Neuron, Vol. 35, 1167–1175, September 12, 2002, Copyright ©2002 by Cell Press

Brain Areas Active during Visual Perception of Biological Motion

Emily D. Grossman[1] and Randolph Blake
Vanderbilt Vision Research Center/Department of Psychology
Vanderbilt University
Nashville, Tennessee 37203

Summary

Theories of vision posit that form and motion are represented by neural mechanisms segregated into functionally and anatomically distinct pathways. Using point-light animations of biological motion, we examine the extent to which form and motion pathways are mutually involved in perceiving figures depicted by the spatio-temporal integration of local motion components. Previous work discloses that viewing biological motion selectively activates a region on the posterior superior temporal sulcus (STSp). Here we report that the occipital and fusiform face areas (OFA and FFA) also contain neural signals capable of differentiating biological from nonbiological motion. EBA and LOC, although involved in perception of human form, do not contain neural signals selective for biological motion. Our results suggest that a network of distributed neural areas in the form and motion pathways underlie the perception of biological motion.

Introduction

It is widely believed that primate vision comprises multiple visual areas organized into hierarchical pathways specialized for registering information about particular aspects of the visual scene (Felleman and Van Essen, 1991). Over the years, this overarching model has taken different forms, with some versions emphasizing distinctions between "sustained" and "transient" aspects of vision (Breitmeyer and Ganz, 1976; Kulikowski and Tolhurst, 1973), others distinguishing "color" and "broadband" channels (Schiller et al., 1990), and still others focusing on distinctions between perceiving objects and acting upon objects (Goodale and Humphrey, 1998). One currently popular version of this theory posits a so-called "motion" pathway extending into more dorsal aspects of extrastriate and posterior parietal cortex, specialized for registering information about the locations of objects and their movements within the visual scene, and an "object" stream pathway in ventral cortex involved in specifying information about the shapes and identities of visual objects (Ungerleider and Mishkin, 1982; Livingstone and Hubel, 1987). This particular version of the multiple pathway model has sparked a wealth of research aimed at testing the notion of object-grounded and motion-grounded neural systems (Haxby et al., 1991; Tanaka, 1996; Bradley et al., 1998; Kourtzi and Kanwisher, 2000).

While not disputing the notion of object-based and motion-based processing streams, several recent neural imaging studies have sought to determine the extent to which neural representations of objects are distributed throughout visual cortex. Sereno et al. (2002) found evidence for cue-invariant, 3D shape representations in multiple brain areas spanning object and motion pathways in anesthetized monkeys. Haxby et al. (2001) found that the distribution of brain activity associated with viewing faces and objects was widespread within ventral temporal cortex, leading these authors to downplay the importance of highly specialized neural areas in object recognition. We, too, have recently become interested in the question of distributed neural representations, in our case, representations associated with a particularly salient class of motion-defined forms, i.e., biological motion. In this paper, we report results from a brain imaging study that examines patterns of neural activity within multiple brain areas implicated in visual perception of bodies and body parts.

In these experiments, we have capitalized on a vivid, remarkable example of motion-defined shape: Johansson's "point-light" animation sequences (Johansson, 1973). These animations convey complex human activities using just a handful of dots placed on the joints of the human body. Single static frames resemble a meaningless cluster of dots portraying no hint of an object, human or otherwise, but when shown in rapid succession, these animated dots are grouped to create the perception of a human form engaged in a readily identified activity (Cutting et al., 1978; Ahlström et al., 1997; Neri et al., 1998; Mather et al., 1992). The compelling sense of human form created by the spatio-temporal integration of these local dot motions would seem to imply that "object" and "motion" pathways are together creating perception of an active person.

Brain imaging studies in humans have pinpointed a region on the posterior superior temporal sulcus (STSp) that is active when observers perceive biological motion in point-light animations (Bonda et al., 1996; Howard et al., 1996; Grossman et al., 2000; Vaina et al., 2001). It seems reasonable to place STSp within the motion pathway, based on its proximity to motion-responsive areas MT and MST (Suneart et al., 1999); moreover, STSp is far removed from ventral temporal cortex and, by implication, brain areas involved in form perception. Nonetheless, point-light animations portraying biological motion create compelling impressions of a class of recognizable *objects*, namely humans, so it is natural to suppose that this unique type of structure from motion also activates "object-selective" ventral stream mechanisms. Thus while STSp may be selectively activated when viewing biological motion perception, the entire network of brain areas involved in registering all aspects of these salient animations may extend to the form pathway.

We have examined this supposition by isolating brain areas generally believed to be involved in the perception of objects, including human body parts, and then measuring BOLD signals in those areas produced by viewing biological motion sequences. These areas are: (1) the

