fMRI Responses to Video and Point-Light Displays of Moving Humans and Manipulable Objects

Michael S. Beauchamp, Kathryn E. Lee, James V. Haxby, and Alex Martin

Abstract

■ We used fMRI to study the organization of brain responses to different types of complex visual motion. In a rapid event-related design, subjects viewed video clips of humans performing different whole-body motions, video clips of manmade manipulable objects (tools) moving with their characteristic natural motion, point-light displays of human whole-body motion, and point-light displays of manipulable objects.

The lateral temporal cortex showed strong responses to both moving videos and moving point-light displays, supporting the hypothesis that the lateral temporal cortex is the cortical locus for processing complex visual motion. Within the lateral temporal cortex, we observed segregated responses to different types of motion. The superior temporal sulcus (STS) responded strongly to human videos and human point-light displays, while the middle temporal gyrus (MTG) and the inferior temporal sulcus responded strongly to tool videos and tool point-light displays.

In the ventral temporal cortex, the lateral fusiform responded more to human videos than to any other stimulus category while the medial fusiform preferred tool videos. The relatively weak responses observed to point-light displays in the ventral temporal cortex suggests that form, color, and texture (present in video but not point-light displays) are the main contributors to ventral temporal activity. In contrast, in the lateral temporal cortex, the MTG responded as strongly to point-light displays as to videos, suggesting that motion is the key determinant of response in the MTG. Whereas the STS responded strongly to point-light displays, it showed an even larger response to video displays, suggesting that the STS integrates form, color, and motion information. ■

INTRODUCTION

A typical visual scene contains dozens of complex objects, many of them in motion. A network of brain regions has been implicated in the cortical processing of these stimuli. The ventral temporal cortex is especially responsive to complex objects (Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001; Grill-Spector et al., 1999; Malach et al., 1995) and within the ventral temporal cortex, different areas respond to different categories of objects (Haxby et al., 1999, 2001; Chao, Haxby, & Martin, 1999; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998). However, the ventral temporal cortex shows little difference in response to static complex objects and the same objects in motion, while the lateral temporal cortex responds much more strongly to complex stimuli when they are moving instead of static (Beauchamp, Lee, Haxby, & Martin, 2002).

In the posterior lateral temporal cortex, area MT is the canonical visual motion processing area, responding to all types of visual motion. Anterior to MT, other regions are responsive to different types of complex object motion. Cortex anterior and superior to MT in the superior temporal sulcus (STS) is especially sensitive to biological motion. In monkeys, single neurons in the STS respond to hand, eye, mouth, or body movements (Oram & Perrett, 1994; Perrett et al., 1988). In humans, mouth, eye movements, and whole-body movements activate the STS (Beauchamp et al., 2002; Puce, Allison, Bentin, Gore, & McCarthy, 1998; Bonda, Petrides, Ostry, & Evans, 1996). Event-related potential studies show that these responses are exquisitely sensitive to the details of the biological motion, such as whether a visually presented hand is opening or closing (Wheaton, Pipingas, Silberstein, & Puce, 2001). The STS is more active in response to points of light displays containing biological motion than similar displays containing scrambled or inverted motion (Grossman et al., 2000; Grossman & Blake, 2001; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001; Howard et al., 1996) and a patient with damage to the lateral temporal cortex was unable to decode point-light displays (Cowey & Vaina, 2000).

Motion is also an important attribute of manmade manipulable objects such as tools and utensils. In addition to eliciting activity in the ventral temporal cortex,

identification of static images of these objects is associated with activity in the lateral temporal cortex (e.g., Chao et al., 1999; Martin, Wiggs, Ungerleider, & Haxby, 1996). However, this activity does not appear in the STS but inferior to the STS in the posterior middle temporal gyrus (MTG), extending inferiorly into the inferior temporal sulcus. The characteristic motion of a tool is an important identifying attribute because it is directly related to the tool's function (Martin, Ungerleider, & Haxby, 2000).

Point-light displays, first described by Johansson (1973), are a striking example of the visual system's ability to use stored knowledge about the visual world to extract information from impoverished displays. For instance, the six points shown in the still frames of Figure 1C and D contain no color or texture information and do not form easily recognizable shapes. However, once put in motion, the rigid up-and-down motion of the points in Figure 1D reveal a hammer, while the fluid motion of the points in Figure 1C lead observers to recognize a human being rising from a seated position. Even infants are able to decode displays of point-light biological motion (Fox & McDaniel, 1982) and they convey a surprising amount of information, including the sex of the actor (Mather & Murdoch, 1994). While previous neuroimaging studies have examined cortical responses to human point-light displays (Grossman et al., 2000; Grossman & Blake, 2001, 2002; Servos, Osu, Santi, & Kawato, 2002; Grezes et al., 2001; Vaina et al.,

2001; Howard et al., 1996), none have compared point-light displays of human motion with point-light displays of tool motion. Because motion is the sole carrier of visual information in point-light displays, comparing the brain responses to point-light displays and displays containing identifiable object motion allows us to better understand the contribution of motion relative to other visual attributes (such as color, form, and texture) in different regions of the temporal cortex. The use of human and tool stimuli allows us to assess the importance of visual motion in the brain's representation of different categories of objects.

