Neuron, Vol. 34, 149–159, March 28, 2002, Copyright ©2002 by Cell Press

Parallel Visual Motion Processing Streams for Manipulable Objects and Human Movements

Michael S. Beauchamp,[1] Kathryn E. Lee,
James V. Haxby, and Alex Martin
Laboratory of Brain and Cognition
National Institute of Mental Health
Bethesda, Maryland 20892

Summary

We tested the hypothesis that different regions of lateral temporal cortex are specialized for processing different types of visual motion by studying the cortical responses to moving gratings and to humans and manipulable objects (tools and utensils) that were either stationary or moving with natural or artificially generated motions. Segregated responses to human and tool stimuli were observed in both ventral and lateral regions of posterior temporal cortex. Relative to ventral cortex, lateral temporal cortex showed a larger response for moving compared with static humans and tools. Superior temporal cortex preferred human motion, and middle temporal gyrus preferred tool motion. A greater response was observed in STS to articulated compared with unarticulated human motion. Specificity for different types of complex motion (in combination with visual form) may be an organizing principle in lateral temporal cortex.

Introduction

Studies in human and nonhuman primates have identified a region in posterior lateral occipital-temporal cortex, named area MT, that responds preferentially to moving visual stimuli. A region anterior and superior to MT in the superior temporal sulcus (STS) has been identified that responds to a specific type of visual motion, namely biological motion. In monkeys, single neurons in STS respond to hand, eye, mouth, or body movements (Oram and Perrett, 1994). In humans, mouth and eye movements activate STS (Bonda et al., 1996; Puce et al., 1998). STS is more active in response to point-light displays containing biological motion than similar displays containing scrambled or inverted motion (Grossman et al., 2000; Grossman and Blake, 2001). STS is also active during viewing of static biological stimuli, such as static faces (Chao et al., 1999a, 1999b; Haxby et al., 1999; Hoffman and Haxby, 2000; Kanwisher et al., 1997) or animals (Chao et al., 1999a, 1999b), suggesting that while STS processes biological motion, it is also automatically engaged whenever biological stimuli are viewed.

Motion is also an important attribute of man-made manipulable objects like tools and utensils, and identification of these objects is also associated with activity in the lateral temporal cortex (Chao et al., 1999a; Martin et al., 1996). However, this activity appears not in STS but inferior to STS in posterior middle temporal gyrus (MTG) extending inferiorly into the inferior temporal sulcus. Static pictures of faces and animals tend to activate STS, and static pictures of tools activate MTG, while static pictures of objects not commonly associated with motion, such as houses and buildings, do not activate lateral temporal cortex (Aguirre et al., 1998; Chao et al., 1999a; Epstein and Kanwisher, 1998; Haxby et al., 1999). An additional piece of evidence comes from the finding that biological or nonbiological static pictures containing implied motion (such as a cup photographed mid-fall) produce greater activation in lateral temporal cortex than pictures without implied motion (such as a cup resting on a table top) (Kourtzi and Kanwisher, 2000; Senior et al., 2000). Taken together, these findings suggest that lateral temporal cortex is involved in perceiving and storing information about motion.

In addition to activity in lateral temporal cortex, ventral occipitotemporal cortex also responds to visual objects (Grill-Spector et al., 1999; Lerner et al., 2001; Malach et al., 1995), and different object categories produce distinct patterns of activity in ventral temporal cortex. Because of its location in the ventral object processing stream, it has been proposed that ventral temporal cortex is tuned to features of object form shared by members of a category (Haxby et al., 2000; Martin et al., 2000). In addition to sensitivity to visual form, lateral temporal cortex might also be tuned to the object motion properties shared by category members (Martin et al., 2000). This argument is based on the known sensitivity of STS to biological motion, and the proximity of lateral temporal cortex to area MT. However, this idea has not been directly tested.

In our first experiment, we tested the hypothesis that lateral temporal cortex is differentially sensitive to the motion properties of different visual categories. We examined the brain response to low- and high-contrast moving gratings, to biological motion consisting of moving human figures, and to object motion consisting of moving manipulable objects (tools) (Figure 1). Three regions in posterior lateral temporal cortex—area MT, STS, and MTG—responded to the motion stimuli, and all three areas preferred human and tool motion to moving gratings. Area MT showed no preference for human or tool motion. Anterior and superior to MT in STS, a larger response was observed for human compared with tool motion. While anterior and inferior to MT in MTG, a larger response was observed for tool motion compared with human motion.

In our second experiment, we sought to demonstrate that the STS and MTG activity was related to the motion contained in the stimuli and did not simply reflect the category-related responses observed for static stimuli described above. Subjects viewed static and moving humans and tools, and we compared the response to each type of stimulus in ventral and lateral temporal cortex. While ventral temporal cortex responded similarly to moving and static stimuli, lateral temporal areas responded much more strongly to moving than static stimuli, supporting the hypothesis that lateral temporal cortex is the cortical locus of complex motion processing.

[1] Correspondence: mbeauchamp@nih.gov

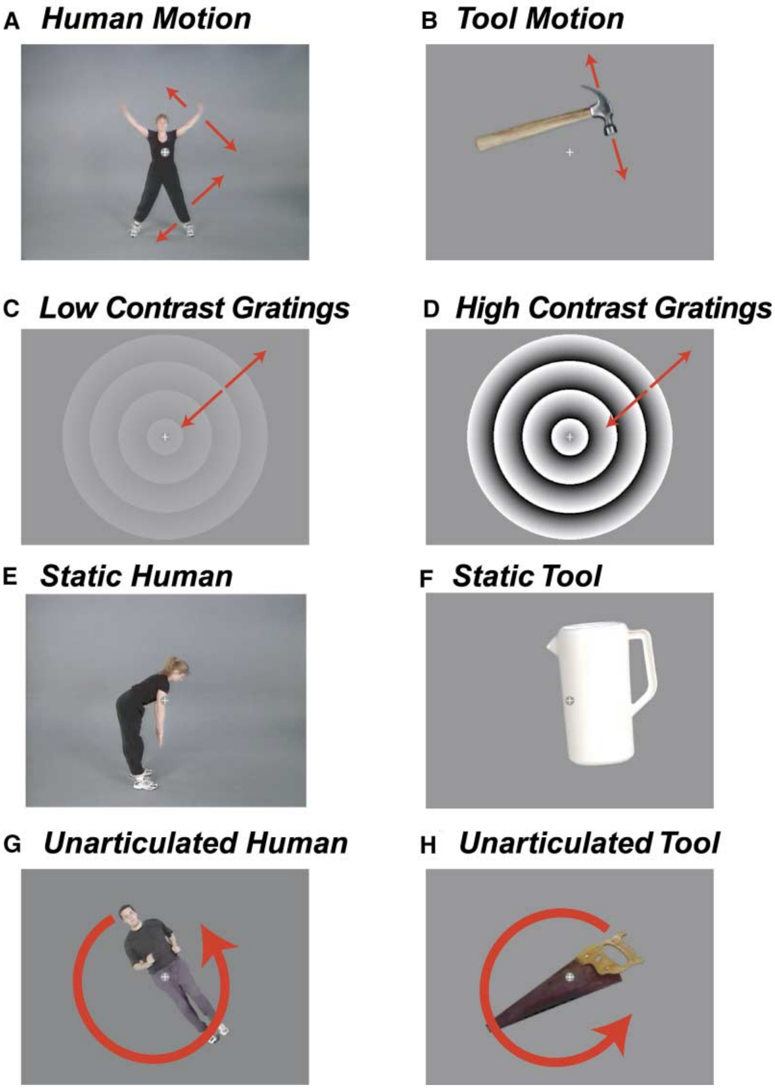


Figure 1. Sample Still Frames from the Eight Types of Visual Stimuli

Red arrows (not present in actual stimuli) illustrate the direction of motion for motion-containing stimuli (all except [E] and [F]). Each stimulus lasted for 2.5 s (continuous motion at 30 frames per second for motion stimuli, unchanging for static stimuli) and contained a small central fixation crosshairs. Human stimuli (A, E, and G) were edited from footage of two actors performing 28 different motions filmed from five viewpoints. Tool stimuli (B, F, and H) were edited from footage of 28 tools, each performing its characteristic motion without visible manipulandum. Radial gratings stimuli (C and D) moved either inwards or outwards during each 2.5 s presentation.