[1]Correspondence: e.grossman@vanderbilt.edu

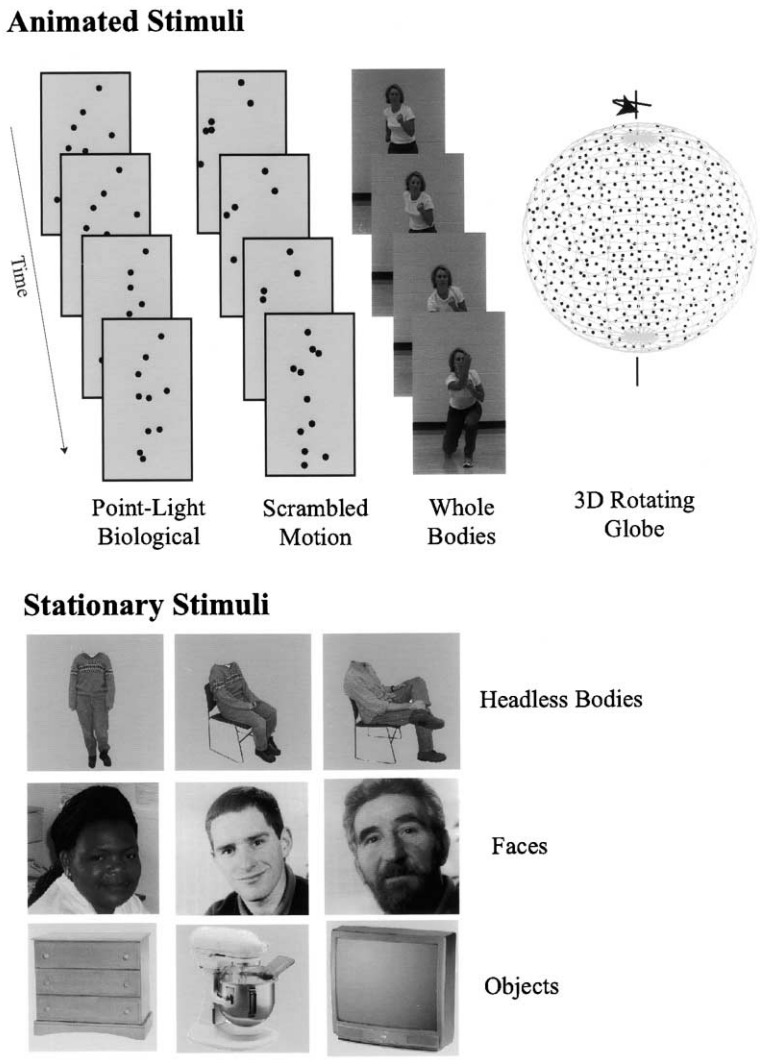

Figure 1. Schematics of the Stimuli

Dynamic human activity was portrayed by the motion twelve dots located on the joints and head of an actor performing various activities. Occlusions of some dots naturally occur as limbs pass behind the body. These occlusions help convey normal, three-dimensional body movement, and so were retained in the animations (i.e., not all dots were visible at all times). The scrambled animations contained the same motion vectors as the biological ones, but the initial starting positions of the dots are randomized within a region approximating the size of the body. The whole-body animations depicted an actor performing the same activities as the point-light animations. The rotating, structured globe was created by moving 200 dots sinusoidally within a circular aperture. The variable speed of the dots conveyed three-dimensional structure, but the direction of the rotation was ambiguous. The wire frame pictured above to denote three-dimensional structure was not visible in the experiment. The stationary, headless body condition consisted of images of bodies standing or sitting, with the heads erased. Observers also viewed images of faces and common household objects.

lateral occipital complex (LOC), which has been implicated in the perception of forms regardless of the visual cues used to define those forms (Grill-Spector et al., 1999); (2) the extrastriate body area (EBA), which has recently been implicated in the perception of images of bodies and body parts, but not faces (Downing et al., 2001); and (3) the occipital (OFA) and fusiform face areas (FFA), both of which have been implicated in the perception of faces (Kanwisher et al., 1997), as well as in the perception of other highly familiar objects (Gauthier et al., 1999). We used widely accepted stimuli and subtraction conditions to isolate these object-responsive brain areas, including pictures of headless bodies to activate the EBA, pictures of objects to activate the LOC, and pictures of faces to activate the FFA and OFA (Figure 1).

We also measured neural activity in STSp, the brain area implicated in perception of biological motion. In localizing STSp, observers viewed point-light actors performing a variety of human activities. However, these animations, composed of only 12 dots, are visually sparse in comparison to the pictures of headless bodies, faces, and objects. To evaluate the consequence of this sparseness on BOLD signal levels, we included animations in which the entire figure of the actor was visible. Also, to determine if neural responses were due to the presence of an object, or specifically because the object is biological, we included animations of a nonbiological object (a rotating, 3D globe) defined solely by dot motions.

Results

Our strategy entailed several steps: (1) use standard subtraction conditions to localize STSp, EBA, LOC, OFA, and FFA (Figure 2A), (2) assess patterns of activation across all of these areas by comparing evoked BOLD signals to a common baseline of fixation, and (3) determine whether neural signals within those regions are capable of discriminating between biological and scrambled point-light animations. To control for attention, which is known to modulate BOLD signals in many of the neural areas included in this study (Corbetta et al., 1991; Wojciulik et al., 1998), observers performed a 1-back task on the individual stimuli within each block.

Results from our measurements are discussed brain region by brain region in the following sections and are