In a previous study, we demonstrated that the lateral temporal cortex prefers video clips of moving objects to static images of the same objects. Within the lateral temporal cortex, the STS prefers video clips of human motion, whereas the MTG prefers video clips of tool motion. In addition, the STS responds more strongly to humans moving in an articulated fashion than to humans moving with unarticulated motion (Beauchamp et al., 2002). We hypothesized that if the lateral temporal cortex is specialized for processing complex object motion and the ventral temporal cortex is specialized for processing visual form, the lateral temporal cortex should respond similarly to video and point-light displays, while the ventral temporal cortex should prefer the extra form information contained in video stimuli. Within the lateral temporal cortex, we hypothesized that

Figure 1. Sample still frames from the four types of motion stimuli: video clips of real motion (A and B) and computer-generated point-light displays (C and D). Red arrows (not present in actual display) illustrate the direction of motion. Each stimulus lasted 2.5 sec (30 frames per second). The human motion clips (A) consisted of two actors performing different whole-body motions. The tool clips (B) consisted of tools performing their characteristic motion without visible manipulandum. The point-light displays (C and D) contained points that moved with motion vectors equivalent to the video stimuli. C is a still frame of a human standing from a seated position, D is a still frame of a hammer in the middle of a downstroke. Human and tool point-light displays were matched for luminance and velocity profiles. All displays contained a central fixation crosshairs on a gray background.


A **Human Motion** B **Tool Motion**

C **Human Point Light** D **Tool Point Light**
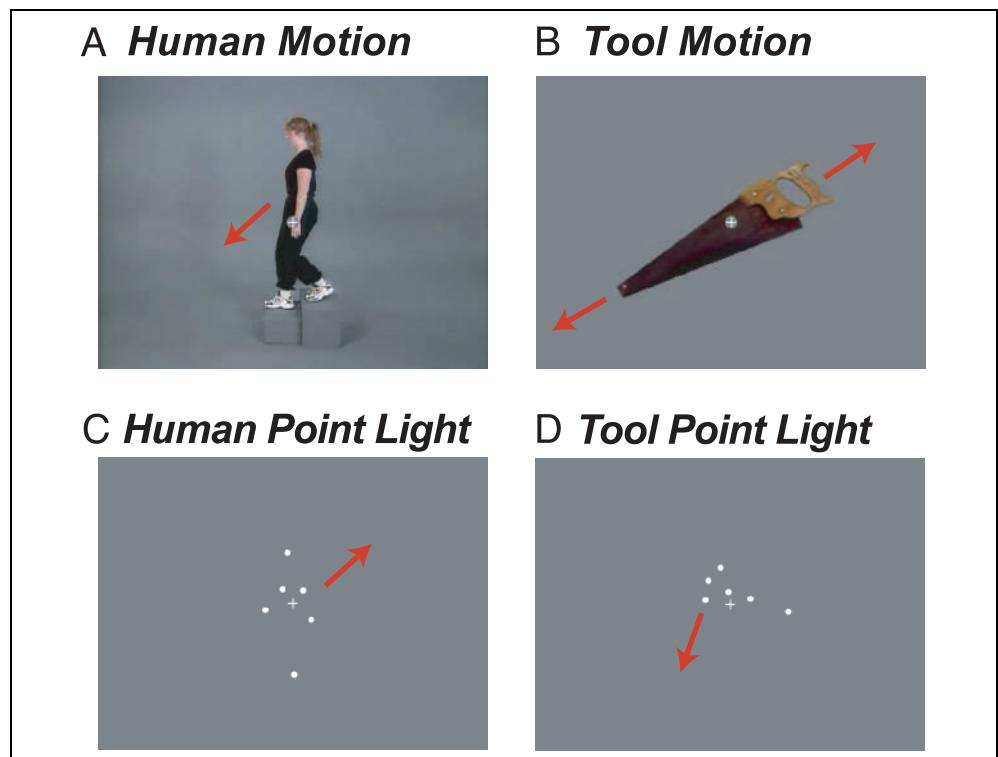
the STS should prefer human point-light displays while the MTG should prefer tool point-light displays.

RESULTS

A number of brain regions responded more strongly to video clips of humans than to video clips of tools and vice versa (illustrated for individual subjects in Figure 2A and B and for the group average dataset in Figure 2C). In the lateral temporal cortex, voxels that preferred tool videos to human videos were concentrated in the MTG and in the inferior temporal sulcus, while voxels that preferred human videos to tool videos were found in the superior temporal gyrus and sulcus. In the ventral

