Images implying motion would activate STS cells that are tuned to articulated postures, and this information could be relayed to area hMT+/V5. The presence of an equal number of cells tuned to both articulated and standing postures in the STS result in a reduced net activation of this region with the contrast (implied motion — static) used by Kourtzi and Kanwisher to detect a response to implied motion.

One aspect we chose not to investigate here was the relationship between the direction the human figure faced in the static images (view) and the direction of movement during walking. This action was taken as our analysis consisted of repeatedly classifying responses and subdividing into separate cell populations. With an initial population of 55 cells, a further subdivision of cells into groups sensitive to each possible permutation of articulation and view would make each sub-group too small for meaningful statistics. Figure 4, however, illustrates the responses of a cell that could not only code motion (or in this cell's case, absence of motion), but also direction. The cell responded more to the static image of a human figure facing to the left; static images of humans facing to the right were ineffectual. The sensitivity to static images was consistent with the sensitivity to walking stimuli in only one direction. Thus, the responses of this cell can be interpreted as coding 'not moving', or 'not moving to the left'. More extensive studies of larger populations of neurons will enable analyses to determine if the neurons we describe here can signal motion or absence of motion in specific directions from static images.

Acknowledgments

This work was funded by grants from the EU and the Wellcome Trust.

References

Barraclough, N.E., Xiao, D.-K., Oram, M.W. and Perrett, D.I. (2005) Integration of visual and auditory information by STS neurons responsive to the sight of actions. J. Cogn. Neurosci., 17: 377–391.

Baylis, G.C., Rolls, E.T. and Leonard, C.M. (1987) Functional subdivisions of the temporal neocortex. J. Neurosci., 7: 330–342.

Beintema, J.A. and Lappe, M. (2002) Perception of biological motion without local image motion. Proc. Natl. Acad. Sci. USA, 99: 5661–5663.

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

Bruce, C.J., Desimone, R. and Gross, C.G. (1986) Both striate cortex and superior colliculus contribute to visual properties of neurons in superior temporal polysensory area of macaque monkey. J. Neurophysiol., 55: 1057–1075.

Desimone, R. and Gross, C.G. (1979) Visual areas in the temporal cortex of the macaque. Brain Res., 178: 363–380.

Desimone, R. and Ungerleider, L.G. (1986) Multiple visual areas in the caudal superior temporal sulcus of the macaque. J. Comp. Neurol., 248: 164–189.

Distler, C., Boussaoud, D., Desimone, R. and Ungerleider, L.G. (1993) Cortical connections of inferior temporal area TEO in macaque monkeys. J. Comp. Neurol., 334: 125–150.

Dubner, R. and Zeki, S.M. (1971) Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res., 35: 528–532.

Edwards, R., Xiao, D.-K., Keysers, C., Foldiak, P. and Perrett, D.I. (2003) Color sensitivity of cells responsive to complex stimuli in the temporal cortex. J. Neurophysiol., 90: 1245–1256.

Foldiak, P., Xiao, D.-K., Keysers, C., Edwards, R. and Perrett, D.I. (2003) Rapid serial visual presentation for the determination of neural selectivity in area STSa. Prog. Brain Res., 144: 107–116.

Giese, M.A. and Poggio, T. (2003) Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci., 4: 179–192.

Hikosaka, K., Iwai, E., Saito, H. and Tanaka, K. (1988) Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. J. Neurophysiol., 60: 1615–1637.

Jellema, T., Maassen, G. and Perrett, D.I. (2004) Single cell integration of animate form, motion and location in the superior temporal cortex of the macaque monkey. Cereb. Cortex, 14: 781–790.

Jellema, T. and Perrett, D.I. (2003) Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion. Neuropsychologia, 41: 1728–1737.

Johansson, G. (1973) Visual perception of biological motion and a model for its analysis. Percept. Psychophys., 14: 201–211.

Kourtzi, Z. and Kanwisher, N. (2000) Activation in human MT/MST by static images with implied motion. J. Cogn. Neurosci., 12: 48–55.