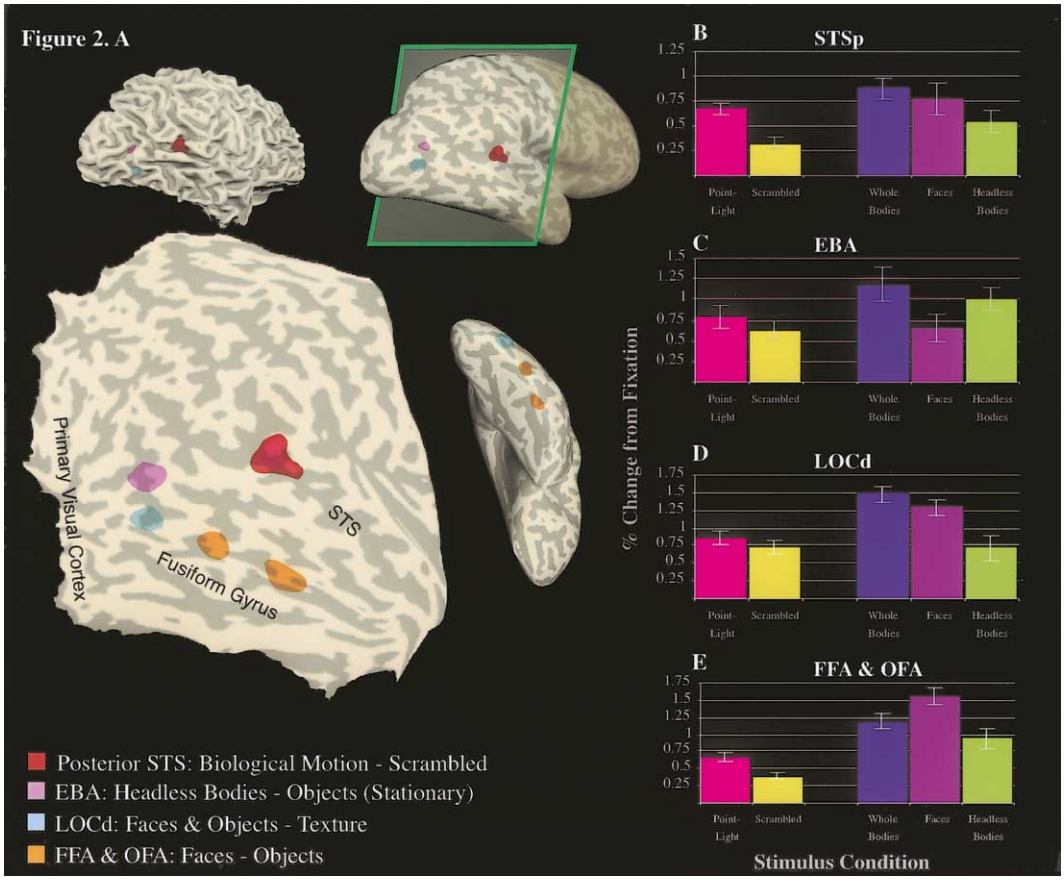

Figure 2. Summary of Regions of Interest (ROIs) and BOLD Responses in Biological Related Stimulus Conditions

(A) The ROIs in the right hemisphere of observer D.L. are displayed on the lateral and ventral surfaces of the gray matter. A cut, as indicated by the green plane, was made and the posterior end of the cortex flattened. We examined BOLD signals in four regions of interest: STSp (red), EBA (purple), LOCd (blue), FFA and OFA (orange). (B–E) The average BOLD activity levels for these ROIs (with FFA and OFA averaged) during the stimulus conditions depicting some kind of biological object, or the scrambled biological motion vectors. These stimulus conditions included animations of point-light biological motion (pink), point-light scrambled motion (yellow), whole-bodies (dark purple), pictures of faces (magenta), and stationary images of headless bodies (green). The percent change activation levels are relative to a fixation baseline. Error bars indicate 1 standard error.

shown in Figures 2B–2E. Using some of the same data from Figure 2, Figures 3 and 4 highlight contrasts important for evaluating differences in selectivity among the areas.

The Posterior Superior Temporal Sulcus

STSp is located at the posterior end of the superior temporal sulcus, near the junction of the STS and the inferior temporal sulcus (ITS). This brain area responds

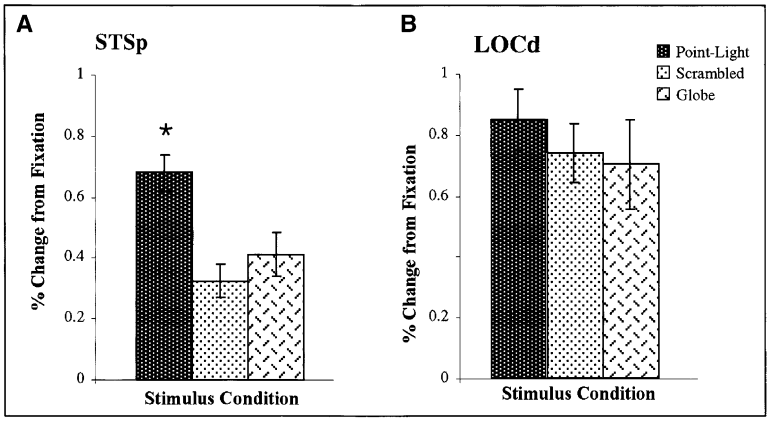

Figure 3. Average Percent Change in the STSp and LOCd ROIs for the Animated Dot Displays

Point-light biological, scrambled, and the rotating globe were the visually sparsest of all the stimuli used in these experiments. The biological and object identities were constructed by the movements of dots only, and could not be ascertained from a single, stationary frame. Even when animated, the scrambled animation looks like an incoherent cloud of dots. Asterisk indicates a significant difference for the biological motion condition over both the scrambled and rotating globe. All other contrasts were nonsignificant.

Table 1. Frequency of Activated Regions during the Point-Light Biological Motion Localizer

	Num Observed Hemispheres		Talaraich Coordinates					
			Left			Right		
ROI	Left	Right	X	Y	Z	X	Y	Z
STSp	10/10	9/10	−41.3	−52.8	11.8	46.1	−48.5	12.4
ITS	6/10	5/10	−43.1	−58.5	11.2	48.8	−53.3	10.8
STSa	2/10	6/10	−46.9	−41.9	6.6	50.0	−33.0	4.3
FFA	3/10	4/10	−33.3	−40.5	−13.9	37.1	−36.0	−12.2

A total of ten observers viewed the biological and scrambled point-light animations. Simulations of false positive rates were used to determine the threshold for the activation maps. Table indicates the number of observers and Talairach coordinates for each hemisphere of the activated clusters. Abbreviations: STSp = superior temporal sulcus, the posterior end, ITS = inferior temporal sulcus, specifically the small region of the ITS between the STS and MT+, STSa = anterior region on the superior temporal sulcus, FFA = fusiform face area.

more strongly to point-light animations portraying biological motion than it does to "scrambled" animations, created with the same motion vectors but whose starting positions are spatially randomized (Grossman and Blake, 2001). STSp does not respond well to coherent, translational motion, nor does it respond to the presence of kinetic boundaries (Grossman et al., 2000). In other words, this region is functionally and anatomically distinct from other motion-responsive areas in the human brain, including human MT+ (Watson et al., 1993; Tootell et al., 1995; Orban et al., 1995) and the kinetic occipital region (KO, or LO, as it is sometimes called; Van Oostende et al., 1997; Malach et al., 1995).