In our third experiment, we explored the visual properties underlying the differential sensitivity of STS and MTG to human and tool motion. When humans move, different body parts typically move with complex motion trajectories connected by articulated joints, while tools typically move with simple motion trajectories and few degrees of articulation. We reasoned that if STS prefers articulated motion, then STS should respond more to humans moving with articulated motion (as during a jumping jack) than to humans moving with unarticulated motion (such as rotating about their center of mass). Similarly, if MTG prefers the unarticulated motion characteristic of tools, it should respond more to unarticulated human motion than to articulated human motion. We constructed stimuli containing humans and tools moving with artificial motion trajectories consisting of simple translation, and rotation with no articulation. STS preferred humans moving with many degrees of articulation to humans moving with unarticulated motion vectors, while MTG responded similarly to natural and artificially generated unarticulated tool motion and showed a trend toward larger responses for unarticulated than articulated human movements. This demonstrates that STS prefers the complex articulated motion characteristic of biological motion and suggests that MTG is selective for the unarticulated motion characteristic of tools. However, as shown in the second experiment, STS prefers static images of humans to static images of tools, and MTG prefers static images of tools to static images of humans. Therefore, a combination of selectivity for motion and form underlies responses in lateral temporal cortex.

Results

Experiment 1: Moving Radial Gratings, Humans, and Tools

As illustrated in Figure 2A, a number of brain regions, primarily in occipitotemporal cortex, responded differently to moving radial gratings and moving humans and tools. Medial occipital areas strongly preferred moving gratings, with a peak preference in calcarine cortex (coordinates for all foci in Table 1). In both hemispheres, lateral occipitotemporal cortex showed the strongest

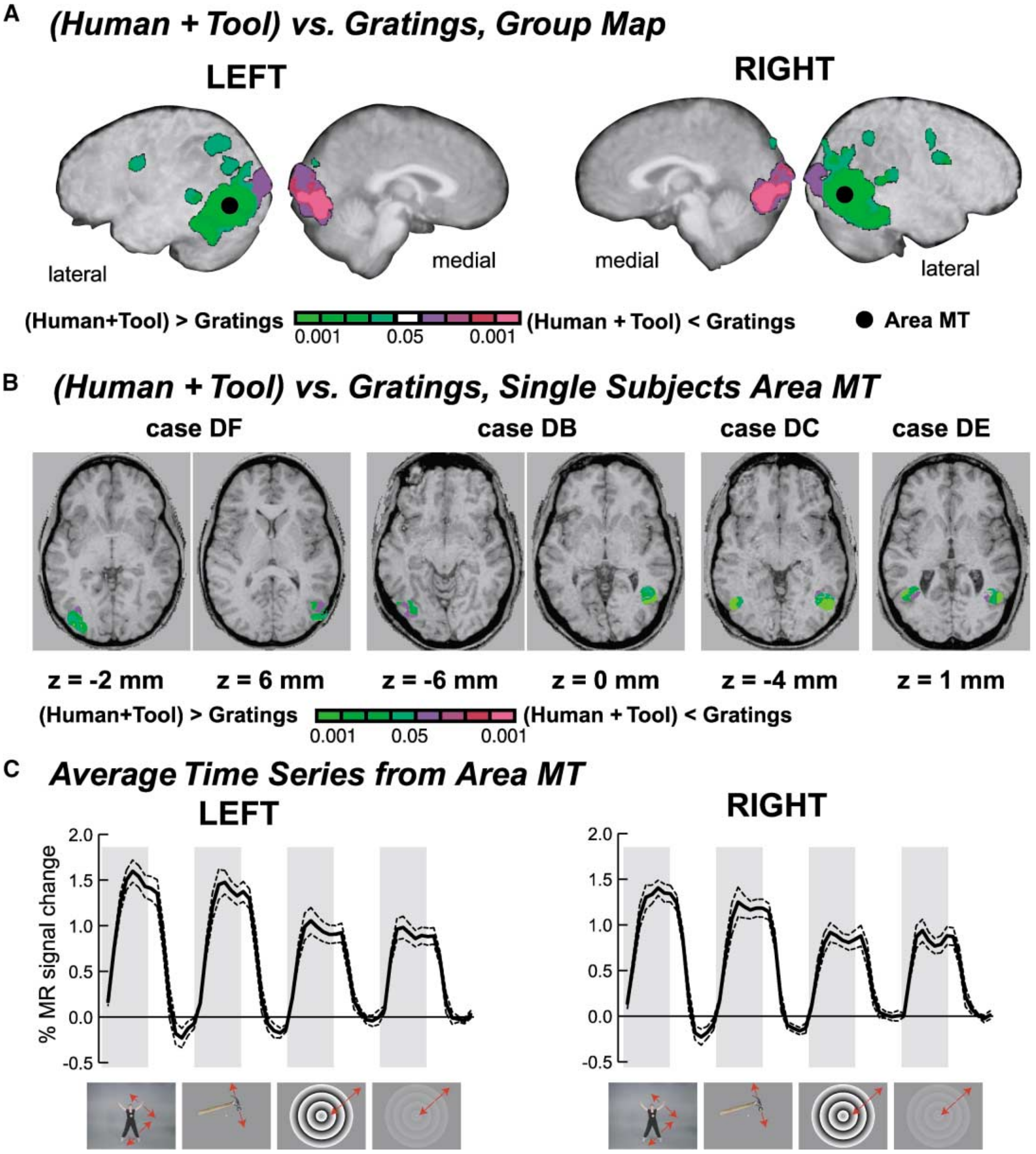


Figure 2. Cortical Response to Human and Tool Motion Compared with the Response to Radial Gratings

(A) Volume renderings illustrating the contrast of human and tool motion versus radial gratings. Average functional data from eight subjects (in color) is shown projected onto a rendering of the average anatomical dataset from the same subjects (grayscale). Lateral and medial views of left and right hemispheres are shown. Voxels are colored by their preference for human and tool motion (green color scale) or moving gratings (violet color scale). Activation within 15 mm of the midline is projected onto the medial surface; all other activation is projected onto the lateral surface. The average position of area MT (projected to the cortical surface) is shown as a black circle in the lateral views.

(B) Axial slices from four subjects (subject code labeled above each slice, Talairach coordinates below) showing the position of left and right area MT. Color scale indicates contrast of human and tool versus gratings.

C) The response to four types of motion stimuli in left and right area MT, averaged across eight subjects. The dark line in each graph shows the response to a 21 s stimulation block (indicated by gray bars) of each stimulus type, illustrated by a sample stimulus underneath each bar: human, tool, high-contrast grating, and low-contrast grating. Each stimulation block contained seven stimuli, all of the same type, and was followed by 21 s of fixation baseline. The dashed lines indicate one SE above and below the mean across subjects.

preference for moving human and tool stimuli compared with moving gratings. Ventral occipitotemporal cortex, intraparietal sulcus, and precentral sulcus also preferred human and tool motion to moving gratings.