Figure 2. Activation maps for four single subjects (A and B) and the mixed-effects group map from nine subjects (C). Anatomical information is shown in gray scale, functional information is shown in color. Regions with a greater response to human or tool motion are colored by their preference for human motion (orange-to-yellow color scale) or tool motion (blue color scale); same color scale for (A), (B), (C). In (A) and (B) surface models of each subject's gray–white matter boundary were created and inflated. The gray-to-white shades illustrate curvature of the folded brain (black = depths of the sulcus, white = crowns of the gyrus). In (A), the left top panel shows a lateral view of the left hemisphere, the left bottom panel shows an inferior view of the left hemisphere, the right top panel shows a lateral view of the right hemisphere, and the right bottom panel shows an inferior view of the right hemisphere. In (B), lateral views of the left and right hemispheres are shown for three subjects. Dashed lines indicated the extent of the ROIs used to generate average timeseries in Figure 3. In (C), an ANOVA was performed across all subjects in each voxel to find significant differences between conditions. The results are shown overlaid on a volume rendering of a single subject's brain. The left panel shows a volume rendering of the left hemisphere, the right panel shows a volume rendering of the right hemisphere. IPS = intraparietal sulcus; LFG = lateral fusiform gyrus; MFG = medial fusiform gyrus; STS = superior temporal sulcus; MTG = middle temporal gyrus.
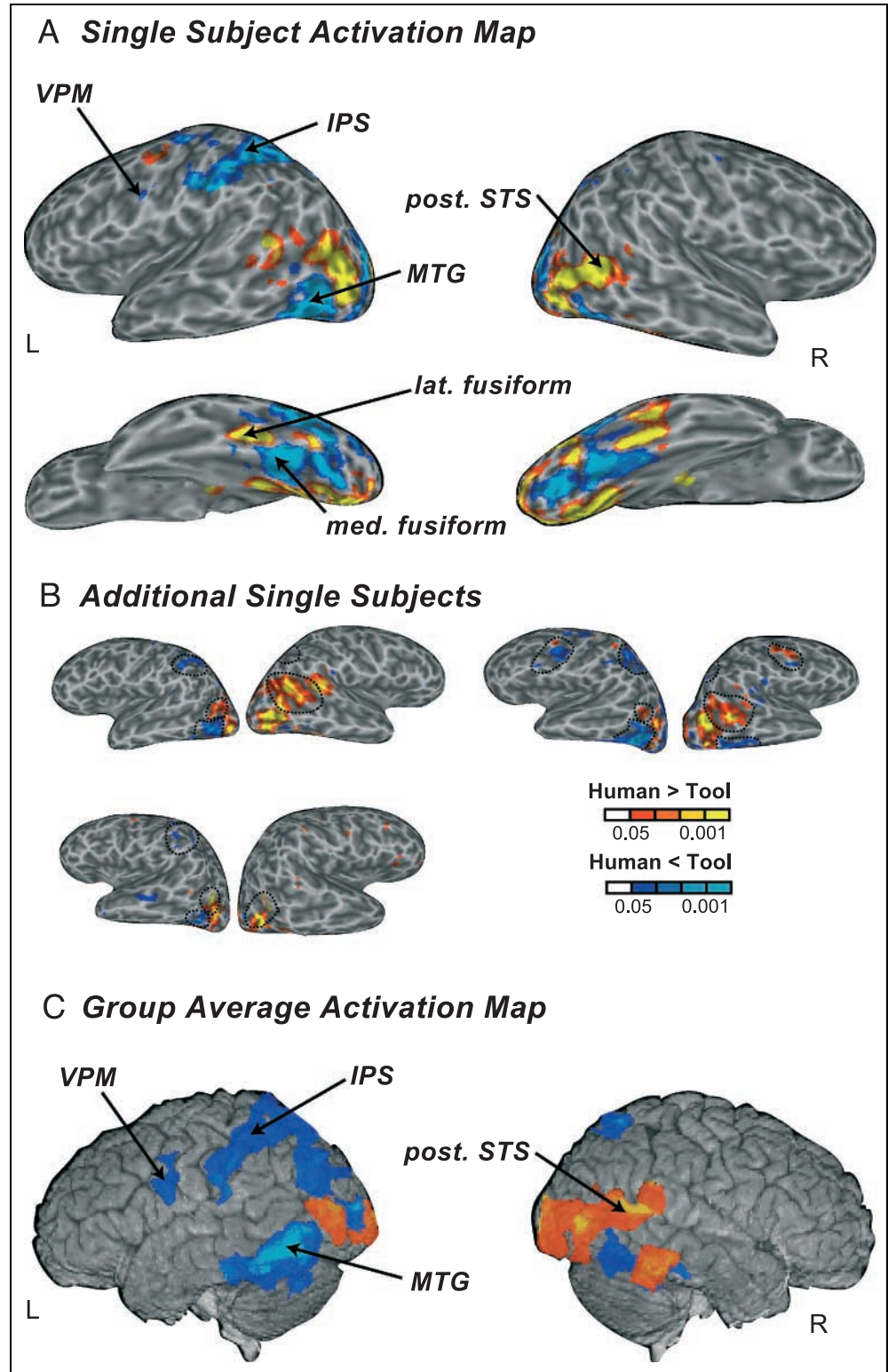
temporal cortex, the lateral fusiform preferred human videos whereas the medial fusiform preferred tool videos. In the parietal and frontal lobes, the intraparietal sulcus (IPS) and the ventral premotor cortex (VPM) preferred tool videos (summary of regions, with stereotactic coordinates, in Table 1).

The category-related responses were lateralized, with a significantly greater number of tool-preferring voxels in the left versus right MTG, the left versus right medial fusiform, and the left versus right IPS, with a significantly greater number of human-preferring voxels in the right versus left STS and the right versus left lateral fusiform (values in Table 1). Although the volume of tool or human selective cortex differed across hemispheres, the response magnitude was similar, so voxels from both hemispheres were grouped into bilateral regions of interest (ROIs) for further analysis (e.g., human video-preferring voxels in the left and right STS were grouped into a single STS ROI; see Methods section for details).

In order to determine how these areas responded to point-light displays of humans and tools, we created MR timeseries showing the response to the different stimulus types in each region, averaged across subjects (Figure 3). Voxels in the lateral temporal cortex that preferred human videos to tool videos (peak response 0.82% vs. 0.47%) were located in the STS. This cortex also preferred human point-light displays to tool point-light displays (0.69% vs. 0.48%, $p < 10^{-5}$). Lateral temporal voxels that preferred tool videos to human videos (0.69% vs. 0.46%) were located in the MTG. This region also preferred tool point-light displays to human point-light displays (0.68% vs. 0.54%, $p = .01$).

