A Dynamic Object-Processing Network: Metric Shape Discrimination of Dynamic Objects by Activation of Occipitotemporal, Parietal, and Frontal Cortices

Johannes Schultz[1], Lewis Chuang[1] and Quoc C. Vuong[1,2]

[1]Department of Cognitive and Computational Psychophysics, Max Planck Institute for Biological Cybernetics, Tübingen, Germany and [2]Division of Psychology, Newcastle University, Newcastle upon Tyne, UK

Both the first and the third author contributed equally to this work.

Shape perception is important for object recognition. However, behavioral studies have shown that rigid motion also contributes directly to the recognition process, in addition to providing visual cues to shape. Using psychophysics and functional brain imaging, we investigated the neural mechanisms involved in shape and motion processing for dynamic object recognition. Observers discriminated between pairs of rotating novel objects in which the 3-dimensional shape difference between the pair was systematically varied in metric steps. In addition, the objects rotated in either the same or the different direction to determine the effect of task-irrelevant motion on behavior and neural activity. We found that observers' shape discrimination performance increased systematically with shape differences, as did the hemodynamic responses of occipitotemporal, parietal, and frontal regions. Furthermore, responses in occipital regions were only correlated with observers' perceived shape differences. We also found different effects of object motion on shape discrimination across observers, which were reflected in responses of the superior temporal sulcus. These results suggest a network of regions that are involved in the discrimination of metric shape differences for dynamic object recognition.

Keywords: fMRI, metric shape, motion, object recognition, parametric design, psychophysics

Introduction

For active organisms, shape perception is important for recognizing and interacting with objects in a dynamic environment. A number of behavioral studies have shown that humans have an exceptional ability to estimate the shape of objects from a combination of visual cues such as shading, texture gradients, stereo disparity, and motion (Bülthoff 1991). Although shape plays a dominant role in object recognition (Tarr and Bülthoff 1998), other cues, particularly the motion of an object (e.g., rigid rotation in depth), also contribute to the recognition process. For example, motion information can be used to estimate the 3-dimensional structure of an object (Ullman 1979) that can subsequently be used for object recognition or motion information can serve as a direct cue to object identity (Stone 1998; Liu and Cooper 2003; Vuong and Tarr 2006).

The roles of shape and motion cues in object recognition seem to be reflected at the neural level as well. Evidence from functional magnetic resonance imaging (fMRI) studies points to a network of occipitotemporal and parietal cortical regions that are involved in the processing and integration of shape and motion cues and that may ultimately contribute to object recognition. These regions are shown in Figure 1. First, there is a large region in the posterior part of the occipital lobe—the latero-occipital complex (LOC)—that responds more to objects (Malach et al. 1995) than to textures or scrambled images, irrespective of the cues that define the objects' shape (e.g., Grill-Spector et al. 1998, 1999, 2000; Kourtzi and Kanwisher 2000; Kourtzi et al. 2003; Hayworth and Biederman 2006). Second, there is a region at the junction between the inferior temporal sulcus and the lateral occipital sulcus (hMT+/V5) that responds more to moving than to stationary stimuli (Zeki et al. 1991; Tootell et al. 1995). Third, a large posterior portion of the superior temporal sulcus (STS) seems to integrate shape and motion cues for important classes of visual stimuli, such as facial, body, or animate motion (e.g., Grossman et al. 2000; Puce et al. 2003; Schultz et al. 2005). Fourth, there are regions along the intraparietal sulcus (IPS) that appear to play a role in recovering 3-dimensional structure from 2-dimensional motion signals projected onto the retinas (e.g., Paradis et al. 2000; Kriegeskorte et al. 2003; Murray et al. 2003; Peuskens et al. 2004). Again, the estimated 3-dimensional shape can feed into an object recognition system. Finally, the frontal lobe may play a role in object recognition through its involvement in cognitive control, working memory, and attention (Goldman-Rakic 1987; Kanwisher and Wojciulik 2000; Miller 2000; Cabeza et al. 2003). Recent human neuroimaging studies also show the involvement of prefrontal cortex, in conjunction with parietal cortex, for object categorization and mental rotation tasks (Gauthier et al. 2002; Ganis et al. 2007; Jiang et al. 2007; Schendan and Stern 2007).