In the present study, observers viewed blocks of 1 s animations of point-light biological motion interleaved with blocks of 1 s animations of scrambled motion. In nine of the ten observers, a bilateral region on the posterior end of the STS, at the junction with the ITS, was more active during the biological epochs than during the scrambled epochs ($p < .001$). In the tenth observer, this region was only found in the left hemisphere (Table 1). Based on the anatomical location and the functional response during the localizer condition, these regions became the STSp region of interest (ROI) for each observer.

Within these ROIs, we compared BOLD signals during the four stimulus conditions depicting biological events (Figure 2B). Two of these conditions—whole body and point-light biological—were animated sequences depicting human activities, while the other two conditions—headless bodies and bodiless heads—were stationary images. BOLD activity levels were highest during the epochs of whole bodies, point-light bodies, and faces, and lowest during the epochs of headless bodies. Pairwise comparisons revealed that only the difference between whole bodies and headless bodies was significant ($p < .001$). It is noteworthy that bodies defined by only twelve points of light were as effective as whole body animations in activating STSp. The is the only brain area in which these two dynamic biological motion animations resulted in equivalent neural responses. This is testimony to the vividness of perception produced from these simple, sparsely sampled animations and to the importance of STSp in the perception of dynamically defined complex activity.

In contrast to BOLD signals evoked by biological motion, the rigidly rotating globe defined by moving dots evoked trivially small BOLD signals that were no different from those found during the scrambled motion epochs (Figure 3A). The weak, nonspecific neural response to the kinetic globe implies that STSp is not simply registering structure from motion but, instead, is specialized for the kinematics portraying biological motion. At the same time, motion is not absolutely crucial for activating STSp, for pictures of faces also produced reliable responses from this area. Face-responsive regions on the STS have been previously reported in the literature (Chao et al., 1999; Puce et al., 1998; Hoffman and Haxby, 2000), and it is possible that the face-responsive STS region (STS-FA) and STSp overlap, resulting in strong BOLD responses to the images of faces. Using the FFA localizer (faces–objects), we attempted to determine the extent of overlap between STSp and STS-FA. However, we were able to localize STS-FA in only one observer, and there was no overlap between the two ROIs. Further testing is needed to determine conclusively the relationship between the face- and body-responsive regions on the STS.

The Extrastriate Body Area

Following the lead of Downing et al. (2001), we localized the EBA by subtracting activation evoked by pictures of stationary headless bodies from activity evoked by pictures of stationary household objects. Using this subtraction, we were able to localize a region in the anterior extent of occipital cortex, dorsal to the inferior occipital sulcus. This region was bilateral in five of the six observers tested, and only in the left hemisphere of the sixth (Talairach coordinates: left hemisphere: −39.3, −70.1, 13.5; right hemisphere: 40.6, −65.7, 10.6).

Results for EBA are summarized in Figure 2C. Consistent with previous reports, this region was most active during stimulus conditions in which pictures of the human body were shown ($p < .05$), which included the epochs of stationary headless bodies and animations of whole bodies (including the heads). We found that faces alone were slightly less effective (though not significantly different) in activating this region. Body shapes depicted by the point-light sequences were also slightly less effective than the explicit body images in activating the EBA.

Both STSp and EBA have been implicated in the perception of bodies, albeit using very different visual stimuli. How are there two areas different? To answer this question, it is interesting to compare the responses of EBA and STSp during the two localizers used to identify

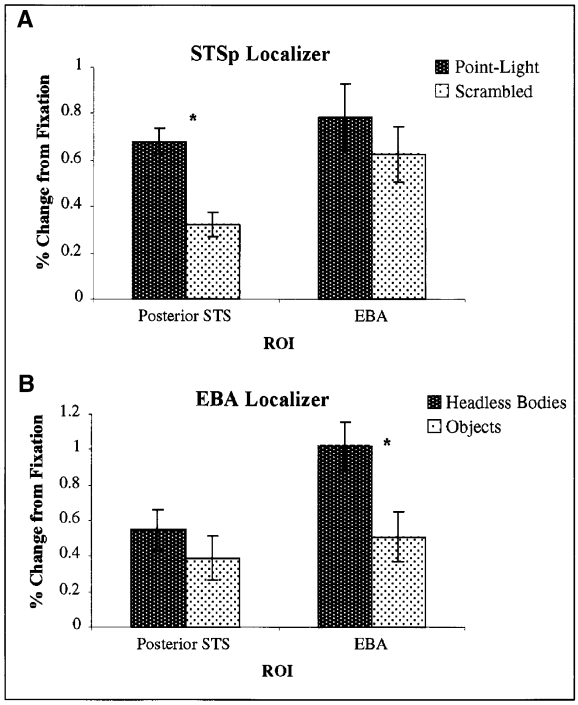

Figure 4. Average Percent BOLD Signal Change during the STSp and EBA Localizer Scans

(A) Subtracting BOLD signal levels during the point-light scrambled (light bars) from point-light biological (dark bars) motion intervals results in a significant contrast in STSp, but not in EBA. (B) However, subtracting activity levels elicited by pictures of household objects (light bars) from that elicited by pictures of stationary headless bodies (dark bars) results in higher percent signal change and more contrast in EBA.

these regions (Figure 4). BOLD signal levels in STSp during point-light biological animations were significantly higher than during the scrambled intervals ($p < .01$), but this was not true for EBA ($p = .41$). Conversely, during the EBA localizer, the images of headless bodies produced significantly more activation in EBA than did images of objects ($p < .01$), but there was no significant difference between the two in STSp ($p = .53$). It is unlikely that the poor discrimination between headless bodies and objects in STSp is due to the absence of a face in the headless bodies images. Quite the opposite, this region responded equally well to animated displays with no facial information (i.e., the point-light sequences) and to images of faces. Instead, we believe it is the absence of dynamic signals, not the absence of a face, that is responsible for STSp's weak response to images of headless bodies.