In the ventral temporal cortex, we also observed category-related responses that were consistent across display type. Voxels that preferred human to tool videos (0.69% vs. 0.39%) were observed in the lateral fusiform gyrus. The lateral fusiform also preferred human to tool point-light displays (0.40% vs. 0.31%, $p < 10^{-4}$). Voxels that preferred tool to human videos (0.62% vs. 0.43%) were located in the medial fusiform gyrus and also preferred tool to human point-light displays (0.40% vs. 0.28%, $p = .004$).
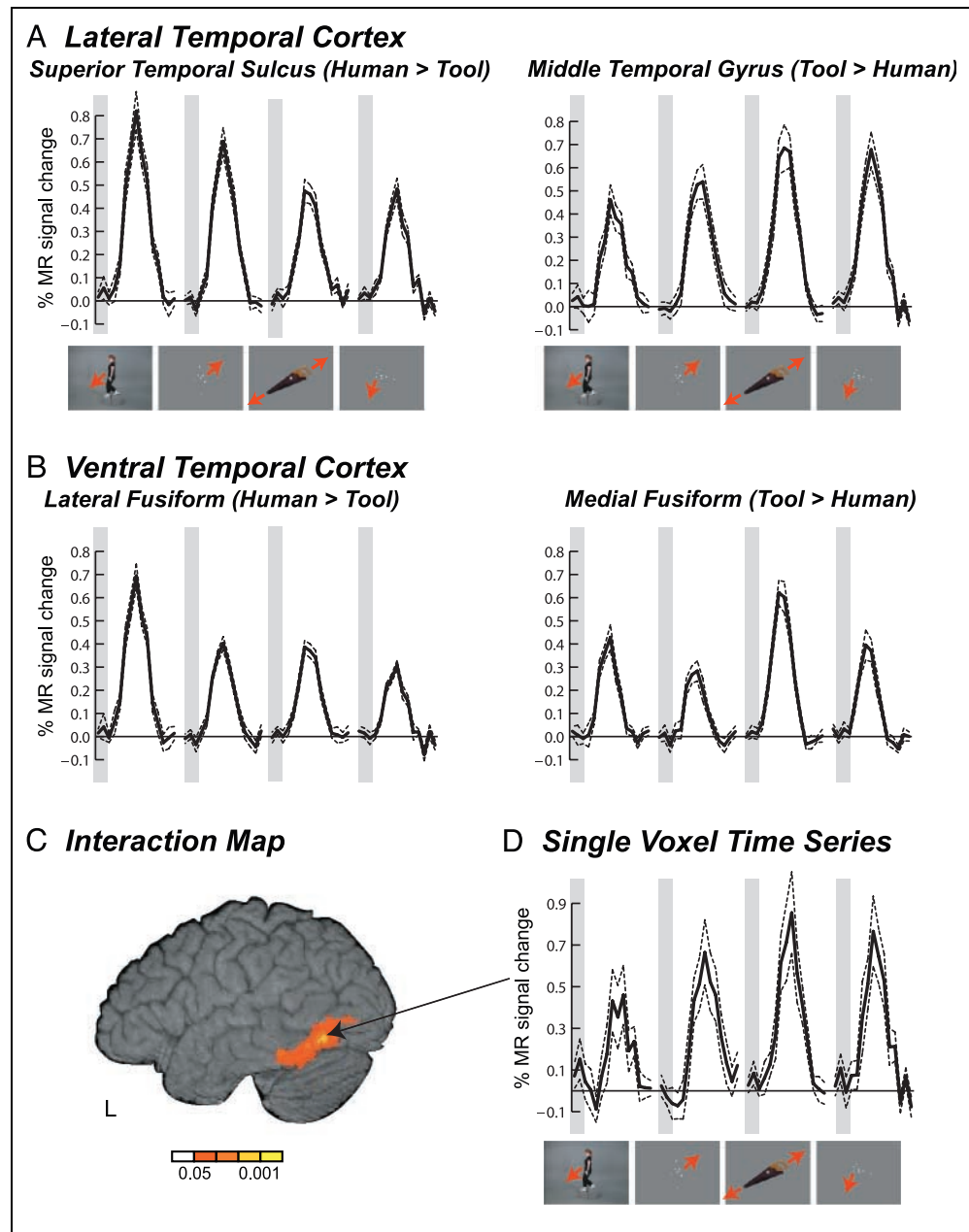
Both lateral and ventral temporal regions showed category differences that were consistent between video and point-light displays. However, the relative amplitude of response to video and point-light stimuli differed. In the ventral temporal cortex, the lateral fusiform cortex strongly preferred human videos to human point-light displays (0.69% vs. 0.40%, $p < 10^{-4}$) and tool videos to tool point-light displays (0.39% vs. 0.31%, $p = .007$), while the medial fusiform strongly preferred tool videos to tool point-light displays (0.62% vs. 0.40%, $p < 10^{-4}$) and human videos to human point-light displays (0.43% vs. 0.28%, $p = .002$). In the

Table 1. Talairach Coordinates of Activation Foci

Anatomical Description		Coordinates			Hemispheric Asymmetry (L vs. R)
		x	y	z	
Regions That Preferred Tool Video Clips to Human Video Clips					
Middle temporal gyrus and inferior temporal sulcus	L	−50	−62	−6	4037 vs. 1681,
	R	45	−63	−9	$p = .0002$
Medial fusiform gyrus	L	−21	−40	−16	4347 vs. 3401,
	R	28	−42	−24	$p = .001$
Intraparietal sulcus	L	−23	−37	51	6003 vs. 4272,
	R	25	−62	54	$p = .009$
Ventral premotor cortex	L	−48	2	21	2596 vs. 2413,
	R	41	−4	19	ns
Regions That Preferred Human Video Clips to Tool Video Clips					
Posterior superior temporal sulcus	L	−39	−59	15	2446 vs. 7752,
	R	47	−56	15	$p = .005$
Lateral fusiform gyrus	L	−36	−42	−11	1097 vs. 2641,
	R	44	−42	−15	$p = .003$

Coordinates are locations of peak significance in the group activation map. Hemispheric asymmetry was measured as the volume of activation (in mm^3) in the same ROI in the left versus right hemispheres; p values from a paired t test across subjects.

Figure 3. The MR response to each type of motion stimulus. The dark line in each graph shows the response to a single presentation of each stimulus (averaged across subjects) with the thin dashed lines illustrating one SE above and below the mean. The light gray bars overlaid on the graph illustrate the 2.5-sec duration of each stimulus, followed by the slow blood-oxygenation level dependent response to the stimulus. Stimulus categories are illustrated with sample stimuli underneath each response in (A) and (B): human video, human point-light, tool video, and tool point-light. In (C), brain regions that showed a significant interaction between stimulus category (human or tool) and display type (video or point-light); display as in Figure 2. In (D), the MR timeseries, averaged across subjects, from a single voxel (position indicated by arrow), coordinates (−50,−65,−4).