Krekelberg, B., Dannenberg, S., Hoffmann, K.-P., Bremmer, F. and Ross, J. (2003) Neural correlates of implied motion. Nature, 424: 674–677.

Krekelberg, B., Vatakis, A. and Kourtzi, Z. (2005) Implied motion from form in human visual cortex. J. Neurophysiol., 94(6): 4373–4386.

Lorteije, J.A.M., Kenemans, J.L., Jellema, T., van der Lubbe, R.H.J., de Heer, F. and van Wezel, R.J.A. (2006) Delayed response to animate implied motion in human motion processing areas. J. Cogn. Neurosci., 18: 158–168.

Newsome, W.T., Mikami, A. and Wurtz, R.H. (1986) Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. J. Neurophysiol., 55: 1340–1351.

Newsome, W.T. and Pare, E.B. (1988) A selective impairment of motion perception following lesions of the middle temporal visual area (MT). J. Neurosci., 8: 2201–2211.

Oram, M.W. and Perrett, D.I. (1992) Time course of neural responses discriminating views of the face and head. J. Neurophysiol., 68: 70–84.

Oram, M.W. and Perrett, D.I. (1994) Responses of anterior superior temporal polysensory (STPa) neurons to ''biological motion'' stimuli. J. Cogn. Neurosci., 6: 99–116.

Oram, M.W. and Perrett, D.I. (1996) Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. J. Neurophysiol., 76: 109–129.

Pavlova, M., Krageloh-Mann, I., Birbaumer, N. and Sokolov, A. (2002) Biological motion shown backwards: the apparent-facing effect. Perception, 31: 435–443.

Perrett, D.I., Harries, M.H., Bevan, R., Thomas, S., Benson, P.J., Mistlin, A.J., Chitty, A.J., Hietanen, J.K. and Ortega, J.E. (1989) Frameworks of analysis for the neural representation of animate objects and actions. J. Exp. Biol., 146: 87–113.

Perrett, D.I., Oram, M.W., Harries, M.H., Bevan, R., Hietanen, J.K., Benson, P.J. and Thomas, S. (1991) Viewer-centred and object-centred coding of heads in the macaque temporal cortex. Exp. Brain Res., 86: 159–173.

Perrett, D.I., Rolls, E.T. and Caan, W. (1982) Visual neurons sensitive to faces in the monkey temporal cortex. Exp. Brain Res., 47: 329–342.

Perrett, D.I., Smith, P.A.J., Mistlin, A.J., Chitty, A.J., Head, A.S., Potter, D.D., Broennimann, R., Milner, A.D. and Jeeves, M.A. (1985) Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: a preliminary report. Behav. Brain Res., 16: 153–170.

Riesenhuber, M. and Poggio, T. (1999) Hierarchical models of object recognition in cortex. Nat. Neurosci., 2: 1019–1025.

Riesenhuber, M. and Poggio, T. (2002) Neural mechanisms of object recognition. Curr. Opin. Neurobiol., 12: 162–168.

Saleem, K.S., Suzuki, W., Tanaka, K. and Hashikawa, T. (2000) Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. J. Neurosci., 20: 5083–5101.

Seltzer, B. and Pandya, D.N. (1994) Parietal, temporal and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. J. Comp. Neurol., 15: 445–463.

Senior, C., Barnes, J., Giampietro, V., Simmons, A., Bullmore, E.T., Brammer, M. and David, A.S. (2000) The functional neuroanatomy of implicit-motion perception 'representational momentum'. Curr. Biol., 10: 16–22.

Tootell, R.B., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R. and Belliveau, J.W. (1995) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J. Neurosci., 15: 3215–3230.

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

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

Watson, J.D., Myers, R., Frackowiak, R.S., Hajnal, J.V., Woods, R.P., Mazziotta, J.C., Shipp, S. and Zeki, S. (1993) Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. Cereb. Cortex, 3: 79–94.

Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C. and Frackowiak, R.S. (1991) A direct demonstration of functional specialization in human visual cortex. J. Neurosci., 11: 641–649.