Response of Area MT

The area with the greatest response for human and tool motion compared with moving gratings was found bilaterally in lateral occipitotemporal cortex, a region that includes area MT. We sought to analyze whether area MT showed a differential response to human or tool motion. First, we identified area MT using well-established functional and anatomical criteria (see Experimental Procedures) in both hemispheres of every subject, with an average position in the left hemisphere of (−42, −67, 1) and in the right hemisphere of (43, −64, 2). The position of area MT in four individual subjects is

Table 1. Talairach Coordinates of Activation Foci in Experiment 1

Regions that Preferred Radial Gratings to Human and Tool Motion

Anatomical Description		Coordinates		
		x	*y*	*z*
Calcarine cortex		−2	−81	−1

Regions that Preferred Human and Tool Motion to Radial Gratings

Anatomical Description		Coordinates		
		x	*y*	*z*
Posterior lateral temporal cortex	L	−45	−79	−7
	R	46	−69	2
Ventral occipitotemporal cortex	L	−39	−48	−19
(fusiform gyrus)	R	35	−46	−20
Inferior parietal cortex	L	−22	−57	44
(intraparietal sulcus)	R	22	−50	42
Premotor cortex	L	−44	12	32
(precentral sulcus)	R	52	14	29

Coordinates are locations of peak significance in the group activation map.

shown in Figure 2B. The response from left and right MT to each stimulus category was averaged across subjects (Figure 2C). While MT showed a stronger response for human and tool motion compared with gratings (1.3% versus 0.9%, $p = 0.0001$), it responded similarly to human and tool motion.

Response to Human and Tool Motion

Brain areas anterior to MT did exhibit differential responses to human and tool motion (Figure 3). Bilateral superior temporal sulcus and superior temporal gyrus showed a stronger response for human motion, while bilateral middle temporal gyrus and inferior temporal sulcus responded more strongly to tool motion. As shown in the average activation map (Figure 3A), the brain area with the largest differential response for human motion was found in right posterior STS with coordinates (47, −64, 10), while the strongest differential response for tool motion was found in left MTG with coordinates (−38, −63, −6). This hemispheric asymmetry was also observed in the volume of activation, with more cortex in the right hemisphere than left hemisphere showing a greater response to human than tool motion (average volume across subjects, 7868 mm^3 versus 4561 mm^3, $p = 0.03$) and more cortex in the left hemisphere preferring tool motion (2754 mm^3 versus 1500 mm^3, $p = 0.06$).

To determine the consistency of the functional distinction between superior and middle temporal visual motion areas, we created a volume rendering of the contrast between human and tool motion for each subject (Figure 3B). Every subject showed the same segregation observed in the group map, with STS responding most strongly to human motion and MTG responding most strongly to tool motion.

To determine the selectivity of the response, average time series were created from human motion preferring voxels (concentrated in STS) and tool motion preferring voxels (concentrated in MTG). As shown in Figure 4, both STS and MTG responded to all stimulus types, with the largest response to the preferred object motion, followed by the response to the nonpreferred object motion (STS, 1.2% for human versus 0.7% for tools; MTG, 1.1% for tools versus 0.7% for humans). Responses to the moving gratings stimuli in STS and MTG were significantly weaker than the responses to the nonpreferred object motion (0.3% versus 0.7%, $p = 10^{-5}$ for STS; 0.5% versus 0.7%, $p = 0.03$ for MTG).

In addition to the distinction between superior and middle temporal cortex, segregation between voxels with stronger responses to human or tool motion was also observed in other brain regions. In parietal cortex, human motion evoked a larger response than tool motion in medial superior parietal lobe, while tool motion produced a larger response in intraparietal sulcus. In ventral occipitotemporal cortex, the lateral aspect of the fusiform gyrus showed a larger response to human motion than tool motion, while the more medial region of the medial fusiform gyrus responded more strongly to tool motion than to human motion.

Experiment 2: Static versus Moving Images of Humans and Tools

Segregation between voxels with stronger responses to human or tool motion was observed in two regions of posterior cortex: lateral temporal lobe (superior versus middle temporal gyrus) and ventral temporal cortex (medial versus lateral fusiform gyrus). Previous studies have reported a similar pattern of category-related responses during the presentation of static pictures of humans and tools (Chao et al., 1999a; Haxby et al., 1999; Kanwisher et al., 1997). We wished to determine whether any of the observed responses in ventral and lateral temporal cortex were specific to stimulus motion. Hence, in a second experiment, we presented both still and moving images of humans and tools.

First, the average response to moving and still humans was compared with the average response to moving and still tools. We found the same pattern of differential sensitivity to stimulus category that was observed in the first experiment: in lateral temporal lobe, STS responded more strongly to human stimuli with peak coordinates (51, −69, 10), while MTG responded more strongly to tools with peak coordinates (−46, −70, −4). In ventral temporal lobe, lateral fusiform gyrus showed a stronger response to humans than tools, with peak Talairach coordinates in the group activation map of (39, −46, −11), while medial fusiform gyrus responded more strongly to tools than humans with peak coordinates (−28, −66, −11).

Next, we made average time series from voxels showing a differential response to human and tool motion. As shown in Figure 5, separate time series were created for STS and MTG in lateral temporal cortex and lateral fusiform gyrus (LFG) and medial fusiform gyrus (MFG) in ventral temporal cortex. Then, we compared the response to different stimuli in human and tool-preferring lateral and ventral areas (STS versus LFG and MTG versus MFG) using a three-factor ANOVA (region × motion × stimulus category) in which each subject was a replication. For human-preferring and tool-preferring regions in lateral and ventral cortex, the only significant two-way interaction was between region and motion ($p = 0.007$ for MTG/MFG, $p = 0.0007$ for STS/LFG). Middle temporal cortex responded much more strongly to moving tools than to still tools (0.55% versus 0.36%, $p = 0.02$), while medial fusiform showed no difference in response (0.51% versus 0.51%, $p = 0.97$). Superior