A **Lateral Temporal Cortex**
Superior Temporal Sulcus (Human > Tool)   *Middle Temporal Gyrus (Tool > Human)*

B **Ventral Temporal Cortex**
Lateral Fusiform (Human > Tool)   *Medial Fusiform (Tool > Human)*

C **Interaction Map**   D **Single Voxel Time Series**

lateral temporal cortex, the MTG showed an equal or greater response for point-light compared with video displays (69% vs. 0.68%, $p = .3$ for tool stimuli; 0.54% vs. 0.46%, $p = .04$ for human stimuli). The STS showed an intermediate pattern of response, with similar responses to tool video and tool point-light displays (0.47% vs. 0.48%, $p = .7$) but slightly greater responses for human video than human point-light displays (0.82% vs. 0.69%, $p = .04$).

Ventral temporal regions showed a strong preference for video compared with point-light displays, whereas lateral temporal regions responded similarly for the two display types. To quantify this interaction, we performed a two-way ANOVA (Area × Display type)

across subjects on the response to the preferred category in the lateral and ventral temporal regions. That is, we compared the response to human video and human point-light displays in the STS and lateral fusiform gyrus, and the response to tool video and tool point-light displays in the MTG and medial fusiform gyrus. We found a significant interaction between area and display type (STS vs. lateral fusiform, $F = 26.1$, $p = .001$; MTG vs. medial fusiform, $F = 52.6$, $p = 10^{-4}$) reflecting the greater response for video than point-light displays in ventral compared with lateral temporal regions (Figure 3).

To test more generally for interactions between stimulus category (human vs. tool) and display type (video

vs. point-light), we performed a voxel-wise ANOVA across subjects and tested the interaction [(tool video − tool point) − (human video − human point)]. The right lateral fusiform and the STS both displayed an interaction, but the strongest interaction effects were found in the left MTG, largely overlapping tool-selective regions (compare Figures 3C and 2B). As shown in the MR timeseries from the region of peak interaction (Figure 3D), this interaction arises from the MTG's similar response to tool video and tool point-light displays and significantly larger response to human point-light displays than to human videos.

DISCUSSION

An important finding of our study is that the lateral temporal cortex (STS and MTG) responds strongly to point-light displays and video displays containing equivalent amounts of visual motion. This suggests that visual motion is an important contributor to activity in the lateral temporal cortex. In contrast, the ventral temporal cortex (lateral and medial fusiform) responds weakly to point-light compared with video displays, suggesting that the additional form, color, and texture information found in video but not in point-light displays are important contributors to activity in the ventral temporal cortex.

Previous neuroimaging studies have shown that facial motion elicits a larger response in the STS than does static faces (Puce et al., 1998). Studies using point-light displays of whole-body motion compared with scrambled or inverted displays have also shown STS activation (Grossman & Blake, 1999, 2001; Vaina et al., 2001; Howard et al., 1996). In the present study, we observed strong responses to both video and point-light displays of human motion, confirming that the STS is a crucial brain locus for perceiving biological motion (Allison, Puce, & McCarthy, 2000). Although the STS preferred human stimuli, the MTG showed a greater response to stimuli containing images of tools, as in previous studies (Chao et al., 1999; Chao, Weisberg, & Martin, 2002; Devlin et al., 2002; reviewed in Martin et al., 2001; Martin, 1996). The present study extends this finding to show that the MTG also prefers tools even when they are represented only by impoverished point-light displays.

Why does the STS prefer human motion and the MTG prefer tool motion? In a previous study, the STS and the MTG responded more strongly to moving humans and tools than to static humans and tools (Beauchamp et al., 2002). However, visual motion could act in a nonspecific way to increase the category-related activity observed for static stimuli (STS for static biological stimuli and MTG for static tool stimuli). The effects of different types of motion were studied in an experiment in which subjects viewed humans moving with natural articulated motion vectors, tools moving with natural unarticulated motion

vectors, and humans and tools moving with artificial unarticulated motion vectors. The STS preferred articulated moving humans to unarticulated moving humans, whereas the MTG preferred unarticulated to articulated humans. To explain these findings, we proposed a model in which the STS is more sensitive to the articulated motion with many degrees of freedom characteristic of biological objects, and the MTG is more sensitive to the rigid-body unarticulated motion characteristic of manmade objects (Beauchamp et al., 2002). The results of the present study are consistent with this model. Human videos contained the most articulation, as limbs and trunk articulated about every joint. Human point-light videos contained less articulation: The point-light displays contained only a single marker on legs and arms (to equate the total number of points between human and tool point-light displays) so that legs and arms moved relative to the trunk, but articulation around their internal joints (elbow, wrist, knee, ankle) was not visible. Tool videos and point-light displays both contained little or no articulation. The STS responded most to human videos, responded less to human point-light displays, and responded weakly to tool video and point-light displays. The MTG showed large responses to tool video and point-light displays, responded less to human point-light displays, and showed the weakest response to human videos. These are precisely the pattern of responses that would be expected if neurons in the STS preferred articulated motion, whereas neurons in the MTG preferred unarticulated motion.

Other factors also likely contributed to the observed responses in the STS and the MTG. Because the STS is thought to integrate information about form, color, and motion (Puce, Castiello, Syngeniotis, & Abbott, 2003; Vaina et al., 2001; Oram & Perrett, 1996), human videos containing form and color would be expected to evoke greater responses in the STS than human point-light displays containing little form or color. However, this explanation does not account for the similar responses to tool video and tool point-light displays in the STS, nor for the greater response to human point-light than human video displays in the MTG. However, the preference of the STS for human point-light displays and the MTG for tool point-light displays cannot be definitively attributed to the motion properties of the displays, because motion properties and stimulus category were confounded in our experiment.