The broad aim of the present work is to understand the functional organization of the cortical network involved in visual object recognition (see Fig. 1). The previous studies reviewed above highlight distinct regions that process different visual cues. Our goal was to integrate these results by investigating the contribution of individual regions to dynamic object recognition using a single paradigm. Specifically, we focused on how observers discriminate metric differences in 3-dimensional shape between pairs of dynamic objects. Observers were shown 2 rotating objects in sequence and had to decide if these were the same or different objects. The rotation of the objects in depth provided static views of those objects and retinal motion signals for shape estimation. Furthermore, the objects rotated in either the same or the different directions to test whether task-irrelevant motion direction can modulate neural activity during shape processing. One possibility is that the same rotation direction may facilitate shape discrimination, particularly when discrimination between object pairs is made more difficult by increasing shape similarity, as has been suggested from behavioral data (Vuong and Tarr 2006). In addition, the modulation of shape discrimination performance may also depend on individual observers' sensitivity to shape and motion cues. Stone et al.
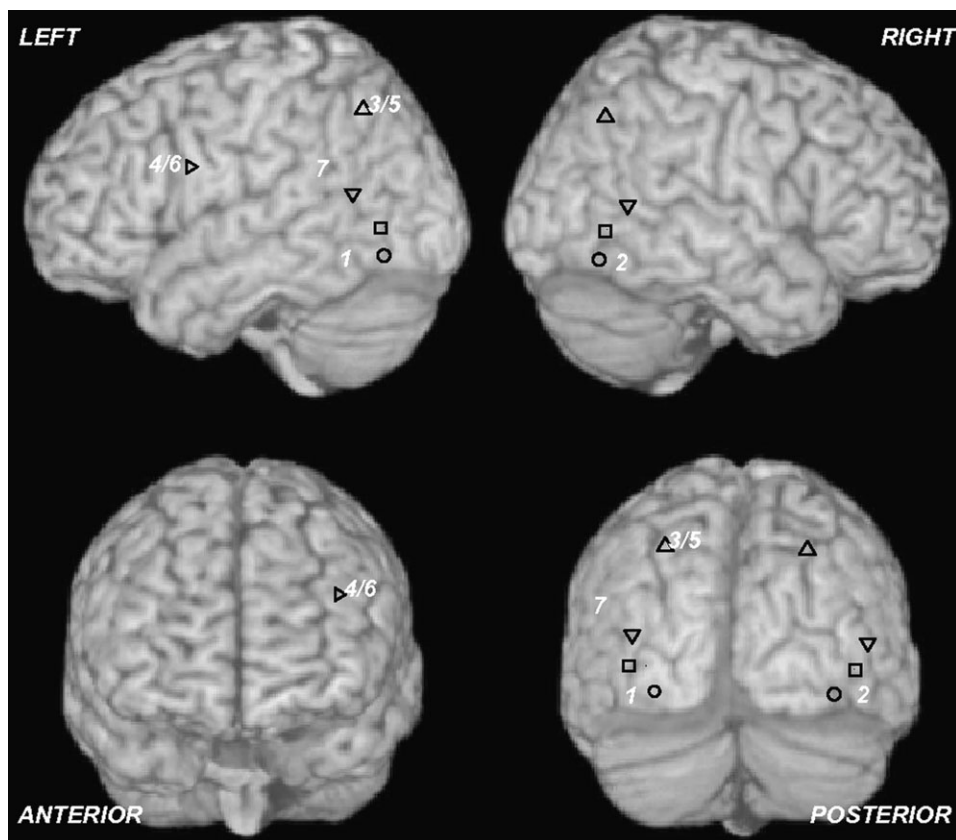
Figure 1. A network of cortical regions involved in object processing. The shape symbols are Talairach coordinates of peak activations averaged across previous reports (circle = LOC; square = hMT+/V5; triangle down = STS; triangle up = IPS; triangle right = frontal). The following references were selected: Zeki et al. 1991; Malach et al. 1995; Tootell et al. 1995; Grill-Spector et al. 2000; Grossman et al. 2000, 2004; Paradis et al. 2000; Dukelow et al. 2001; Grill-Spector and Malach 2001; Kourtzi et al. 2002; Beauchamp et al. 2003; Kourtzi et al. 2003; Kriegeskorte et al. 2003; Murray et al. 2003; Peuskens et al. 2004; Sack et al. 2006; and Schendan and Stern 2007. It should be noted that this list is far from exhaustive. The Talairach coordinates of the peak activations for the current study are shown as white numeric symbols corresponding to Table 1 (1–4 = clusters from the perceived shape difference analysis; 5–6 = clusters from the objective shape difference analysis; and 7 = cluster from the effect of motion analysis).

(2000), for example, have shown that performance on an object recognition task could be explained by observers' sensitivity to shape and motion cues.

In our experiment, we constructed multipart objects whose 3-dimensional shape was controlled by a set of parameters. These parameters were motivated by the early debate in theories of object recognition that tried to distinguish between object representations based on nonaccidental image properties (such as a curved edge vs. a straight edge; Biederman and Gerhardstein 1993, 1995; Biederman and Bar 1999) and those based on metric image properties (such as edges with different degrees of curvature; Tarr and Bülthoff 1995; Hayward and Tarr 2000). Biederman (1987), in his influential paper, proposed that a small set of qualitative shape primitives (i.e., geons) could serve as the building blocks of object representations. These primitives can be rapidly identified from binary contrasts (e.g., straight vs. curved edges) of 3 or 4 nonaccidental image properties projected by objects. Alternatively, other researchers have proposed that observers encode metric variations of image features (Tarr and Bülthoff 1998). In the present study, we used nonaccidental properties but allowed their values to vary in a continuous rather than binary manner (see also Kayaert et al. 2003, 2005).

Figure 2 illustrates examples of the objects used in the present study and how our parameterization allowed us to systematically vary the shape difference between 2 objects. The
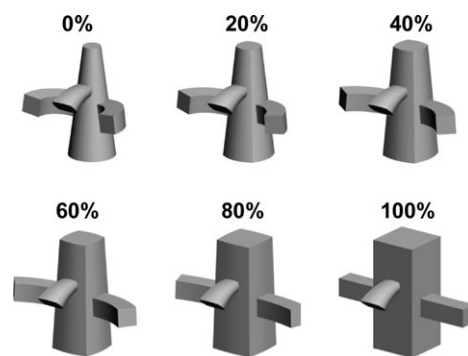

Figure 2. Examples of morphs between 2 exemplar objects (0% and 100%). For illustration purposes, these examples were rendered in 3D Studio Max and in a single color. The stimuli used in the experiment were rendered in color using custom software.

parametric manipulation of shape differences served 3 purposes. First, it allowed us to measure the extent to which brain and behavioral responses vary systematically with the magnitude of shape difference. Such a relationship between response and stimulus parameter would support a metric object representation, as suggested by behavioral and computational work (e.g., Cutzu and Edelman 1998; Lawson et al. 2003). Second, the shape parameterization allowed us to directly

compare brain activation to objective shape differences as measured by our parameterization and perceived shape differences as measured by observers' responses. Finally, the parametric design increased the statistical and interpretative power of our fMRI analysis (Friston 2005).

Here, we used an fMRI adaptation paradigm to study the network of regions that may be involved in dynamic object recognition. fMRI adaptation refers to the reduction in blood oxygen level–dependent (BOLD) response that occurs when a stimulus is repeated or when the presented stimulus shares a property with a previous stimulus. It is generally thought that this adaptation is due to reduced responses of neurons selective for that property (Grill-Spector and Malach 2001; Grill-Spector et al. 2006). Importantly, researchers have shown that the magnitude of adaptation can be varied by systematically changing stimulus parameters of interest, leading them to suggest a tight functional association between a neural region and the processing of those parameters. For example, increasing the visual dissimilarity between 2 faces results in a corresponding reduction in adaptation in face-selective regions (e.g., Rotshtein et al. 2005; Fang et al. 2007; Gilaie-Dotan and Malach 2007). We used this logic to investigate whether there are regions that show adaptation to parametric differences in 3-dimensional shape between 2 dynamic objects, thereby identifying a network of regions that process dynamic metric shape differences. We further examined neural activation time courses in significant clusters to explore possible segregation of functional roles within this network. Consistent with previous work, our results suggest that occipitotemporal, parietal, frontal, and superior temporal regions are involved in metric shape discrimination of dynamic objects and that different regions within this network process different yet complementary aspects of dynamic stimuli.