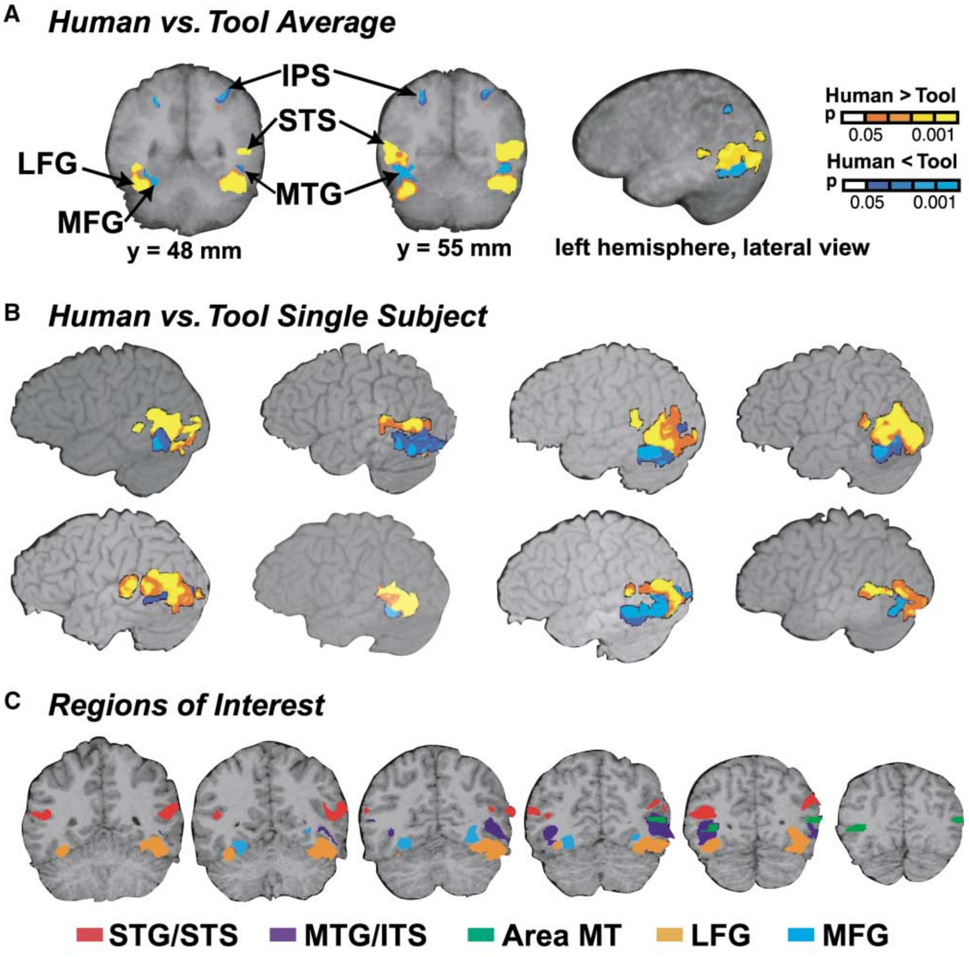


Figure 3. Anatomical Segregation of the Response to Human and Tool Motion

(A) Average functional data from eight subjects (in color) are shown overlaid on the average anatomical dataset from the same subjects (grayscale). Voxels are colored by their preference for human motion (orange-to-yellow color scale) or tool motion (blue color scale). The left two panels are coronal slices (Talairach location shown below each slice, left-is-left) and the right panel is a volume rendering (all activity greater than 15 mm from the midline projected to the surface). IPS, intraparietal sulcus; LFG, lateral fusiform gyrus; MFG, medial fusiform gyrus; STS, superior temporal sulcus; MTG, middle temporal gyrus.

(B) Contrast between human and tool motion in each subject's left lateral temporal cortex. Voxels in lateral temporal cortex showing a preference for human or tool motion are rendered on each subject's MR anatomical dataset. Human motion preferring voxels cluster in superior temporal cortex; tool motion preferring voxels cluster in middle temporal cortex. Same color scale as in (A).

(C) Regions of interest in a single subject. Coronal slices (left-is-left) progressing from anterior (leftmost slice, $y = -47$ mm) to posterior (rightmost slice, $y = -77$ mm) at 6 mm intervals. Different colors show location of active voxels within each anatomically defined ROI (STG/STS, superior temporal gyrus and sulcus; MTG/ITS, middle temporal gyrus and inferior temporal sulcus; LFG, lateral fusiform gyrus; MFG, medial fusiform gyrus).

temporal cortex showed a much stronger response to moving humans than to static humans (0.54% versus 0.31%, $p = 0.0005$), while lateral fusiform showed a much smaller difference (0.57% versus 0.50%, $p = 0.06$). While lateral temporal regions (STS and MTG) strongly preferred moving stimuli, they showed the same category specificity for moving and still stimuli. STS preferred moving humans to moving tools and also preferred still humans to still tools, while MTG preferred moving tools to moving humans and also preferred still tools to still humans.

Experiment 3: Humans and Tools Moving with Articulated and Unarticulated Motion

In experiments 1 and 2, STS responded maximally to moving humans, while MTG responded maximally to moving tools. This could occur because of a specialization in STS and MTG for processing different kinds of object motion, or it could reflect a response to the visual form of each stimulus category plus a nonspecific sensitivity to any type of visual motion. In order to distinguish between these possibilities, in a third experiment, we separated the form and motion properties of the stimuli. Tools typically move with few degrees of articulation, while human motions usually have many degrees of articulation. Still frames from the human and tool stimuli were animated with identical artificial motion vectors consisting of simple translation and rotation. Therefore, human forms could move with either articulated motion (e.g., jumping jacks, as shown in experiment 2) or unarticulated motion (translation and rotation). If the STS preference for human motion consists only of a form

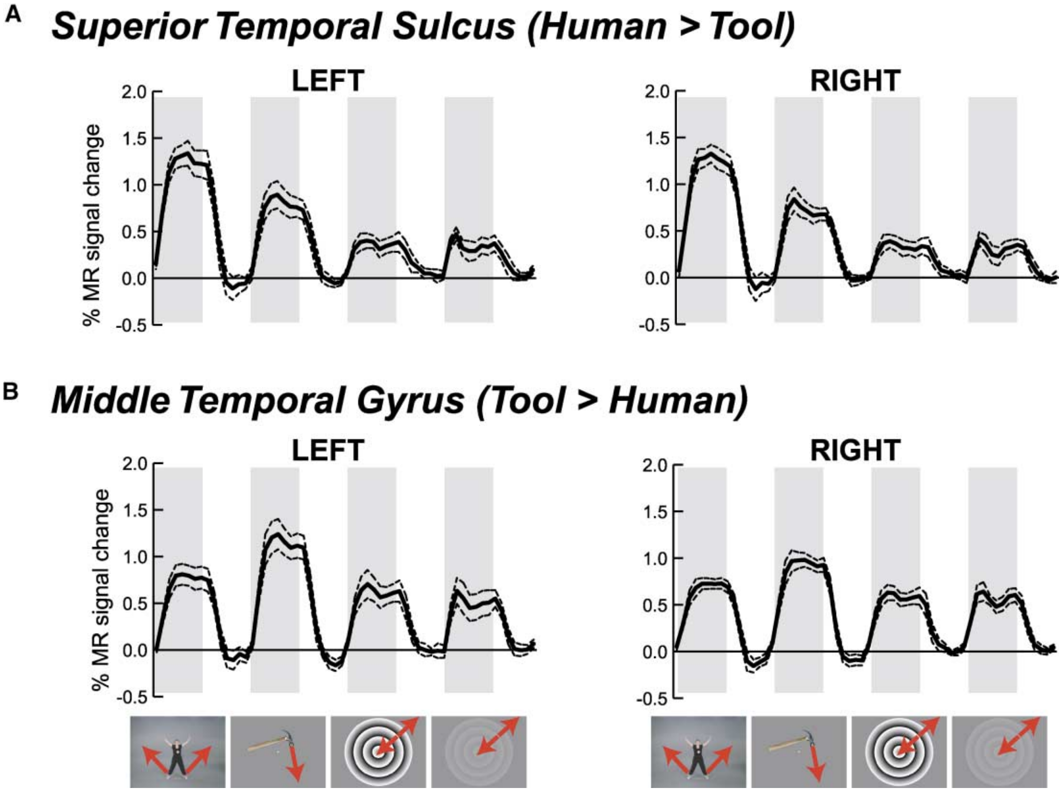


Figure 4. Lateral Temporal Responses to Human Motion, Tool Motion, and Radial Gratings

The response to four types of motion stimuli in four regions of interest (superior temporal sulcus and middle temporal gyrus in left and right hemispheres) averaged across eight subjects. The dark line in each graph shows the response to a 21 s stimulation block (indicated by gray bars) of each stimulus type, illustrated by a sample stimulus underneath each bar: human, tool, high-contrast grating, and low-contrast grating. Each stimulation block contained seven stimuli, all of the same type, and was followed by 21 s of fixation baseline. The dashed lines indicate one SE above and below the mean across subjects.

preference plus a nonspecific motion response, it should respond similarly to both articulated and unarticulated human motion. Alternately, if STS prefers typically human motion qualities, it should show a greater response to articulated than unarticulated human motion stimuli.