The Lateral Occipital Complex (Dorsal)

The lateral occipital complex (LOC) is a large area of cortex starting most dorsally at the posterior end of the lateral occipital sulcus and extending through the ventral temporal cortex. We localized this region using the technique of Kourtzi and Kanwisher (2000), in which neural activity during periods of viewing texture patterns is subtracted from activity associated with viewing faces and objects. In our study, this subtraction revealed a large region of activation, the most posterior and dorsal extent of which we will refer to as LOCd (Talairach coordinates: left hemisphere: −28.4, −77.8, −2.5; right hemisphere: 33.9, −74.7, −1.0). The more anterior extent of the activated region lies on the ventral surface of the temporal lobe and encompasses other object- and face-responsive brain areas on the posterior fusiform gyrus, including the occipital face area (OFA). Based on previous work showing clear functional differences between anterior and posterior regions of LOC (Grill-Spector et al., 1999; Bar et al., 2001), we felt it important to analyze the BOLD signals separately for the LOCd (discussed in this section) and the OFA (discussed in the following section).

BOLD signal levels in the LOCd were significantly higher when observers viewed faces and objects than when they viewed images of texture patterns created from scrambling the face and object pictures ($p < .001$). BOLD signals in LOCd were highest during the conditions showing whole-body animations and face images (no significant difference between the two conditions, $p = .29$), almost twice that for conditions when observers viewed stationary pictures of headless bodies or the point-light animations depicting biological motion (Figure 2D; $p < .0001$). There were no significant differences between BOLD signal levels while observers viewed point-light biological motion, scrambled motion, or the structured rotating globe animation (Figure 3B).

The Occipital and Fusiform Face Areas

Situated more anterior than LOC on the ventral surface of the temporal lobe is a cluster of foci related to object and face recognition. The most widely studied of these regions is the fusiform face area (FFA), located on the ventral medial aspect of the temporal lobe abutting the cerebellum. The FFA is functionally identified by its greater response to images of faces than to common household objects (Kanwisher et al., 1997). There is also another, more posterior region on the fusiform gyrus that responds strongly to faces and objects, and may lie within or just anterior to the LOC. This occipital face area (OFA), like the FFA, is often found by contrasting activation when observers view pictures of faces versus pictures of objects. We were able to localize the FFA in both hemispheres of all five observers tested using this contrast (Talairach coordinates: left hemisphere: −34.0, −39.7, −15.3; right hemisphere: 37.6, −38.9, −14.2); the OFA was localized in both hemispheres of three of these five observers, and only the left hemisphere of a fourth (Talairach coordinates: left hemisphere: −35.0, −55.0, −11.7, right hemisphere: 37.5, −54.2, −9.9). Across the stimulus conditions tested, we found no differences in the patterns of activity levels between the FFA and OFA, with the exception of overall slightly higher percent signals changes in the OFA. The results from these two regions, therefore, were averaged and are presented together.

Of all the stimulus conditions tested, the images of faces alone produced the highest responses in the fusiform areas (Figure 2E). This "face-only" activation was significantly higher than that produced when observers viewed pictures of objects ($p < .001$), headless bodies ($p < .01$), or whole bodies with the faces intact ($p < .05$).

Although faces also activated the fusiform face areas significantly more than the animations of point-light biological motion ($p < .001$), the contrast between point-light biological and point-light scrambled was also significant ($p < .001$). Critically, although activation levels are overall lower, the biological organization of the point-light animations is sufficient to activate this region, as evidenced by the contrast with scrambled animations. Incidentally, our failure to identify reliable FFA activation in an earlier study (Grossman et al., 2000) may be attributable, at least in part, to non-optimal slice placement for capturing the ventral temporal cortex in all observers. It is also noteworthy that in the present study, unlike the earlier one, FFA was localized using the standard stimulus contrast (faces minus objects).

Discussion

People watch other people all the time, trying to deduce intentions and moods based on dynamic visual information. Indeed, recognizing what others are up to is one of the most important perceptual activities we engage in. It is not surprising, therefore, that people are remarkably adept at perceiving intentions and affect even when those personal characteristics are portrayed in point-light animations devoid of static form cues. Befitting such a crucial perceptual skill, neural representations of biological activity are widely distributed throughout visual cortex. Based on converging lines of evidence, Allison et al. (2000) propose that a major component in this distributed representation is a large expanse of cortex spread across the STS.

The results from our study, while confirming the importance of the STS in perception of biological activity, reveal that neural areas in the ventral stream are also activated when one views point-light biological motion animations. The fusiform and occipital face areas in ventral temporal cortex contain neural signals capable of discriminating between point-light animations that organize into biological motion and those that organize into scrambled motion. Our results also demonstrate that other neural areas previously found to be selectively involved in the perception of bodies and objects, specifically LOC and EBA, can not discriminate between biological and scrambled point-light animations. To determine the discriminability of the neural signals within these five regions—STSp, FFA, OFA, EBA, and LOC—we probed each with a variety of biological and nonbiological visual stimuli. What conclusions can we draw from the patterns of activations in these distributed areas? To answer this question, we need to reconsider the response selectivity of these neural regions.