Materials and Methods

Participants
Thirteen observers (5 females and 8 males) from the Tübingen community volunteered as subjects for pay. Two of the authors also served as subjects (J.S. and Q.V.). Naive observers did not know the purpose of the experiment and had not seen the stimuli used. All participants provided informed consent and filled out a standard questionnaire approved by the local ethics committee for experiments involving a high-field magnetic resonance (MR) scanner to inform them of the necessary safety precautions.

Stimuli
Figure 2 shows examples of the novel multipart objects used as stimuli (Biederman and Gerhardstein 1993; Vuong and Tarr 2006). Each object consisted of a large central body part with 3 smaller parts attached to it. These appendages were approximately 50-70% smaller in volume than the body. Two appendages, both of the same shape, were attached laterally to the central body so that the object was symmetric about the vertical axis of the body. The remaining appendage attached to the body defined the front of the object (i.e., 0° view). The body, lateral appendages, and front appendage had different shapes. Each of these part was a geon (Biederman 1987) specified by continuous values along 3 parameters: the 2-dimensional shape of its cross section (from circle to square), the magnitude of bending perpendicular to its axis of elongation (from –45° to 45°), and the tapering of its cross section size along the axis (from –0.6 to 0.6 arbitrary unit). The effect of varying these parameters on a geon is shown in Figure 3.

We created 6 arbitrary sets of pairs by fixing the parameter values of the 4 component parts. The distance between the 2 exemplars of a pair was normalized to 100%, and new points were sampled along this
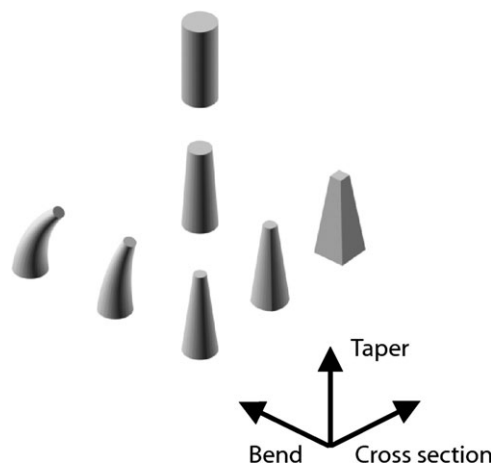

Figure 3. Each part of a multipart object is controlled by 3 shape parameters: cross-section shape, bending, and tapering.

"identity vector" in equal 5% intervals. Each new point defined a set of parameter values to create a multipart object that was effectively a "morph" between the 2 end points. Thus, there were 21 objects in each set including the 2 end points. The objective similarity between any 2 objects in a set was defined as the percentage of shape difference along this identity vector. Figure 2 shows intermediate objects for one set of exemplar pair. There were a total of 126 multipart objects.

The objects were created in 3D Studio Max v8 (Autodesk, Montreal, Canada). The 3-dimensional coordinates of the vertices and their corresponding surface normals were imported into custom software that rendered the parts of the objects with different matte colors. The body was red, the lateral appendages were yellow, and the frontal appendage was green. The same color scheme was used for all objects; therefore, color was not a cue to identity. Rather, the color was provided to facilitate segmentation of the object into its constituent parts. The objects were illuminated by several constant light sources. All objects were rendered against a uniform black background. The object models and a viewing program are available at: http://www.staff.ncl. ac.uk/q.c.vuong/smx.html

Design and Procedure
The experiment consisted of a "same—different" discrimination task in which observers judged whether 2 sequentially presented objects were the same object or different objects. It was emphasized to observers that shape differences could sometimes be very small. Therefore, they should respond as accurately as possible. The experiment conformed to a 2 × 6 within-subjects factorial design with the motion direction of the 2 objects (same direction and different direction) and the percentage of shape difference of the 2 objects (0% [same object] to 50% in 10% increments) as repeated measures. There were 4 experimental runs conducted while observers were in the scanner. Each run lasted approximately 7 min. In each run, the 6 sets were used once in each of the 12 experimental conditions for a total of 72 experimental trials. There were an additional 12 fixation trials in which there was only a fixation cross to allow hemodynamic responses to decrease toward baseline levels that increases the power of the experimental design (Josephs and Henson 1999). There were thus a total of 84 trials per run (14% of the trials were fixation conditions). All trials, including fixation trials, were randomly presented for each run and for each observer. Across the 4 runs, there were 24 repetitions of each experimental condition. The 4 runs were run sequentially with a short break (2-3 min) between runs to setup the experiment and give observers a short rest.

Objects were presented rotating in depth about the vertical axis at an angular velocity of 60°/s. The starting angle of both the first and the second object was randomly determined between –90° and 90°, with 0° representing the frontal view of the objects and ±90° representing the side views. Thus, observers saw all component parts

of an object on most trials. The direction of the first object (clockwise or counterclockwise) was randomly determined per trial. The direction of the second object was either in the same direction or in the opposite direction with respect to the rotation direction of the first object. The 2 objects were always from the same exemplar pair set. As with the motion factor, the first object was randomly selected from 1 of the 21 possible objects in a set on each trial. The second object was then selected so that the percentage of shape difference between it and the first object was between 0% (same object) and 50% (different objects). Note that only about 17% of the experimental trials were same trials (0% shape difference).

Each trial lasted either 4400 or 4500 ms for each observer because 2 different stimulus durations were used. The sequence of events on a given trial was as follows: there was a 500-ms fixation cross, followed by the first object, followed by a 500-ms blank period, followed by the second object, and finally followed by a second blank period. For 4 of the 15 observers, both the first and the second objects were presented for 750 ms. For the remaining observers, both objects were presented for 700 ms. Therefore, the objects rotated a total of 45° or 42° about the vertical axis on each trial. The observers' task was to respond same or different using a scanner-compatible response box at any time after the onset of the second object (or do nothing during fixation trials). The mapping between response and button was counterbalanced across observers. If observers did not respond before 2000 ms after the onset of the second object, the experiment continued to the next trial. No feedback was provided. Prior to being put in the scanner, observers were shown some example trials from the experiment to familiarize them with the stimulus, task, and response.

Observers laid supine on the scanner bed. The stimuli were back projected onto a projection screen situated behind the observers' head and reflected into their eyes via a mirror mounted on the head coil. The projection screen was 140.5 cm from the mirror so that the stimuli subtended a maximum visual angle of approximately 9.0°. A JVC LCD projector with custom Schneider-Kreuznach long-range optics, a screen resolution of 1024 × 768 pixels, and a 60-Hz refresh rate was used. The experiment was run on a 3.2-GHz Pentium 4 Windows PC with 2 GB RAM and an NVIDIA GeForce 7800 GTX graphics card with 256 MB video RAM. The program to present the stimuli and collect responses was written in C and relied on the OpenGL 1.2 interface to the PC's graphics hardware.