Figure 6 illustrates the average response in STS and MTG to natural and artificially moving humans and tools. STS showed a strong preference for natural articulated human motion compared with artificial unarticulated motion (0.56% versus 0.46%, $p = 0.01$). In contrast, MTG showed a trend toward preferring unarticulated to articulated human motion (0.28% versus 0.23%, $p = 0.08$) and MTG showed equivalent responses to tools moving with natural or artificial unarticulated motion (0.45% versus 0.44%, $p = 0.8$). These differential responses suggest that STS and MTG are specialized for processing the types of motion common to their preferred stimulus categories.

Discussion

Both lateral and ventral temporal cortex showed category-related responses. While ventral temporal cortex showed similar responses to moving and static human and tool stimuli, lateral temporal cortex showed larger responses to moving compared with static humans and tools. Lateral temporal cortex also strongly preferred moving humans and tools to moving radial gratings, suggesting that lateral temporal cortex is the cortical locus of complex motion processing. Our results suggest that within lateral temporal cortex there are two parallel visual motion processing streams that begin just anterior to area MT. The superior stream, in STS, is more responsive to biological motion, while the inferior stream, in MTG, is more responsive to tool motion.

In order to understand the responses to human motion in STS and tool motion in MTG, we consider the contributions of three different factors: the form of the stimulus, the role of stored information about motion, and the motion properties of different visual stimuli.

Responses to Stationary Stimuli

Our results are consistent with a model in which lateral temporal cortex is specialized for processing object motion, while ventral temporal cortex is specialized for object form. Since the form of an object (such as a hammer) remains the same when it is still and when it is moving, areas that respond only to form should exhibit similar responses regardless of motion, as we observed in ventral temporal cortex (medial and lateral fusiform gyrus) in experiment 2.

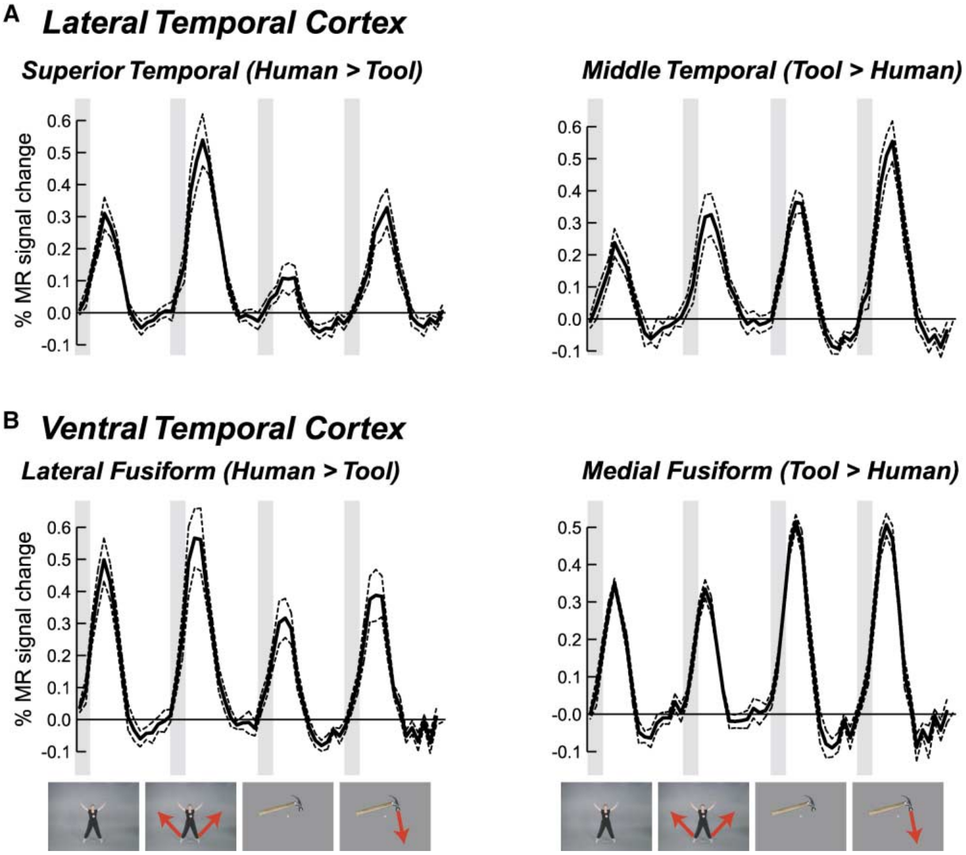


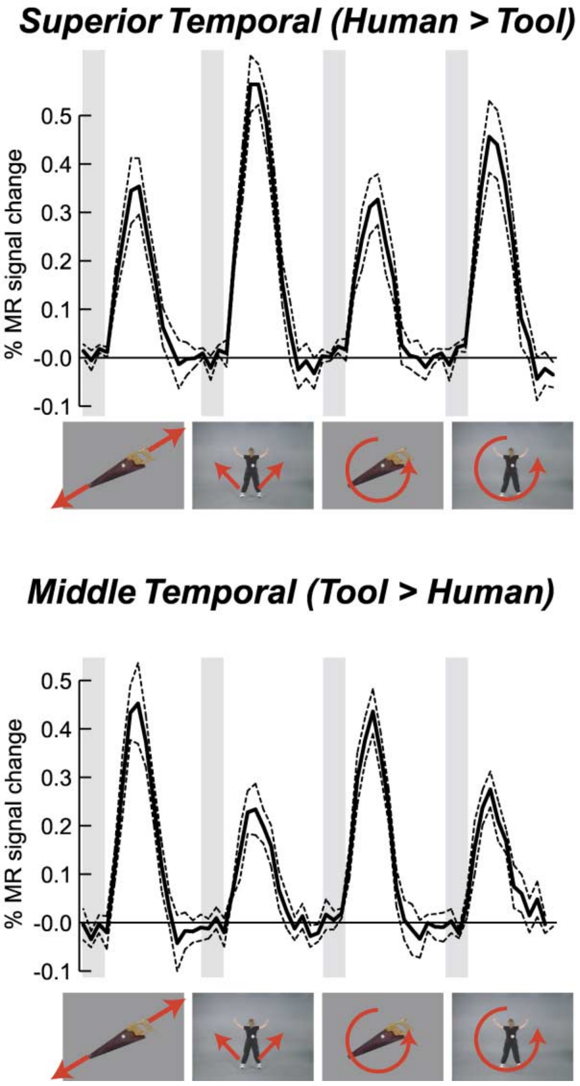
Figure 5. Lateral and Ventral Temporal Responses to Moving and Static Images of Humans and Tools

A rapid event-related design was used to determine the hemodynamic response to a single 2.5 s presentation of each stimulus type. The stimulus presentation duration is indicated by a thin gray bar and corresponds to the stimulus type shown by the sample below each bar in the lower panel (static human, moving human, static tool, and moving tool). Stimuli were presented in pseudo-random order, and the impulse response to each stimulus type was estimated with 1 s resolution in a 15 s window following stimulus onset. Responses were averaged across active voxels in four anatomically defined regions of interest traced on high-resolution MR images for each subject. The thick line in each graph shows the mean response within each ROI averaged across six subjects and the thin dashed lines shows one SE above and below the mean across subjects.

While lateral temporal cortex (STS and MTG) responded much more to moving than static objects, they still exhibited a significant category-selective response to both moving and static objects. Although STS responds to facial motion and to point-light displays of whole-body motion (Grossman et al., 2000; Grossman and Blake, 2001; Puce et al., 1998) it also responds to static images of human faces and animals (Chao et al., 1999a, 1999b; Haxby et al., 1999; Kanwisher et al., 1997) (Figure 5). One cause of this activity is likely to be sensitivity to visual form, with MTG neurons preferring the form of tool-like objects and STS preferring the form of biological objects. However, the response to static images also reflects an additional component: stored information about different visual stimuli.