Although the STS preferred human videos to human point-light displays, the MTG responded equally to video and point-light tools, despite the greater amount of color and form information contained in the videos, suggesting that visual motion, not color or form, is the key determinant of activity in the MTG. This may explain the MTG's role in identifying tools, because motion, but not color or form, is an invariant property for many tools (different hammers may have very different colors or

shapes, but it is their rhythmic down-and-up striking motion that defines them as hammers). In addition to the MTG, the IPS and the VPM showed strong responses to tool videos, perhaps related to their role in grasping and manipulating tools (Chao & Martin, 2000; Chao et al., 2002; Grafton, Fadiga, Arbib, & Rizzolatti, 1997).

The human ventral temporal cortex is known to be sensitive to color (Beauchamp, Haxby, Jennings, & DeYoe, 1999; Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998; Zeki, 1990) and texture (Beason-Held et al., 1998). Object form is also an important attribute in the ventral temporal cortex, and is thought to be the crucial contributor to category-related activity (Martin et al., 2000). Initial reports described a single focus within the lateral temporal cortex selective for faces, known as the fusiform face area (Kanwisher, McDermott, & Chun, 1997) and a more medial focus selective for visual scenes (Epstein & Kanwisher, 1998). Additional experiments have shown that a broad swath of the lateral temporal cortex prefers biological forms such as faces, bodies, and animals to other visual stimuli, whereas the more medial cortex prefers forms of tools, buildings, and manipulable objects (Levy, Hasson, Avidan, Hendler, & Malach, 2001; Chao et al., 1999; Haxby et al., 1999; Ishai et al., 1999; Aguirre et al., 1998; Epstein & Kanwisher, 1998; Kanwisher et al., 1997). The idea that color, texture, and form are important determinants of response in the ventral temporal cortex is supported by the finding that ventral temporal areas respond similarly to static and moving complex objects (Beauchamp et al., 2002). The present study used point-light displays that did not contain color, texture, or form cues, and the ventral temporal cortex responded about 40% less to point-light displays than to video stimuli, which did contain these features. Although little form information is directly available in the point-light display, connectivity and shape can be inferred from the coherent motion of neighboring points. This information could be extracted by motion-sensitive neurons in the ventral temporal cortex (Cheng, Hasegawa, Saleem, & Tanaka, 1994; Dupont, Orban, De Bruyn, Verbruggen, & Mortelmans, 1994; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991) or it could be calculated by the lateral temporal areas and then transferred to the ventral temporal cortex via the abundant connections between the two regions (Ungerleider & Desimone, 1986). The finding that patients with lesions that damage area MT (but spare the STS and ventral temporal cortex) can still decode biological point-light displays suggests that ventral motion processing areas serve as important inputs to the STS (McLeod, Dittrich, Driver, Perrett, & Zihl, 1996; Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990). Information about the precise timing of lateral and temporal activations via electrical or magnetic measures should permit a better understanding of the integration of lateral and ventral motion processing during viewing of point-light displays (Singh, Barnes, Hillebrand, Forde, & Williams, 2002).

METHODS

Human Subjects and MR Data Collection

Nine subjects (3 women, 6 men, average age 29.6 years) underwent a complete physical examination and provided informed consent (World Medical Association, 1997). Subjects were compensated for participation in the study and anatomical MR scans were screened by the NIH Clinical Center Department of Radiology in accordance with the NIMH human subjects committee. MR data were collected on a GE Signa 3 Tesla scanner using standard imaging procedures. A high-resolution SPGR or MP-RAGE 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 msec. Twenty-four axial slices (slice thickness of 4.5 or 5.0 mm as necessary) with a TR of 2 sec were acquired, with 8–10 series for each subject and 132 brain volumes per series.

Visual Stimuli

Four types of complex motion stimuli were presented: video clips of moving humans, video clips of moving tools, point-light displays of moving humans, and point-light displays of moving tools. Each type of stimulus lasted 2.5 sec at 30 frames per second, followed by a 0.5-sec fixation for an effective stimulus duration of 3.0 sec. Single still frames from each stimulus category are shown in Figure 1. Sample stimuli may be viewed online at http://lbc.nimh.nih.gov/people/mikeb/stimuli.html. All stimuli were presented on a gray background and contained a central fixation cross overlaid on the stimulus.

Point-light displays were created using a motion capture system (Vicon Motion Systems, Lake Forest, CA) located in the NIH Clinical Center. Small reflective targets were attached to the joints of an actor performing different whole-body movements (jumping jack; stair climb and descend, leftwards and rightwards; walk, left and right; jog, left and right; soccer kick, left and right; sit-up; standing broad jump, left and right; stationary jog; rise from seated position to standing and vice-versa; touch toes and rise). The targets were illuminated with camera-mounted LEDs and their motion captured by six high-speed cameras. The position of each target in three-space was determined by triangulation and stored for later manipulation. A similar procedure was used to capture motion vectors for manipulable objects (chef's knife, hammer, needle-nose pliers, paintbrush, plastic drink pitcher, pliers, ratchet, wood saw, scissors, utility knife.) Real video clips of human actors and tools were collected on a gray background; tools moved realistically without visible manipulandum (details in Beauchamp et al., 2002). Point-light displays of moving humans and moving tools were equated (across all stimuli in each category) to ensure the same average retinotopic extent,

the same average number of moving points in each stimulus category, and the same average point velocity.

Visual stimuli were presented using MATLAB (Mathworks, Natick, MA) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) running on a Macintosh G4 (Apple Computer, Cupertino, CA). The source code for the stimulus program is freely available at http://lbc.nimh.nih.gov/people/mikeb/matlab.html. Stimuli were back-projected onto a Lucite screen using a three-panel LCD projector (Sharp, Mahwah, NJ). Stimulus presentation was synchronized with MR data acquisition using a DAQ board (National Instruments, Austin, TX) and subject responses were recorded using Super-Lab software (Cedrus, San Pedro, CA).