Image Acquisition
All participants were scanned at the MR Centre at the Max Planck Institute for Biological Cybernetics. All anatomical T_1-weighted images and functional gradient-echo echo-planar T_2^*-weighted images (EPI) with BOLD contrast were acquired from a Siemens Trio 3-T scanner with an 8-channel phased-array head coil (Siemens, Erlangen, Germany). The imaging sequence for functional images had a repetition time of 3000 ms, an echo time of 40 ms, a flip angle of 90°, a field of view of 256 × 256 mm, and a matrix size of 64 × 64 pixels. Each functional image consisted of 36 axial slices. Each slice had a thickness of 3.0 × 3.0 × 2.5 mm with a 0.5-mm gap between slices. This volume was positioned to cover the whole brain based on the information from a 13-slice parasagittal anatomical localizer scan acquired at the start of each scanning session. For each observer, 137 functional images (or 140 for 4 observers who were presented with slightly longer stimulus durations) were acquired in a single session lasting approximately 7 min, including a 12- and 16-s blank period at the beginning and end of each run. The first 4 of these images were discarded as "dummy" volumes to allow for equilibration of T_1 signal. A high-resolution anatomical scan was also acquired for each observer with a T_1-weighted MDEFT sequence lasting approximately 12 min.

fMRI Data Preprocessing
Prior to any statistical analyses, the functional images were realigned to the first image and resliced to correct for head motion. The aligned images were then normalized into a standard EPI T_2^* template with a resampled voxel size of 3 × 3 × 3 mm = 27 mm³ (Friston et al. 1995). Following normalization, the images were convolved with an 8-mm full width at half maximum Gaussian kernel to spatially smooth the data.

This smoothing enhanced the signal-to-noise ratio and allowed comparisons across observers.

fMRI Statistical Analyses
Processed fMRI data were analyzed using the general linear model (GLM) framework implemented in the SPM2 software package from the Wellcome Department of Imaging Neuroscience (www.fil.ion. ucl.ac.uk/spm). A 2-step mixed-effects analysis was used. The first step used a fixed-effects model to analyze individual data sets. The second step used a random-effects model to analyze the group aggregate of individual results. No additional smoothing was used in the second step.

For each observer, a temporal high-pass filter with a cutoff of 128 s was applied to the preprocessed data to remove low-frequency signal drifts and artifacts, and an autoregressive model (AR 1 + white noise) was applied to estimate serial correlations in the data and adjust degrees of freedom accordingly. Following that, a linear combination of regressors in a design matrix was fitted to the data to produce beta estimates (Friston et al. 1995) that represent the contribution of a particular regressor to the data.

For this study, we modeled the full trial duration from the onset of the first stimulus to simplify the analyses of neural adaptation across our experimental conditions (see Appendix for rationale and more details). There were 12 experimental conditions (2 motion × 6 shape difference) and 1 fixation condition. Two sets of regressors were created for each of these conditions in the following manner. For each condition, we first modeled the onsets of the first stimulus of each trial (or the onset time for the fixation trials) as a series of delta functions. The first set of regressors was created by convolving this series of delta functions with a canonical hemodynamic response function (HRF). The HRF was implemented in SPM2 as a sum of 2 gamma functions. The second set was created by convolving the delta functions with the first temporal derivative of the HRF. Therefore, there were a total of 26 regressors per experimental run in the part of the design matrix used to model experimentally induced effects. In addition, the design matrix also included a constant term and 6 realignment parameters (yaw, pitch, roll, and 3 translation terms). These parameters were obtained during motion correction and used to correct for movement-related artifacts not eliminated during realignment.

For our statistical analysis, contrasts of beta estimates were then used to create contrast images to assess the main effects of motion (same motion and different motion), shape difference (0–50%, in 10% increments), and the interaction between these 2 factors. For all contrasts involving the shape difference between the 2 objects, 2 sets of contrast weights were used. One set of contrasts consisted of linearly increasing weights over shape difference. For the other set, these linear weights were scaled by each observer's proportion of different responses at each level of shape difference. Consequently, the first set of weights represents the "objective" shape difference, whereas the second set represents observers' "perceived" shape difference. Note that the contrast weights for the perceived shape difference were necessarily different for each observer, whereas the contrast weights for the objective shape difference were necessarily the same for every observer. For each motion condition, all weights were mean subtracted so that they summed to zero, as required for a GLM. We only report the results from the group analysis. In this second level of analysis, 1-sample t-tests were performed on observers' contrast images for specific contrasts. For all statistical tests, we used $P < 0.05$ corrected for multiple comparisons across the whole brain at the cluster level and a cluster size threshold of 20 voxels (Poline et al. 1997).

We then compared brain activation between the objective and perceived shape difference as follows. First, we thresholded the statistical maps from both analyses at $P < 0.001$, uncorrected. Then for each cluster identified in the perceived shape difference contrast, we selected the closest cluster in the objective shape difference contrast. Finally, we calculated the proportion of overlap and nonoverlap for each pair of clusters. A similar overlap analysis was recently used by Schendan and Stern (2007), for example, to compare patterns of activation for saccades, mental rotation, and object decision tasks.

Lastly, we adapted a correlation method used by Haynes et al. (2005; see also Macaluso et al. 2000) as a simple means to test for possible synchrony between brain regions while observers performed the task.

This synchrony may further help segregate functional roles in a network of regions by finding regions that have similar time courses. Briefly, after fitting the BOLD signal data for each observer with the GLM using SPM2, we calculated the residuals, that is, the nonmodeled signal, by subtracting the fitted data from the real data from all voxels in our regions of interest. We then averaged these residuals across voxels within each region and computed pairwise correlations on the averaged data. The residuals were used to rule out the possibility that correlations were driven by our stimulus manipulation or by observers' responses (i.e., the residuals represent the variance not explained by these factors).

Results

Behavioral Effects of Shape Difference and Motion

Figure 4 presents the behavioral results. The data consisted of the proportion of trials observers responded different in each of the experimental conditions. An omnibus analysis of variance (ANOVA) on these proportions with motion (same motion and different motion) and shape difference (0% [same] – 50%) as repeated measures showed only a main effect of shape difference, $F_{5,70} = 216.5$, $P < 0.01$. As evident in Figure 4, the proportion of different responses increased with shape differences between the 2 objects, irrespective of whether the 2 objects rotated in the same direction or in different directions.