The Role of Stored Information

Activity in STS is greater when subjects attend to the direction of eye gaze of a static face than when they attend to its identity (Hoffman and Haxby, 2000), even though the form of the two stimuli is identical. Similarly, activity in lateral temporal regions is greater when subjects view still pictures of scenes that contain implied motion than when they view scenes without implied motion (Kourtzi and Kanwisher, 2000; Senior et al., 2000). These findings suggest that in addition to the actual motion present in the visual scene, lateral temporal cortex responds to the potential for motion in the visual scene. In other words, it contains stored information about motion: the knowledge that eyes are a moveable feature in a face, that a person looking toward an object may soon reach to grasp it, or that an upraised hammer might soon begin a down stroke. Additional evidence that lateral temporal areas store information about motion in addition to simply processing visual motion comes from the result that generating action words activates middle temporal gyrus, as does naming tools or answering written questions about tools (for review, see Martin and Chao, 2001). The notion that lateral temporal cortex may store information about motion as well as process visual motion in the current scene is consistent with models of visual motion processing as a Bayesian system in which information about previously observed scenes puts strong constraints on the analysis of the current visual input (Rao, 1999; Snowden and Verstra-



Superior Temporal (Human > Tool)

Middle Temporal (Tool > Human)

Figure 6. Lateral Temporal Responses to Natural and Artificial Human and Tool Motion

Humans and tools moved with either characteristic natural motions or with artificially generated translation and rotation. The motion vectors for the artificially generated motions were identical for human and tool stimuli. The stimuli were presented in random order, and deconvolution was used to estimate the response to a single 2.5 s stimulus of each type in a 15 s window following stimulus onset. The thin gray bars show the stimulus duration and correspond to the stimulus type shown below each bar (natural tool, natural human, artificial tool, and artificial human). The thick line in each graph shows the mean response from five subjects to a single presentation of each stimulus within anatomically defined ROIs, with the thin dashed lines illustrating one SE above and below the mean across subjects.

ten, 1999). This stored information about motion may be crucial for the identification of manipulable objects, as suggested by studies of patients with damage to posterior temporal cortex (Tranel et al., 1997) who were found to have deficits that impaired their ability to recognize and name tools.

Selectivity for Motion Properties

While selectivity for visual form and a nonspecific motion response could explain the lateral temporal responses to moving and static objects observed in experiment 2, they cannot account for the differential responses to different types of motion observed in experiment 3, in which stimuli with the same form moved with different motions. MTG showed a trend toward a larger response to unarticulated than to articulated human motion, supporting the hypothesis that middle temporal neurons may prefer the rigid-body motion characteristic of tools, in which connected elements move in unison, maintaining the same spatial relationship (such as the fixed relationship of handle, head, and claw of a hammer during its down stroke). Consistent with this notion, MTG showed no difference between tools moving with natural and artificially created unarticulated motion vectors. In contrast, STS preferred articulated to unarticulated human motion, suggesting that neurons in superior temporal cortex may prefer visual stimuli in which different points in the object move without maintaining a consistent spatial arrangement (such as the differing arrangement of head, hand, and foot during a jumping jack.) An organization by motion properties may allow an explanation of the patterns of response to moving objects based on a small set of visual motion primitives.

Recently, a study by Downing et al. (2001) demonstrated that a region in lateral temporal cortex (which they label the "extrastriate body area" or EBA) responds to images of human body parts. In their study, the EBA was located in the superior temporal sulcus directly posterior to the region of the STS that responds strongly to images of faces. Because our stimulus set contained images of faces and bodies, it seems likely that the region of superior temporal cortex active in our study (Figure 3) contains the STS face area, EBA, and perhaps other areas important for decoding complex biological motion.

Lateral temporal cortex almost certainly encodes multiple stimulus dimensions, two of which seem to be motion and form. The results of experiments 2 and 3 reveal that MTG responds to the rigid-body motion characteristic of tools while STS responds to human-like articulated motion trajectories. The results of experiment 2 demonstrate a parallel organization by visual form, with MTG preferring static tool forms and STS preferring static human forms. The interaction between motion and form processing remains largely unexplored. Recent evidence (Kourtzi et al., 2002) demonstrates that small regions of lateral temporal cortex prefer images of intact objects to scrambled images and prefer moving to stationary stimuli, suggesting a common neural substrate for form and motion processing. When objects move, they assume different configurations within the visual field. Therefore, in cortical areas that are both form and motion selective, responses to object motion could be a combination of motion responses (to the moving object) and form responses (to the different configurations the object assumes while in motion).

Responses to Moving Stimuli: Occipital Motion Responses

Low-level visual areas in medial occipital cortex showed a strong preference for moving gratings compared with human and tool motion. These areas are retinotopically organized (DeYoe et al., 1996; Engel et al., 1997; Sereno et al., 1995), and the radial gratings and human and tool stimuli were not matched for retinotopic extent. The grating stimuli contained moving contours at all locations in the visual field, as opposed to the human and tool stimuli, which contained moving contours only at the spatially restricted boundary of the human or tool. Therefore, low-level visual areas would be expected to show a greater response to the visually more extensive radial grating. In addition, early visual areas are sensitive to motion energy and image contrast: the radial gratings contained motion at all points in the visual field, while the high-contrast grating contained higher contrast than the human or tool stimuli.

However, these low-level stimulus features cannot explain the responses of area MT and more anterior lateral temporal regions. Area MT is only weakly retinotopically organized and responded less, not more, to the retinotopically more extensive grating stimuli. Area MT's contrast response curve saturates at low image contrast and MT responded equally to low- and high-contrast gratings.

There are several possible explanations for MT's preference for human and tool stimuli compared with gratings. First, area MT is strongly modulated by attention (Beauchamp et al., 1997, 2001; O'Craven et al., 1997). Human and tool motion may be inherently more attention eliciting than moving gratings (even though subjects performed an equally difficult, attention-demanding task with both types of stimuli) leading to the observed response pattern. Second, area MT receives reciprocal connections from more anterior temporal areas (Desimone and Ungerleider, 1986; Maunsell and Van Essen, 1983). The observed modulation could reflect a feedback enhancement of lower-level motion areas by more selective anterior areas. Finally, the area that we have labeled "MT" likely contains MST and other associated motion areas. Because MST is known to prefer more complex motion stimuli, such as optic flow (Duffy and Wurtz, 1991; Vaina, 1998), the observed greater response in and around MT could reflect single-unit response properties that favor the more complex motion associated with objects over the simpler motion properties of moving gratings.

Experimental Procedures

Visual Stimuli

In the experiment 1, four types of motion stimuli were presented: human motion, tool motion, and low- and high-contrast radial gratings (Figures 1A–1D). The motion stimuli were presented as short video clips with duration 2.5 s at 30 frames per second. A single still frame from each stimulus type is shown in Figure 1. For video clips of each stimulus type, see Supplemental data at http://www.neuron.org/cgi/content/full/34/1/149/DC1. All motion stimuli were presented on a gray background and contained a central fixation cross overlaid on the motion stimulus. In the experiments 2 and 3, four additional stimulus categories were created. Static images of humans and tools (created by selecting single still frames from the human motion and tool motion video clips) were presented for 2.5 s (Figures 1E and 1F). Video clips (2.5 s long) of artificially moving humans and tools were created by translating and rotating still images of humans and tools (Figures 1G and 1H).

Visual stimuli were presented using MATLAB (Mathworks Inc., Natick, MA) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) running on a Macintosh G4 (Apple Computer, Cupertino, CA). Stimuli were back-projected onto a Lucite screen using a 3-panel LCD projector (Sharp Inc., Mahwah, NJ). Stimulus presentation was synchronized with MR data acquisition using a DAQ board (National Instruments, Austin, TX), and subject responses were recorded using SuperLab software (Cedrus Corp., San Pedro, CA).