In the ventral stream, point-light bodies certainly produce weaker responses than those evoked by faces or, for that matter, by animated whole bodies in FFA and OFA. However, when observers view point-light bodies, the fusiform activations are significantly stronger than those resulting from the same point-lights scrambled to destroy the impression of a human body. In principle, then, neural activity in the fusiform region could form part of a distributed representation of human bodies, including bodies defined by motion. But how does this "object" pathway region—the FFA—acquire its selectivity for motion-defined bodies? After all, the human form in point-light animation sequences is portrayed exclusively by motion signals associated with hierarchical, pendular motions of the limbs: perception of a human form emerges only from the integration of motion signals over space and time. To the extent that the motion analyses underlying point-light animations are performed exclusively by neural mechanisms within the motion pathway, our results would imply that the outputs of those mechanisms project to the FFA; this idea, incidentally, is one component of a recently published model of face recognition (O'Toole et al., 2002). Of course, it is possible that FFA itself contains neural machinery for registering the motion signals necessary for constructing the body representation in point-light animations. From our results, we cannot distinguish between these two hypotheses, and to do so may require using analytic techniques that reveal the strength of functional connectivity among areas (Friston et al., 1995).

In contrast to FFA and OFA, LOC has been characterized as a general object recognition area, responding invariantly to images that can be interpreted as shapes regardless of the visual cues creating the shapes' contours (Grill-Spector et al., 2001). Our results, however, place some limits on the generality of that characterization. In particular, point-light animations, although readily organized into a visible human form, produced no greater activation in LOCd than did scrambled animations which resembled a disorganized cloud of dots. Similarly, the rotating globe, also easily organized into a structured shape, did not increase brain activity beyond levels found during viewing of scrambled motion. Evidently, the cue invariance of LOC does not extend to all forms defined by motion.

Finally, biological and scrambled motion produced equivalent BOLD responses in EBA, implying that this brain region does not carry signals capable of registering the presence of human forms as depicted in point-light animations. This is not to say that EBA does not register the presence of a human form—indeed, this area is functionally defined by its stronger response to images of headless bodies than to nonhuman objects (Downing et al., 2001). Nor do our results imply that EBA is not involved in viewing bodies in motion, as evidenced by the equivalent BOLD responses to whole bodies in motion and to images of stationary, headless bodies. Instead, our results indicate that the BOLD signals in EBA are contingent upon the shape of the body being explicitly represented in the image.

It is interesting to note that STSp behaves in a fashion complementary to EBA. Unlike in EBA, the BOLD responses in STSp are selectively driven by the dynamics of the human form. This is evidenced by (1) the equivalent BOLD responses produced by stationary images of bodies and by nonhuman objects, and (2) the stronger BOLD responses produced by dynamic human bodies than by a moving, nonhuman object. Together, these results imply that STSp is specialized for a particular class of dynamic events, namely moving human bodies. Also, unlike in EBA, the BOLD responses in STSp do not depend on the explicit representation of the human body, for STSp responded just as strongly to point-light animation sequences as to animations showing whole bodies in motion.

To end on a speculative note, it is reasonable to wonder why the human brain would contain multiple areas (e.g., EBA and STSp) dedicated to the perception of human bodies. Perhaps identification of an individual constitutes a different perceptual task than perceiving an individual's mood or intentions. Intentions can be judged based on gestures, actions, and expressions independent of identity. Accurate identification, on the other hand, must generalize across gesture, mood, and activity. Given these divergent demands, extraction of the visual information subserving identification and perception of intention may require different neural operations, perhaps most efficiently embodied in separate neural architectures (EBA and STSp, respectively).

Experimental Procedures

Participants

Ten individuals (6 men, 4 women) with normal or corrected to normal vision participated in this study. All observers had experience viewing point-light animations and easily recognized all the biological motion sequences as human activities. The observers gave informed consent as approved by Vanderbilt University Institutional Review Board.

Stimuli

Visual Stimuli were displayed using Matlab (Mathworks, Inc.) together with routines from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Point-light biological motion sequences were created by videotaping an actor performing various activities, including running, jumping, throwing, and kicking. The segments were digitized, and the joint positions in each frame were encoded as motion vectors with initial starting positions. Scrambled biological motion animations were created by randomizing the starting positions of each joint within a region approximating that covered by the biological sequences. The motion vectors were left intact so that the joints moved naturally as they would in the biological animations. This manipulation ensured the individual motion components were identical in both the biological and scrambled animations; only the hierarchical relations among the dots in the scrambled displays were destroyed. For both kinds of animations, the joints were displayed as small, black dots subtending approximately 9 arc min of visual angle against a gray background. Each biological activity sequence consisted of 20 frames displayed in a 1 s interval (33 ms interframe interval, 60 Hz). At this display rate, the biological animations generated smooth apparent motion, depicting natural body movement.

Whole body animations depicted the same activities as the point-light sequences, with the entire body of the actor being visible. These animations sequences were recorded with a digital video camera, edited into 1 s clips that were then converted to Quicktime files displayed with Matlab QT routines.

Random dot cinematograms depicting a structured "globe" and coherent planar motion were created with 200 dots horizontally displaced within a circular aperture. Approximately half the dots moved leftwards within the aperture while the other half moved rightwards. To create the structured globe, the speed of the dots was sinusoidally modulated such that the dots moved fastest on the outer edges of the aperture and slowest in the center. These animations create the vivid impression of a 3D transparent globe rotating about its vertical axis (Doner et al., 1984). Two coherent planes of transparent motion were created by moving half the dots leftward and half rightward, with the constant speed throughout the animation. The average speed of the dots was 4.8°/s (48 ms interframe interval, 60Hz), and the aperture subtended 7.2° × 7.2° of visual angle.