We also estimated each observer's 75% shape discrimination threshold separately for the same motion and different motion conditions by fitting a cumulative Gaussian distribution to individual data using the psignifit toolbox (Wichmann and Hill 2001). This threshold represents the amount of objective shape difference needed by that observer to discriminate between the 2 dynamic objects with 75% accuracy. Consistent with the ANOVA, there was no difference in shape discrimination thresholds for the 2 motion conditions, $t_{14} = 1.3$, $P = 0.22$. The mean percentage of shape difference averaged across observers at threshold was 40.3% (standard error [SE] = 2.4%) for the same motion condition and 42.5% (SE = 2.0%) for the different motion condition.

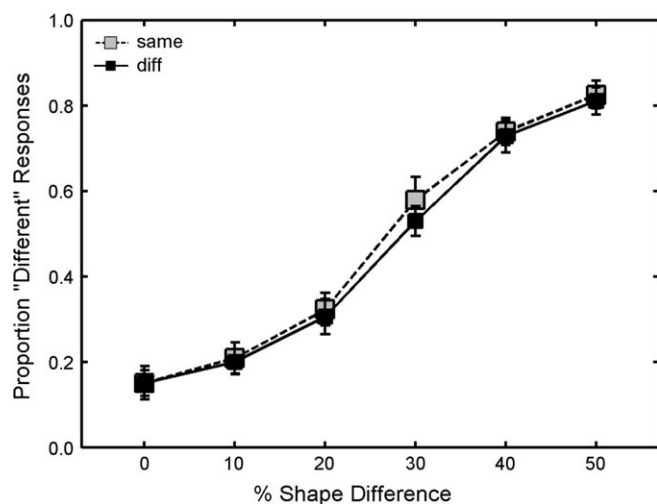However, when we calculated the difference between the same motion threshold and the different motion threshold for each observer, we found an effect of motion on shape discrimination thresholds that varied across observers. This distribution is shown in Figure 5. For some observers, their discrimination threshold decreased if the 2 stimuli had the same motion pattern. For others, a reverse pattern was observed. Therefore, this distribution suggests that each individual's shape discrimination could be modulated by irrelevant motion information, which would reflect a form of shape-by-motion interaction. That is, subtle shape discrimination may depend on individual observers' sensitivity to shape and motion cues (Stone et al. 2000). For example, individuals may vary in their ability to derive shape estimates from the rotation of the object.

Although these individual differences can be the result of chance, we also found similar individual differences in brain activity, which correlated with these behavioral differences (see below). This correlation argues against a purely chance account of the observed individual differences in behavior.

fMRI Data

In parallel with the behavioral results, we found main effects of both objective and perceived shape difference on BOLD responses. These main effects could not be accounted for by task difficulty because we did not find regions that significantly correlated with observers' accuracy on shape discrimination performance (as a measure of task difficulty). Furthermore, in a region of interest analysis described below, we also found a motion effect on shape discrimination. These findings are discussed in the following paragraphs, and further details about the clusters are provided in Table 1. None of the other contrasts led to any significant clusters. The supplementary material presents the statistical parametric maps for the individual clusters identified in these separate analyses.

Neural Correlates of Shape Differences

Figure 6 shows orthogonal projections of voxels specific to objective shape differences (light gray, outlined in black), voxels specific to perceived shape differences (dark gray), and nonspecific voxels that responded to both objective and perceived shape differences (black). The peak activation of each cluster is also plotted in Figure 1 to show their spatial
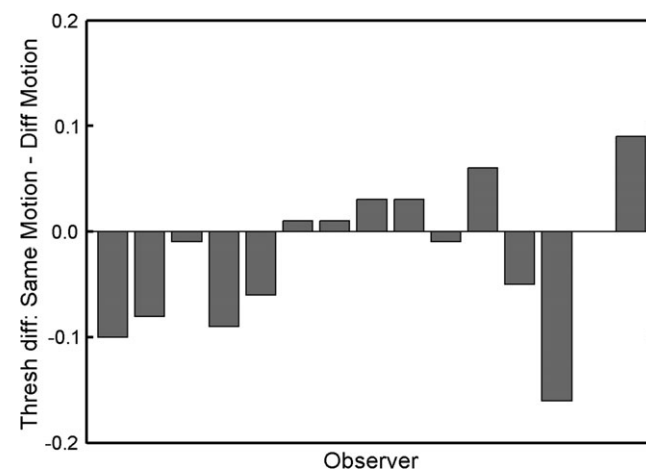

Figure 4. The proportion of different responses as a function of the percentage of shape difference averaged across observers. Error bars represent the standard errors of means across observers.

Figure 5. The distribution of 75% same motion threshold — 75% different motion threshold for each observer. Thresholds were estimated using a cumulative Gaussian distribution.

Table 1
The 4 clusters identified by the perceived shape difference contrast (1–4), the 2 clusters identified by the objective shape difference contrast (5–6), and the single cluster identified by the effect of motion on shape discrimination (7)

Cluster	Coordinates			Z score	P_{corr}	Volume (mm^3)	Structure
	x	y	z				
1	−45	−56	−5	3.60	0.035	999	Left lateral occipital (inferior temporal gyrus)
2	48	−61	−4	3.78	0.028	1053	Right lateral occipital (inferior temporal gyrus)
3	−27	−61	53	4.18	0.01	1323	Left superior parietal lobule
4	−45	10	27	4.56	0.000	2430	Left middle frontal gyrus
5	−24	−64	53	4.18	0.004	1512	Left superior parietal lobule
6	−48	10	30	4.22	0.000	2106	Left middle frontal gyrus
7	−59	−48	25	3.58	0.036	108	Left posterior superior temporal gyrus