Biological Motion

Twenty-eight common whole-body motions (see Appendix A for listing) were selected. A male and a female actor each performed many repetitions of each movement, which were filmed from five perspectives (facing the camera; profile, facing right; profile, facing left; three quarters profile, facing right; and three quarters profile, facing left). Five samples of each complete movement from each viewpoint (2.5 s each) were saved in Quick Time (Apple Computer) format.

Tool Motion

Twenty-eight common household objects (see Appendix A for listing) were modified so that they could be operated from a distance using wood and metal dowels. The dowels were painted green, and the puppeted tools were videotaped on a green background. Nonlinear digital editing tools (Discreet Logic, Montreal, Canada) were used to replace the moving dowels and stationary background with a uniform gray field. The result was video clips of tools that moved realistically but with no visible controller such as a human hand or glove. Five samples of each tool's motion were saved.

Simple Motion

The simple motion stimuli (Tootell et al., 1995) consisted of radial gratings made up of concentric light and dark rings (spatial frequency 0.3 cycles per degree) that either expanded or contracted at 4° per second in each 2.5 s trial. The low-contrast simple motion stimulus had a contrast between light and dark rings, measured as (high − low)/(high + low) of 0.06, so that the rings were barely visible when stationary, but clearly visible when moving. The high-contrast simple motion stimulus had the highest possible contrast between light and dark rings, 1.0.

Experiment 1: Experimental Design and Visual Task

Within each 275 s MR scan series, subjects viewed 21 s blocks of no visual stimulation (white fixation crosshairs on a gray background) alternating with six 21 s blocks of visual motion stimulation (white fixation crosshairs overlaid on a motion stimulus on a gray background). Each 21 s stimulation block contained seven 3 s trials (2.5 s of motion followed by a 0.5 s intertrial interval) of a single type of motion stimulus (human motion; tool motion; high-contrast grating; low-contrast grating). The blocks were presented in a pseudo-random order that differed in each MR scan series. In order to ensure attention to the visual stimulus, subjects were required to decide if each motion stimulus matched the immediately preceding stimulus, responding with a button press on a response device held in the right hand. During simple motion stimulation blocks, each successive radial grating stimulus either expanded or contracted, and subjects decided if the direction of motion was the same as the previous trial. During human and tool motion blocks, subjects decided if the tool in the current trial was the same or different as the tool in the preceding trial or if the actor was the same or different as in the preceding trial.

Experiment 2: Experimental Design and Visual Task

Each of four stimulus types (static human, moving human, static tool, and moving tool) was presented in a rapid event-related design that allowed the measurement of responses to individual presentations of visual stimuli (Figure 5). Within each 275 s MR scan series, subjects viewed 14 trials of each stimulus type and 34 trials of 3 s fixation baseline, randomly ordered for optimal experimental efficiency (Dale, 1999). The visual task consisted of a simple two-alternative forced choice discrimination, with subjects deciding if the stimulus contained a human or a tool.

Experiment 3: Experimental Design and Visual Task

Each of four stimulus types was presented in a rapid event-related design similar to that used in experiment 2. Tools and humans moved either naturally (as in experiments 1 and 2) or with identical, artificially created motion vectors consisting of translation and rotation (see Figure 6 for illustration), and subjects decided whether the stimulus contained natural or artificial motion.

Behavioral Data

In experiment 1, there was no difference in reaction time (RT) between moving gratings (mean ± SD of 1228 ± 282 ms) and moving humans and tools (1189 ± 223 ms, $p = 0.3$). There was no significant difference for high- versus low-contrast gratings (1230 ± 306 ms versus 1227 ± 263 ms, $p = 0.9$), but there was a significantly longer RT for human motion compared with tool motion stimuli (1273 ± 223 ms versus 1106 ± 227 ms, $p < 0.0001$). In experiment 2, RT was significantly longer for moving than static stimuli (1243 ± 427 ms versus 1140 ± 360 ms, $p = 0.02$) and a trend was observed toward a longer RT for human compared with tool stimuli (1225 ± 424 ms versus 1157 ± 364 ms, $p = 0.09$). In experiment 3, there was no significant difference in reaction time between motion categories (biological motion, 1275 ± 96 ms; tool motion, 1283 ± 74 ms; translating and rotating humans, 1307 ± 35 ms; and translating and rotating tools, 1296 ± 86 ms.) Subjects performed at behavioral ceiling in all tasks ($<$1 error per stimulus category per scan series).

MR Data Collection

Eighteen subjects (eight females, eleven males) recruited from the NIH community were paid for their participation in the study. Eight subjects took part in experiment 1, six subjects took part in experiment 2, and five subjects took part in experiment 3 (one subject took part in both experiments 2 and 3). MR data were collected on a GE Signa 3 Tesla scanner using standard imaging procedures. A high-resolution SPGR anatomical sequence was collected at the beginning of each scanning session. Echo-planar volumes sampled the entire cortex with 3.75 mm in-plane resolution and an echo time (TE) of 30 ms. In experiment 1, between 28 and 34 axial slices (as necessary to achieve whole-brain coverage) with 4 mm slice thickness and a repetition time (TR) of 2.5 s were acquired. In experiments 2 and 3, 24 axial slices (slice thickness of 4.5 or 5.0 mm as necessary) with a TR of 2 s were acquired. In experiment 1, eight scan series were collected for each subject with 110 volumes per series. In experiments 2 and 3, eight or ten series were collected for each subject with 132 volumes per series.

fMRI Data Analysis

MR data were analyzed using a random effects approach within the framework of the general linear model in AFNI 2.29 (Cox, 1996). The first two volumes in each scan series, collected before equilibrium magnetization was reached, were discarded. Then, all volumes were registered to the volume collected nearest in time to the high-resolution anatomy. Next, a spatial filter with a root-mean-square width of 4 mm was applied to each echo-planar volume. The response to each stimulus category compared with the fixation baseline was calculated using multiple regression. All areas that showed a response to any stimulus type were included in the analysis.

For experiment 1, multiple regression was performed using 32 regressors of no interest (mean, linear trend, and second-order polynomial within each scan series to account for slow changes in the MR signal; and eight outputs from volume registration to account for residual variance from subject motion not corrected by registration) and four regressors of interest, one for each stimulus type. Each regressor of interest consisted of a square wave for each stimulation block of that stimulus type, convolved with a γ-variate function to account for the slow hemodynamic response (Cohen, 1997). Four contrasts were used: all regressors of interest, radial gratings versus human and tool motion, high-contrast versus low-contrast radial gratings, and human versus tool motion.

In experiments 2 and 3, the impulse response function to each stimulus category was estimated with 1 s resolution using deconvolution (Glover, 1999). A separate regressor was used to model the response in each 1 s period in a 15 s window following each stimulus presentation. With four stimulus types, this resulted in 60 regressors of interest (each consisting of a series of delta functions) resulting in an estimate of the response to a single stimulus of each type with no assumptions about the dynamics of the hemodynamic response.

Activation maps were created by using the overall experimental-effect (all regressors of interest) to find voxels showing a response to any type of stimulus. This threshold was set extremely high to correct for the multiple comparisons produced by the approximately 25,000 intracranial functional voxels. In experiment 1, the group average F-map was thresholded at $F > 30$, $p < 10^{-23}$ (uncorrected). The overall experimental-effect thresholds for each individual subject were adjusted to account for the differing levels of physiological noise in each subject (Cox and Saad, 2001) (mean $F > 40$, $p < 10^{-30}$, uncorrected). For experiments 2 and 3, maps of the overall experimental effect for individual subjects were thresholded at a lower value ($F > 9$, $p < 10^{-8}$) because of the decreased sensitivity of event-related relative to block designs (Friston et al., 1999).