Experimental Design and Visual Task

Each of four stimulus types (video and point-light humans and tools) was presented in a rapid event-related design. In the rapid event-related design, stimuli from each category and fixation periods were presented in pseudorandom order. Within each 275-sec MR scan series, subjects viewed 14 trials of each stimulus type and 34 trials of 3-sec fixation periods, randomly ordered for optimal experimental efficiency (Dale, 1999). The combination of 3-sec stimuli with 2-sec time for brain acquisition allowed for an effective TR of 1 sec, allowing estimation of the hemodynamic response to a single stimulus of each type with 1 sec resolution (see below).

Behavioral Data

To ensure attention to the stimulus, subjects performed a simple two-alternative forced-choice discrimination, with subjects deciding if the stimulus contained a human or a tool. Subjects responded slightly faster to the tool videos (mean $\pm$ SD of 1089 $\pm$ 127 msec) than to the human videos (1155 $\pm$ 223 msec, $p = .01$) but showed no reaction time difference to the tool versus human point-light displays (1200 $\pm$ 113 vs. 1203 $\pm$ 107 msec, $p = .83$). Averaged across categories, responses to video stimuli were significantly faster than to point-light displays (1122 $\pm$ 146 msec vs. 1202 $\pm$ 107 msec, $p = .01$). Subjects performed the simple behavioral task with accuracy near ceiling (<1 error per stimulus category per scan series).

In a behavioral study performed outside the scanner, 10 subjects (none of whom had been previously exposed to the stimuli or participated in the fMRI experiment) viewed the set of moving point-light videos presented in the fMRI experiment or individual still frames selected from the videos. Subjects classified each moving or still stimulus as either ''human'' or ''tool'' in a two-alternative forced-choice task. Subjects were more accurate in decoding the moving point-light displays than the stationary point-light displays for both human stimuli (98% vs. 80%, $p = .009$) and tool stimuli (100% vs. 93%, $p = .013$). These results show that point-light

displays of tools, such as point-light displays of humans, carry information that is dependent on the pattern of motion of the coherent dots and is not entirely present in the static dot arrangements.

fMRI Data Analysis

MR data were analyzed within the framework of the general linear model in AFNI 2.50 (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.

The impulse response function to each stimulus category was estimated with 1-sec resolution using deconvolution (Glover, 1999). A separate regressor was used to model the response in each 1-sec period in a 15-sec 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 shape of the hemodynamic response.

The response magnitude to each stimulus type was calculated by summing the beta-weights of the regressors from the 4th through the 10th second of the response (capturing the positive blood-oxygenation level dependent response but not any poststimulus undershoot).

In order to minimize type I and type II errors, a two-step thresholding process was used (Chao et al., 1999; Haxby et al., 1999). First, a very high experimental-effects threshold eliminated voxels that did not respond to the stimulus. In voxels that did respond to the stimulus, a more liberal threshold was used for maximum sensitivity. Individual subject 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 at a threshold of $p < 10^{-6}$ to correct for the multiple comparisons produced by 20,000–25,000 intracranial functional voxels. Following stringent thresholding by the experimental-effect contrast, voxels were displayed using a color scale based on the significance of human videos versus tool videos contrast with a threshold of $p < .05$. 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.

To create group maps, a random-effects model was used. For each subject, the regression model provided a single estimate of the response to each stimulus type in each voxel. After stereotactic normalization, a two-way

mixed-effect ANOVA was performed on each voxel in standard space. Planned contrasts on stimulus type (point-light vs. video) and category (human vs. tool) were undertaken (fixed effect), with each individual subject serving as the repeated measure (random effect).

Region of Interest Identification

Any given cortical area spans multiple MRI voxels, whereas the exact anatomical location and size of the same functional area varies from subject to subject. To accurately describe the fMRI response, we created ROIs for six different cortical areas, defined individually for each subject (STS, MTG, lateral fusiform, medial fusiform, ventral premotor, and IPS). The response in each ROI was calculated separately for every subject, followed by computation of a grand mean across subjects. ROIs from three subjects are shown in Figure 2C.

First, anatomical ROIs were used to find all voxels in one of four anatomical neighborhoods (IPS, ventral premotor, lateral fusiform, medial fusiform). Then, functional criteria were used to find only active voxels within each neighborhood.

The IPS anatomical ROI extended 1 cm in each direction from the IPS from $z = 36$ to $z = 57$. A premotor ROI was created in the left VPM along the inferior portion of the precentral sulcus (extending 1 cm anterior and posterior) from $z = 18$ to $z = 36$. Lateral temporal ROIs extended (in the superior to inferior direction) from the superior temporal gyrus to the lower bank of the inferior temporal sulcus. Ventral temporal ROIs extended in the lateral-to-medial direction from the inferior termination of the middle temporal ROI in the inferior temporal sulcus to the collateral sulcus. Because anterior temporal regions are anatomically and functionally distinct from more posterior regions, lateral and ventral temporal ROIs were restricted to posterior regions, from the mid-temporal lobe to the posterior termination of the STS in the occipital lobe (approximately $y = 70$ mm).

Within these anatomically defined ROI, all voxels that showed an effect of visual stimulus compared with fixation (threshold of $p < 10^{-6}$) were considered for further analysis. Inspection of the group map revealed relatively homogenous activations within the premotor and intraparietal ROIs, and no further selection was necessary. Within the ventral temporal ROI, human-preferring cortex was always located in the lateral fusiform gyrus while tool-preferring cortex was always located in the medial fusiform cortex (Figure 2, see Haxby et al., 1999). Because no clear anatomical boundary exists between the medial and lateral fusiform, we used functional differences as a boundary between medial and lateral fusiform ROIs, so that voxels that preferred human videos versus tools videos ($p < .05$) were placed in the lateral fusiform ROI while voxels that preferred tool videos to human videos ($p < .05$) were

placed in the medial fusiform ROI. Similarly, no clear anatomical boundary was found in the lateral temporal cortex between the STS and the MTG, so human video-preferring voxels were placed in the STS ROI and tool video-preferring voxels were placed in the MTG ROI. Of course, using the human versus tool video contrast as a criterion for lateral and ventral temporal ROI selection precludes statistical analysis of this contrast in these ROIs (responses to point-light displays were not used in ROI creation and thus can be safely compared).