Note: Note that the P value for cluster 7 was corrected by a SVC for STS.
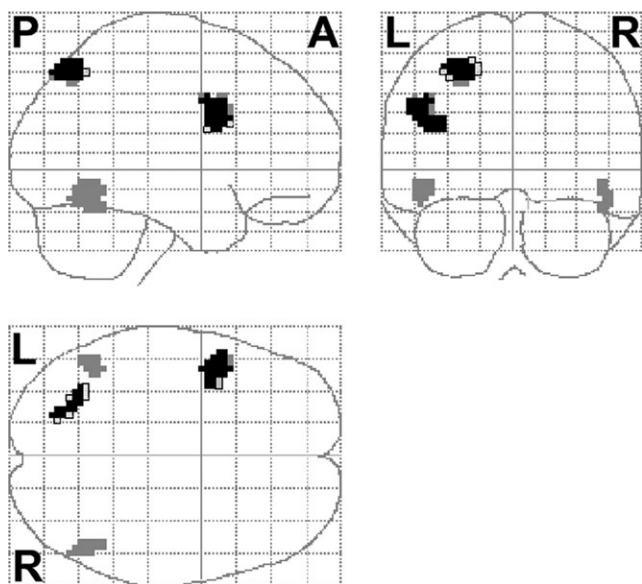

Figure 6. Maximum intensity projection images of the fMRI data (L = left; R = right; A = anterior; P = posterior). Significant activation for the objective shape difference analysis (light gray regions, outlined in black) and the perceived shape difference analysis (dark gray regions). The overlap between these 2 analyses is shown in black. See also Table 2. Thresholds for both analyses were Height: $t_{14} = 3.79$, Extent: $k = 20$.

relationship to previous fMRI studies of shape and motion perception.

The beta estimates for the different levels of shape difference for all significant clusters are shown in Figure 7. For simplicity, the beta estimates are extracted from clusters identified by the perceived shape difference analysis. This analysis was chosen because it identified all 4 clusters. Furthermore, although the objective shape difference contrast identified slightly different parietal and frontal clusters, there is substantial overlap between the 2 contrasts for these clusters so that the beta estimates were essentially the same. As evident in Figure 7, BOLD responses generally increased as a function of the shape difference between the pair of objects, suggesting that all these clusters are involved in processing metric shape differences.

There also appears to be a decrease in the beta estimates in all 4 clusters from the 40% shape difference to the 50% shape difference in Figure 7. Such a nonlinearity would be an evidence for some degree of qualitative processing of shape. To test whether this drop is significant, we submitted the beta estimates for each cluster to separate repeated-measures ANOVA with shape differences as a within-subjects factors (averaging across same motion and different motion conditions, as there was no main effect of motion). Importantly, reverse Helmert contrasts showed no significant drop in beta estimates from the 40% to 50% shape difference levels for any of the clusters. This contrast compares the beta estimate at 1 stimulus level with the beta estimate averaged across all preceding stimulus levels to compare successive levels (and based on the results of the repeated-measures ANOVA). We stress that the results from these contrasts are consistent with the fMRI analyses, which showed a significant correlation between observers' perceived shape difference and their BOLD response. Table 2 summarizes the analysis using the reverse Helmert contrasts.

In both the objective shape difference analysis and the perceived shape difference analysis, we found clusters of voxels in the cortex surrounding the IPS in the left hemisphere (objective: –24, –64, 53; perceived: –27, –61, 53) and clusters in the left frontal lobe (objective: –48, 10, 30; perceived: –45, 10, 27). Clusters from both analyses are almost at the same location (see discussion of overlap below). These findings are consistent with previous studies that showed that parietal regions are involved in structure-from-motion processing (e.g., Paradis et al. 2000; Kriegeskorte et al. 2003; Murray et al. 2003; Peuskens et al. 2004) and that frontoparietal regions may be involved in mental rotation and object categorization (Gauthier et al. 2002; Ganis et al. 2007; Jiang et al. 2007; Schendan and Stern 2007).

We also found bilateral activation in occipitotemporal cortex only in the perceived shape difference analysis, consistent with earlier findings that occipitotemporal regions are directly involved in processing shape (e.g., Grill-Spector et al. 1998, 1999, 2000; Kourtzi and Kanwisher 2000; Kourtzi et al. 2003). The Talairach coordinates of the peak activations in these lateral occipital clusters (right: 48, –61, –5; left: –45, –56, –5) fell within the spread of LOC coordinates reported in previous studies (e.g., Malach et al. 1995) but were displaced more laterally and more anterior relative to previous peaks (see Fig. 1).

Overlap between Objective and Perceived Shape Differences

To further investigate the sensitivity of significant clusters to objective and perceived shape differences, we compared the spatial overlap between voxels across the whole brain that responded to objective shape differences and those that responded to perceived shape differences (Schendan and Stern 2007). Table 3 presents the percentage of overlap between these 2 contrasts (P < 0.001, uncorrected). We highlight 2 main findings from this analysis. First, clusters in the occipitotemporal cortex are predominantly driven by perceived shape differences. There is only a 10.3% overlap of voxels between the objective and the perceived set of contrast weights for the right occipitotemporal cluster and a 10.8% overlap for the left cluster. Second, by comparison, voxels in parietal and frontal cortices are driven by either objective or perceived shape
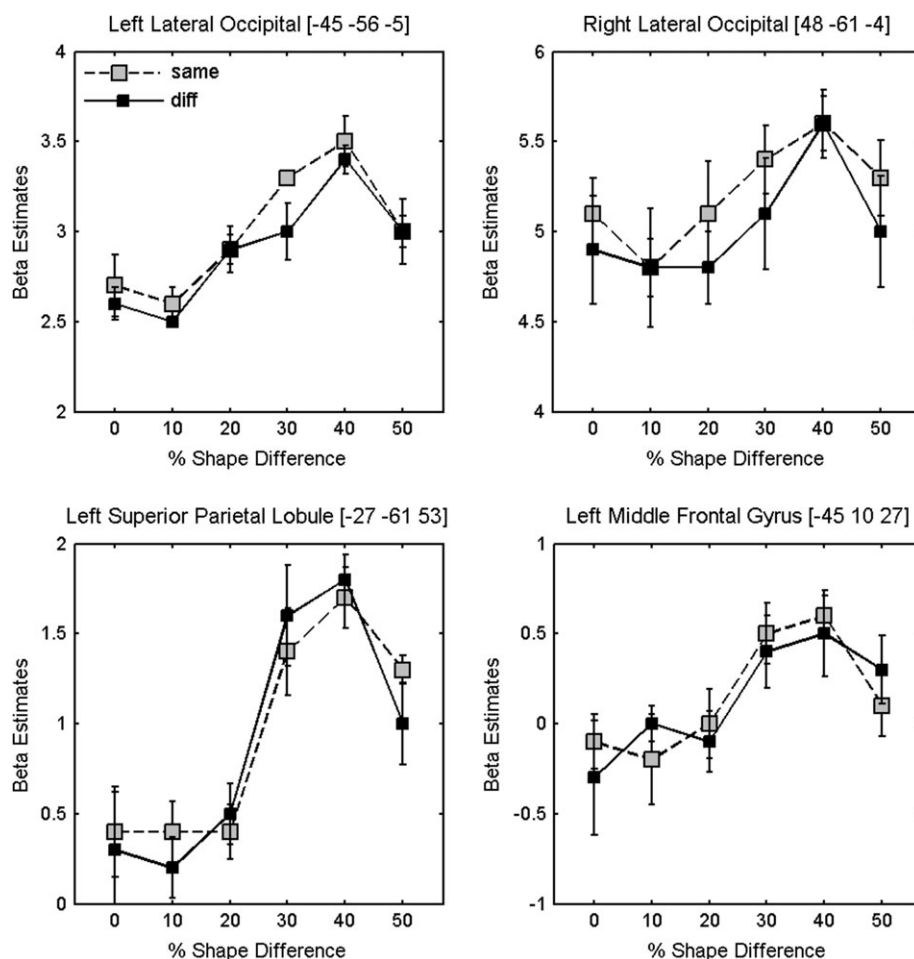
Figure 7. The beta estimates from the perceived shape difference analysis as a function of the percentage of shape difference. The estimates were extracted from the clusters as indicated in Materials and Methods. These estimates were first averaged across voxels in each cluster and then averaged across observers. Error bars represent the standard errors of means across observers. Similar functions were obtained for parietal and frontal regions if the estimates were extracted from these regions based on the objective shape difference contrast.