Some observers also viewed grayscale images of faces, headless bodies, and common household objects. These images were static, not animated, and they subtended 7.5° × 7.5° of visual angle. Texture patterns were created by breaking the images into small pieces that were then spatially shuffled to yield scrambled images lacking coherent spatial structure.

All individuals participated in the STSp localizer experiment in which point-light biological and scrambled motion animations were presented alternately in a block design. The resulting images from this scan were used to determine biological motion responsive ROI. All individuals also viewed alternating blocks of biological, scrambled, and whole-body motion, and in a separate scan the structured rotating globe and planar transparent motion. For five individuals, epochs of point-light biological motion were included within the structured globe and planar motion scan to eliminate any effect of neural adaptation to the coherent dots and still maintain the attention of the observers. We found no difference in BOLD signal levels for the structured globe or planar motion in the two condition or three condition scans. Six of the ten observers (D.L., D.R., F.R., J.L., K.B., M.R.) participated in additional scans designed to localize the EBA (alternating blocks of stationary headless bodies and objects). Five observers (D.L., D.R., F.R., K.B., M.R.) were also scanned to localize LOC and the FFA (alternating blocks of faces, objects, and texture patterns).

Imaging

All brain images were collected on a 3 Tesla GE Signa scanner located within Vanderbilt University Medical School. Observers participated in scanning sessions that lasted approximately 1.5 hr. During this time, we acquired high resolution T1 anatomical images of the observer's head (124 slices, 1.4 × 1.4 × 9375 mm), T2 functional images in the axial plane (single-shot EPI, TR = 1500 ms, TE = 25 ms, flip = 90°, 21 axial slices, 3.75 × 3.75 mm inplane, 5 mm, no gap), and T1 high resolution images of the slice positions. Slices positions were chosen to cover the entire occipital pole, the ventral surface of the temporal lobe, and the posterior extent of the superior temporal sulcus . Because the high resolution slice anatomy images were collected in alignment with the T2 functional images, the high and low resolution images were naturally co-registered with each other. The slice anatomy images were manually aligned with the T1 whole-brain images, subsequently allowing us to register the functional images into the whole-brain coordinate space. Using this alignment method, we were able to have some individuals return to the scanner on different days, align the data into common space, and treat the images as if they had been acquired on the same day.

Functional scans lasted 172.5 s, the initial 6 s (4 volumes) of which were discarded prior to analysis to allow for MR stabilization. The images and animations were blocked into 10.5 s epochs consisting of seven stimuli (1 s each, 500 ms interstimulus interval). The exemplars within the block were chosen randomly on each trial, but with a 50% chance of repeating on successive trials. Observers were instructed to monitor the stimuli carefully, and indicate sequential repetitions with a button press (1-back task), and the task was made more taxing by requiring observers to respond within the short 500 ms interval between the animations. This kind of challenging task maintains the observer's level of attention through a block of trials, an important consideration in view of evidence that attention can modulate BOLD signal in early and extrastriate visual areas (Somers et al., 1999; Watanabe et al., 1998). During the intervals involving the structured globe and planar motion, the near and far surfaces of the flowfield were inherently ambiguous, resulting in spontaneous reversals of depth ordering. Consequently, a 1-back task would have had no objective measure of correctness. To promote attention in these epochs, observers were instructed to monitor the motion direction of the dots in the nearest plane of depth. For all conditions, the order of the blocks was counterbalanced across subjects. Following each stimulus block was a 3 s (2 volume) interval of fixation during which only a small cross in the center of the screen was visible.

Visual displays were back-projected with a DLP projector onto a screen located at the observer's feet. A periscope mirror attached to the birdcage headcoil was adjusted prior to the onset of the scans to maximize viewing angle of the screen.

Analysis

Images were corrected for in- and out-of-plane motion using AIR 3.08 (Woods et al., 1998). All subsequent analyses were done using Brain Voyager (Brain Innovations, Inc.) and Matlab. The realigned images were corrected for linear trend over time then spatially fil-

tered with a 5 mm FWHM gaussian filter. The filtered and unfiltered images were averaged together to created "multifiltered" images, as described by Skudlarski et al. (1999). Multifiltering minimizes sites of single voxel false positive activations while maximizing signal change that may be lost by spatial smoothing alone.

All observers viewed alternating blocks of point-light biological and scrambled motion, allowing us to localize STSp as described in our earlier work (Grossman et al., 2000). Because of spatial and temporal correlations naturally occurring in the data, the actual r-cutoff value for determining regions of interest (ROIs) was empirically derived through repeated simulations of false positive rates occurring in the voxels within the brain. This was done in the following manner: (1) the BOLD signal values within each voxel were randomly shuffled in time, (2) the correlation between the randomized time series and the localizer boxcar was computed for each voxel, (3) the value corresponding to the upper .01% cutoff of the distribution of r values across the brain (corresponds to a 1% Type I error rate) was selected. These steps were repeated 1000 times, and the final r value threshold for determining the activation maps was taken as the mean value of the distribution of cutoff values.

In localizing the ROIs, we used previously published and widely accepted "subtraction" conditions. However, to compare the activity levels during a variety of visual tasks, the raw MR signal from each scan was converted into percent change of the mean BOLD signal activation level during the fixation intervals in the scan. Further, we found that because of the short time between volume acquisitions (1.5), and the relatively short block duration (10.5 s), the MR signal barely reached saturation before the end of the epoch, and was more appropriately fit by a sinusoidal model (i.e., Boynton et al., 1996) than by an on-off boxcar function. Thus in calculating percent change values, we calculated the peak-to-peak differences between the stimulus and fixation intervals.

Acknowledgments

This work was supported by a Vanderbilt University Discovery Grant, NSF BCS0079579 and NSF BCS0121962. We thank Nancy Kanwisher and Paul Downing for their headless bodies and object stimuli, and we thank Marvin Chun, David Lyon, and Duje Tadin for comments on an earlier version of the manuscript.