Every subject showed activation in the ventral temporal, lateral temporal, and IPS ROIs; 7 of 9 subjects showed activation in the premotor ROI.

Surface Modeling

Three-dimensional models of the cortical surfaces were constructed using FreeSurfer software (Cortechs, http://www.cortechs.net). One to five high-resolution MP-RAGE scans for each subject were collected and averaged. An automated segmentation routine then extracted the gray–white boundary and constructed a surface model, which was then inflated to allow inspection of active areas buried deep in the cortical sulci (Fischl, Sereno, & Dale, 1999). The overall model significance was thresholded and blurred with a spatial gaussian filter of root-mean-square width 8 mm before painting to the cortical surface. Only voxels intersecting surface nodes were mapped to the cortical surface. Surfaces were visualized using SUMA software (http://afni.nimh.nih.gov/ssc/ziad/SUMA).

REFERENCES

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

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 4,* 267–278.

Beason-Held, L. L., Purpura, K. P., Krasuski, J. S., Maisog, J. M.,

Daly, E. M., Mangot, D. J., Desmond, R. E., Optican, L. M., Schapiro, M. B., & VanMeter, J. W. (1998). Cortical regions involved in visual texture perception: A fMRI study. *Brain Research. Cognitive Brain Research, 7,* 111–118.

Beauchamp, M. S., Haxby, J. V., Jennings, J. E., & DeYoe, E. A. (1999). An fMRI version of the Farnsworth–Munsell 100-hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cerebral Cortex, 9,* 257–263.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron, 34,* 149–159.

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

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

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

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage, 12,* 478–484.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex, 12,* 545–551.

Cheng, K., Hasegawa, T., Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *Journal of Neurophysiology, 71,* 2269–2280.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: Functional anatomy by positron emission tomography. *Journal of Neuroscience, 11,* 2383–2402.

Cowey, A., & Vaina, L. M. (2000). Blindness to form from motion despite intact static form perception and motion detection. *Neuropsychologia, 38,* 566–578.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

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

Devlin, J. T., Moore, C. J., Mummery, C. J., Gorno-Tempini, M. L., Phillips, J. A., Noppeney, U., Frackowiak, R. S., Friston, K. J., & Price, C. J. (2002). Anatomic constraints on cognitive theories of category specificity. *Neuroimage, 15,* 675–685.

Dupont, P., Orban, G. A., De Bruyn, B., Verbruggen, A., & Mortelmans, L. (1994). Many areas in the human brain respond to visual motion. *Journal of Neurophysiology, 72,* 1420–1424.

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

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9,* 195–207.

Fox, R., & McDaniel, C. (1982). The perception of biological motion by human infants. *Science, 218,* 486–487.

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

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage, 6,* 231–236.

Grezes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., & Decety, J. (2001). Does perception of biological motion rely on specific brain regions? *Neuroimage, 13,* 775–785.

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

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience, 12,* 711–720.

Grossman, E. D., & Blake, R. (1999). Perception of coherent motion, biological motion and form-from-motion under dim-light conditions. *Vision Research, 39,* 3721–3727.

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

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35,* 1167–1175.

Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. H. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience, 1,* 235–241.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

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

Howard, R. J., Brammer, M., Wright, I., Woodruff, P. W., Bullmore, E. T., & Zeki, S. (1996). A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Current Biology, 6,* 1015–1019.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences, U.S.A., 96,* 9379–9384.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14,* 201–211.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

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

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature Neuroscience, 4,* 533–539.

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., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A., 92,* 8135–8139.

Martin, A. (2001). Semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging and cognition* (pp. 153–186). Boston: MIT Press.

Martin, A., Ungerleider, L. G., & Haxby, J. V. (2000). Category specificity and the brain: The sensory/motor model of semantic representations of objects. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 1023–1036). Cambridge: MIT Press.

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

Mather, G., & Murdoch, L. (1994). Gender discrimination

in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London B, 259,* 273–279.

McLeod, P., Dittrich, W., Driver, J., Perrett, D. I., & Zihl, J. (1996). Preserved and impaired detection of structure from motion by a ''motion-blind'' patient. *Visual Cognition, 3,* 363–392.

Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to ''biological motion'' stimuli. *Journal of Cognitive Neuroscience, 6,* 99–116.

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

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

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

Puce, A., Syngeniotis, A., Thompson, J. C., Abbott, D. F., Wheaton, K. J., & Castiello, U. (2003). The human temporal lobe integrates facial form and motion: Evidence from fMRI and ERP studies. *Neuroimage, 19,* 861–869.

Servos, P., Osu, R., Santi, A., & Kawato, M. (2002). The neural substrates of biological motion perception: An fMRI study. *Cerebral Cortex, 12,* 772–782.

Singh, K. D., Barnes, G. R., Hillebrand, A., Forde, E. M., & Williams, A. L. (2002). Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response. *Neuroimage, 16,* 103–114.

Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *Journal of Comparative Neurology, 248,* 190–222.

Vaina, L. M., Lemay, M., Bienfang, D. C., Choi, A. Y., & Nakayama, K. (1990). Intact ''biological motion'' and ''structure from motion'' perception in a patient with impaired motion mechanisms: A case study. *Visual Neuroscience, 5,* 353–369.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 11656–11661.

Wheaton, K. J., Pipingas, A., Silberstein, R. B., & Puce, A. (2001). Human neural responses elicited to observing the actions of others. *Visual Neuroscience, 18,* 401–406.

World Medical Association. (1997). Declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *Journal of the American Medical Association, 277,* 925–926.

Zeki, S. (1990). A century of cerebral achromatopsia. *Brain, 113,* 1721–1777.