Table 2
F values with 1 and 14 degrees of freedom for the reverse Helmert contrasts of beta estimates for the 4 clusters identified by the perceived shape difference contrast

		10%	20%	30%	40%	50%
1	Left lateral occipital (inferior temporal gyrus)	0.52	17.80**	7.27*	24.19**	0.45
2	Right lateral occipital (inferior temporal gyrus)	1.24	0.13	3.20	12.82**	0.08
3	Left superior parietal lobule	0.03	0.21	12.81**	26.47**	1.41
4	Left middle frontal gyrus	0.27	0.48	17.99**	20.68**	0.43

Note: Each *F* value compares the current shape difference level with the mean of all preceding levels.
*$P < 0.05$.
**$P < 0.01$.

differences, as indicated by the large amount of overlapping voxels (66.7% for the parietal cluster and 71.4% for the frontal cluster). It is important to emphasize that although observers' performance are highly correlated with objective shape difference (i.e., the perceived and objective contrast weights are correlated), activation in occipitotemporal regions are almost exclusively driven by perceived shape difference.

Time Course of Residual Activation

To assess whether these clusters show similar activation patterns over and above any similarity of response induced by our experimental conditions, we tested the pairwise correlations between residual time courses (Macaluso et al. 2000; Haynes et al. 2005). These residuals are fluctuations in the BOLD signal not explained by our GLM. We used this residual analysis only to make "relative comparisons" of possible shared pattern of activation between cluster pairs, which may reflect their stimulus- and response-independent neural synchrony for our task. Table 4 shows pairwise correlations across the 4 clusters from the perceived shape difference analysis. Consistent with the overlap analysis above, we found the largest pairwise correlations between the left and the right occipitotemporal clusters, $r = 0.67$, and between the parietal and the frontal clusters, $r = 0.68$. The other pairwise correlations for these 4 clusters ranged from $r = 0.46$ to $r = 0.55$ (see Table 3). That is, regions that responded with the same degree of specificity to objective or perceived shape differences also had similar residual time courses. The higher correlations across hemispheres and across parietal and frontal lobes indicate that these correlations are not necessarily due to artifacts such as smoothing or spatial proximity.

Table 3
The percentage of overlap and specificity for the objective and perceived shape difference contrasts

Structure	% Overlap	% Specificity		Number of voxels	
		Objective	Perceived	Objective	Perceived
Left lateral occipital (inferior temporal gyrus)	10.8	2.7	86.5	5	36
Right lateral occipital (inferior temporal gyrus)	10.3	0.0	89.7	4	39
Left superior parietal lobule	66.7	22.2	11.1	56	49
Left middle frontal gyrus	71.4	8.2	20.4	78	90

Note: The union of these 2 contrasts resulted in a total number of significant voxels in each of the 4 structures indicated. From this total number of voxels, the percentage of overlap was computed as the intersection of the 2 contrasts (i.e., contrast A and contrast B), and the specificity for each contrast was computed as the percentage of voxels unique to a particular contrast (e.g., contrast A and not contrast B). Percentages in each row sum to 100%.

Table 4
Pairwise correlations of residual time courses between clusters identified by the perceived shape difference contrast

	1	2	3	4
1		0.67 (0.09)	0.55 (0.15)	0.53 (0.13)
2			0.48 (0.12)	0.46 (0.08)
3				0.68 (0.05)

Note: These correlations were computed for each observer and then averaged across observers. Parentheses are standard deviations across observers. The shaded correlations reflect the occipitotemporal pair and the frontal–parietal pair. 1 = left lateral occipital; 2 = right lateral occipital; 3 = left superior parietal lobule; 4 = left middle frontal gyrus.

Effect of Motion on Shape Discrimination

As discussed in the behavioral results and shown in Figure 5, motion direction had different effects on individual observers' 75% discrimination threshold, which may reflect an interaction between shape and motion cues. To test for neural regions that responded to this interaction, we used the regression model in SPM2 to find voxels in which there is a correlation between individual observers' difference in threshold (same motion threshold – different motion threshold) and a corresponding difference in their BOLD signal (same motion beta estimate – different motion beta estimate). An initial whole-brain analysis revealed a small cluster in left STS that did not survive our stringent corrections for multiple comparisons across the whole brain ($P < 0.001$, uncorrected; left: –59, –48, 25; Z score = 3.58; 135 mm^3). There was one other small cortical cluster in frontal region (left: –39, 24, 4; Z score = 3.53; 54 mm^3) that reached the same uncorrected P value. Statistical parametric maps for these 2 clusters ($P < 0.001$, uncorrected) are shown in the supplementary material.