Received: March 18, 2002
Revised: August 5, 2002

References

Ahlström, V., Blake, R., and Ahlström, U. (1997). Perception of biological motion. Perception *26*, 1539–1548.

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

Bar, M., Tootell, R.B.H., Schacter, D.L., Greve, D.N., Fischl, B., Mendola, J.D., Rosen, B.R., and Dale, A.M. (2001). Cortical mechanisms specific to explicit visual object recognition. Neuron *29*, 529–535.

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.

Boynton, G.M., Engel, S.A., Glover, G.H., and Heeger, D.J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. *16*, 4207–4221.

Bradley, D.C., Chang, G.C., and Andersen, R.A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. Nature *392*, 714–717.

Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 443–446.

Breitmeyer, B.G., and Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern matching, saccadic suppression, and information processing. Psychol. Rev. *8*, 1–36.

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

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

Cutting, J.E., Proffitt, D.R., and Kozlowski, L.T. (1978). A biomechanical invariant for gait perception. J. Exp. Psychol. Hum. Percept. Perform. *4*, 357–372.

Doner, J., Lappin, J.S., and Perfetto, G. (1984). Detection of three-dimensional structure in moving optical patterns. J. Exp. Psychol. Hum. Percept. Perform. *10*, 1–11.

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

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Friston, K.J., Ungerleider, L.G., Jezzard, P., and Turner, R. (1995). Characterizing modulatory interactions between areas V1 and V2 in human cortex: A new treatment of functional MRI data. Hum. Brain Mapp. *2*, 211–224.

Gauthier, I., Skudlarski, P., Gore, J.C., and Anderson, A.W. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat. Neurosci. *2*, 568–573.

Goodale, M.A., and Humphrey, G.K. (1998). The objects of action and perception. Cognition *67*, 181–207.

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.

Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision Res. *41*, 1409–1422.

Grossman, E.D., and Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. Vis. 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., Grady, C.L., Horwitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B., and Rapoport, S.I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. Proc. Natl. Acad. Sci. USA *88*, 1621–1625.

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

Hoffman, E., and Haxby, J. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat. Neurosci. *2*, 574–580.

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

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

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

Kourtzi, Z., and Kanwisher, N. (2000). Cortical regions involved in processing object shape. J. Neurosci. *20*, 3310–3318.

Kulikowski, J.J., and Tolhurst, D.J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. J. Physiol. *232*, 149–162.

Livingstone, M.S., and Hubel, D.H. (1987). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. Science *240*, 740–749.

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 the human occipital cortex. Proc. Natl. Acad. Sci. USA *92*, 8135–8139.

Mather, G., Radford, K., and West, S. (1992). Low-level visual processing of biological motion. Proc. R. Soc. Lond. B Biol. Sci. *249*, 149–155.

Neri, P., Morrone, M.C., and Burr, D.C. (1998). Seeing biological motion. Nature *395*, 894–896.

Orban, G.A., Dupont, P., De Bruyn, B., Vogels, R., Vandenberghe, R., and Mortelmans, L. (1995). A motion area in human visual cortex. Proc. Natl. Acad. Sci. USA *92*, 993–997.

O'Toole, A.J., Roark, D.A., and Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. Trends Cogn. Sci. *6*, 261–266.

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 activations in viewing eye and mouth movements. J. Neurosci. *18*, 2188–2199.

Schiller, P.H., Logothetis, N.K., and Charles, E.R. (1990). Functions of the colour-opponent and broad-band channels of the visual system. Nature *343*, 68–70.

Sereno, M.E., Trinath, T., Augath, M., and Logothethis, N.K. (2002). Three-dimensional shape representation in monkey cortex. Neuron *33*, 635–652.

Skudlarski, P., Constable, R.T., and Gore, J.C. (1999). ROC analysis of statistical methods used in functional MRI: Individual subjects. Neuroimage *9*, 311–329.

Somers, D.C., Dale, A.M., Seiffert, A.E., and Tootell, R.B. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. Proc. Natl. Acad. Sci. USA *96*, 1663–1668.

Suneart, S., Van Hecke, P., Marchal, G., and Orban, G.A. (1999). Motion responsive regions of the human brain. Exp. Brain Res. *127*, 355–370.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annu. Rev. Neurosci. *19*, 109–139.

Tootell, R.B.H., 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.

Ungerleider, L., and Mishkin, M. (1982). Two cortical visual systems. In Analysis of Visual Behavior, D. Ingle, M. Goodale, and R. Mansfield, eds. (Cambridge, MA: MIT Press), pp. 549–586.

Vaina, L.M., Solomon, J., Chowdhury, S., Sinha, P., and Belliveau, J.W. (2001). Functional neuroanatomy of biological motion perception in humans. Proc. Natl. Acad. Sci. USA *98*, 11656–11661.

Van Oostende, S., Sunaert, S., Van Hecke, P., Marchal, G., and Orban, G.A. (1997). The kinetic occipital (KO) region in man: An fMRI study. Cereb. Cortex *7*, 690–701.

Watanabe, T., Harner, A.M., Miyauchi, S., Sasaki, Y., Nielsen, M., Palomo, D., and Mukai, I. (1998). Task-dependent influences of attention on the activation of human primary visual cortex. Proc. Natl. Acad. Sci. USA *95*, 11489–11492.

Watson, J.D.G., 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.

Wojciulik, E., Kanwisher, N., and Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. J. Neurophys. *79*, 1574–1578.

Woods, R.P., Grafton, S.T., Watson, J.D., Sicotte, N.L., and Mazziotta, J.C. (1998). Automated image registration: II. Intersubject validation of linear and nonlinear models. J. Comput. Assist. Tomogr. *22*, 155–165.