Following stringent thresholding by the experimental-effect contrast, voxels were displayed using a color scale based on the significance of the gratings versus human and tool motion contrast or the human versus tool motion contrast with a threshold of $p < 0.05$. Functional data was interpolated to 1 mm^3 resolution using cubic interpolation and overlaid on single-subject average anatomical data or group average anatomical data.

Identification of Area MT

In experiment 1, area MT was identified in each subject by its anatomical location and strong response to low-contrast motion. First, we found cortical regions showing a strong ($p < 10^{-10}$) response to low-contrast motion and a response of equal amplitude ($p > 0.01$) to low- and high-contrast motion, distinguishing characteristics of area MT (Tootell et al., 1995). Next, we selected the active region closest to the junction of the inferior temporal sulcus with the ascending limb of the inferior temporal sulcus, a reliable anatomical landmark for area MT (Dumoulin et al., 2000; Watson et al., 1993).

Region of Interest Identification

In addition to the anatomical-functional definition of area MT (above) anatomically defined regions-of-interest (ROI) were manually traced on the Talairach-normalized high-resolution anatomy in both hemispheres of every subject. For experiments 1 and 3, left and right lateral temporal ROIs were created that extended (in the superior to inferior direction) from superior temporal gyrus to the lower bank of the inferior temporal sulcus. For experiment 2, additional ROIs in left and right ventral temporal cortex were constructed, extending in the lateral-to-medial direction from the inferior termination of the middle temporal ROI in inferior temporal sulcus to the collateral sulcus. All ROIs covered the same extent in the anterior-to-posterior direction, from mid-temporal lobe (anterior to the observed functional activation) to the posterior termination of the superior temporal sulcus in occipital lobe. The average response to each stimulus category was calculated from all voxels in each ROI that showed a significant experimental effect and a significant effect of the human versus tool contrast ($p < 0.05$).

Appendix A

The common household objects were: bread knife, chef's knife, compass, slotted spatula (flipper), hammer, ladle, needle-nose pliers, paintbrush, pencil, pencil sharpener, ping-pong ball bouncing on ping-pong paddle, plastic drink pitcher, pliers, ratchet, rolling pin, wood saw, scissors, soup-type ladle, screwdriver, scrub brush, shovel, small artist's paintbrush, rubber spatula, trowel, utility knife, wall paper brush, x-acto knife, and yoyo. The common whole-body motions were: toe touch, jumping jack, pushup, stationary jog, sit up, up-and-down jump, skip, salute, side-to-side jump, stair climb, spin, sit down from standing position, lie down from standing position, shadow box, clap, hop, stationary kick, run, tightrope walk pantomime, swim pantomime, ponder pantomime—hand on chin, running kick, stationary leap, running leap, seated leg cross, standup from seated position, wave, and yawn.

Received: August 31, 2001
Revised: February 4, 2002

References

Aguirre, G.K., Zarahn, E., and D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron *21*, 373–383.

Beauchamp, M.S., Cox, R.W., and DeYoe, E.A. (1997). Graded effects of spatial and featural attention on human area MT and associated motion processing areas. J. Neurophysiol. *77*, 516–520.

Beauchamp, M.S., Petit, L., Ellmore, T.M., Ingeholm, J., and Haxby, J.V. (2001). A parametric fMRI study of overt and covert shifts of visuospatial attention. Neuroimage *14*, 310–321.

Bonda, E., Petrides, M., Ostry, D., and Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J. Neurosci. *16*, 3737–3744.

Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433–436.

Chao, L.L., Haxby, J.V., and Martin, A. (1999a). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. *2*, 913–919.

Chao, L.L., Martin, A., and Haxby, J.V. (1999b). Are face-responsive regions selective only for faces? Neuroreport *10*, 2945–2950.

Cohen, M.S. (1997). Parametric analysis of fMRI data using linear systems methods. Neuroimage *6*, 93–103.

Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. *29*, 162–173.

Cox, R.W., and Saad, Z.S. (2001). Direct estimation of DOF in FMRI regression analyses. Neuroimage *13*, S101.

Dale, A.M. (1999). Optimal experimental design for event-related fMRI. Hum. Brain Mapp. *8*, 109–114.

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.

DeYoe, E.A., Carman, G., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., and Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. Proc. Natl. Acad. Sci. USA *93*, 2382–2386.

Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science *293*, 2470–2473.

Duffy, C.J., and Wurtz, R.H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. J. Neurophysiol. *65*, 1329–1345.

Dumoulin, S.O., Bittar, R.G., Kabani, N.J., Baker, C.L., Jr., Le Goualher, G., Bruce Pike, G., and Evans, A.C. (2000). A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb. Cortex *10*, 454–463.

Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb. Cortex *7*, 181–192.

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598–601.

Friston, K.J., Zarahn, E., Josephs, O., Henson, R.N., and Dale, A.M. (1999). Stochastic designs in event-related fMRI. Neuroimage *10*, 607–619.

Glover, G.H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage *9*, 416–429.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187–203.

Grossman, E.D., and Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. Vision Res. *41*, 1475–1482.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., and Blake, R. (2000). Brain areas involved in perception of biological motion. J. Cogn. Neurosci. *12*, 711–720.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L., Hoffman, E.A., and Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. Neuron *22*, 189–199.

Haxby, J.V., Ishai, I., Chao, L.L., Ungerleider, L.G., and Martin, A. (2000). Object-form topology in the ventral temporal lobeResponse to I. Gauthier (2000). Trends Cogn. Sci. *4*, 3–4.

Hoffman, E.A., and Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat. Neurosci. *3*, 80–84.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. *17*, 4302–4311.

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

Kourtzi, Z., Bulthoff, H.H., Erb, M., and Grodd, W. (2002). Object-selective responses in the human motion area MT/MST. Nat. Neurosci. *5*, 17–18.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., and Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. Cereb. Cortex *11*, 287–297.

Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc. Natl. Acad. Sci. USA *92*, 8135–8139.

Martin, A., and Chao, L.L. (2001). Semantic memory and the brain: structure and processes. Curr. Opin. Neurobiol. *11*, 194–201.

Martin, A., Wiggs, C.L., Ungerleider, L.G., and Haxby, J.V. (1996). Neural correlates of category-specific knowledge. Nature *379*, 649–652.

Martin, A., Ungerleider, L.G., and Haxby, J.V. (2000). Category specificity and the brain: the sensory/motor model of semantic representations of objects. In The New Cognitive Neurosciences, M.S. Gazzaniga, ed. (Cambridge, MA: MIT Press), pp. 1023–1036.

Maunsell, J.H.R., and Van Essen, D.C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. J. Neurosci. *3*, 2563–2586.

O'Craven, K.M., Savoy, R.L., and Rosen, B.R. (1997). Attention modulates fMRI activation in human MT/MST. Neuron *21*, 1988.

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.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. *10*, 437–442.

Puce, A., Allison, T., Bentin, S., Gore, J.C., and McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. J. Neurosci. *18*, 2188–2199.

Rao, R.P. (1999). An optimal estimation approach to visual perception and learning. Vision Res. *39*, 1963–1989.

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 or representational momentum. Curr. Biol. *10*, 16–22.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889–893.

Snowden, R.J., and Verstraten, F.A. (1999). Motion transparency: making models of motion perception transparent. Trends Cogn. Sci. *3*, 369–377.

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.

Tranel, D., Damasio, H., and Damasio, A.R. (1997). A neural basis for the retrieval of conceptual knowledge. Neuropsychologia *35*, 1319–1327.

Vaina, L.M. (1998). Complex motion perception and its deficits. Curr. Opin. Neurobiol. *8*, 494–502.

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.