We were, however, motivated to focus our analysis in cortical surfaces along posterior STS because several studies have shown that regions here integrate shape and motion cues (e.g., Grossman et al. 2000). We defined anatomical regions to perform a small volume correction (SVC) for multiple comparisons of voxels within these regions along STS (Poline et al. 1997). Like analyses across the whole brain, SVC uses random field theory to correct for multiple comparisons within the smaller defined regions. These regions were defined on the basis of an anatomical atlas of the human brain (Duvernoy

1999) and drawn using MRIcro software (Rorden and Brett 2000; www.mricro.com). The region in left STS extended from –67 to –51 mm in the x dimension, from –64 to –32 mm in the y dimension, and from 5 to 29 mm in the z dimension. The region in right STS extended from 50 to 69 mm in the x dimension, from –63 to –35 mm in the y dimension, and from 1 to 28 mm in the z dimension. The volumes were 6049 mm^3 for the left STS and 7284 mm^3 for the right STS. As shown in Figure 8, within these search regions, we found again the small cluster in the left posterior STS that was also identified in the whole-brain analysis that showed significant correlation between individual threshold differences and individual BOLD signal differences. This cluster survived correction for multiple comparisons across all voxels of the search regions (left: –59, –48, 25; $P < 0.05$, SVC; Z score = 3.58; 108 mm^3).

To highlight this correlation between brain and behavior, Figure 9 shows a scatterplot of the beta estimate difference and the threshold difference for the same motion and different motion condition per observer. The beta estimates were extracted from the voxel in the cluster that showed the peak activation. There is a significant negative correlation, $r_{13} = -0.80$, $P < 0.001$, which suggests that activations in left posterior STS are modulated by individual observers' sensitivity to shape and motion cues. This modulation was suggested by previous behavioral results (Stone et al. 2000). Again, we stress that the correlation between BOLD signal differences and behavioral threshold differences across observers provides strong support that these individual differences in either brain or behavior are not due to chance.

Task Difficulty
Our main hypothesis is that BOLD responses are driven by either objective or perceived shape differences between object pairs. Another possibility—which is not necessarily mutually exclusive with this hypothesis—is that BOLD responses may be driven by the difficulty of the shape discrimination. To test for this task difficulty, we looked for brain regions that correlated with observers' accuracy on the shape discrimination task.

For this analysis, we scaled linear contrast weights by each observer's proportion of correct responses (i.e., responding same at 0% shape difference and responding different at all other levels of shape difference) rather than their proportion of different responses. Again, for each motion condition, all weights were mean subtracted so that they summed to zero. This analysis revealed no significant clusters using the same stringent threshold used for the other analyses ($P < 0.05$, corrected for multiple corrections across the whole brain). There were, however, small bilateral clusters in the anterior portions of temporal cortex and parahippocampal regions that survived a less stringent threshold, $P < 0.001$, uncorrected (right: 53, –18, –17, Z score = 3.58, 216 mm^3; left: –45, –18, –12, Z score = 3.68, 270 mm^3; left: –24, –33, –21, Z score = 3.65, 405 mm^3). These clusters, unlike the cluster reported earlier in the STS, did not lie within areas known or thought to be of particular importance for object recognition or motion perception. The areas in which these clusters lie were thus not of a priori interest in our study, and we did not have predefined search regions of interest as we did for the STS. As before, statistical parametric maps for these clusters ($P < 0.001$, uncorrected) are shown in the supplementary material.
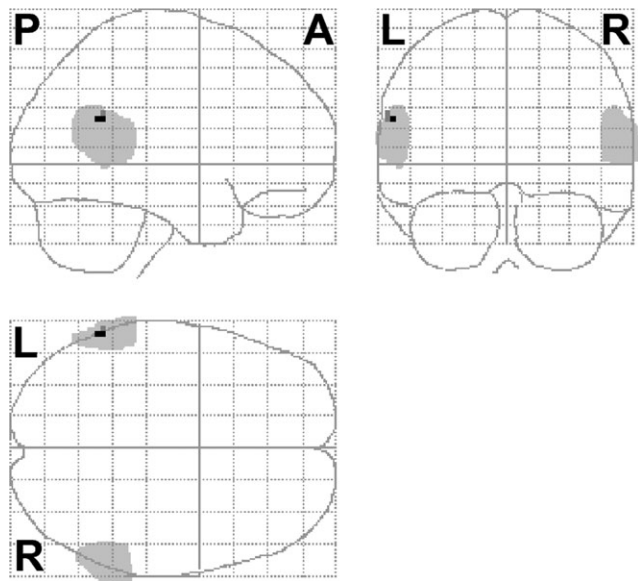
Figure 8. Maximum intensity projection images of the search region in the left and right posterior STS (light gray regions) and the significant cluster from the effect of motion on shape discrimination (black region) (L = left; R = right; A = anterior; P = posterior). The search region was used for SVC ($P < 0.05$, SVC). Threshold was Height: $t_{13} = 3.85$, Extent: $k = 0$.
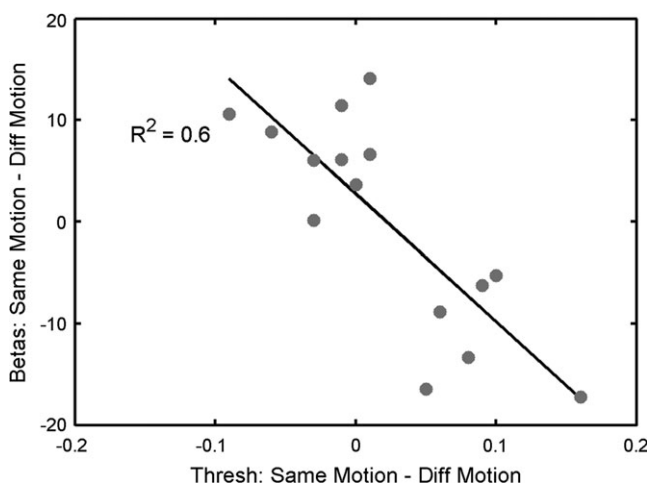

Figure 9. A scatterplot showing the correlation between the difference of beta estimates and the difference of 75% thresholds for the same motion and different motion conditions in the STS cluster. Each point represents an observer.

Discussion

In the present study, observers were required to integrate shape and motion information across low-level visual cues (e.g., particular views) and early visual processes (e.g., structure from motion or segmentation) to successfully discriminate objects that had metric differences in their 3-dimensional shape. We found that performance on this task correlated with neural activity in regions along ventral and dorsal streams, which have previously been shown to play important roles in shape perception and object recognition. Furthermore, this performance is due to the processing of 3-dimensional shape differences rather than the difficulty of the task per se.

Our main findings are as follows. First and most critically, we found that lateral occipital regions early in the visual hierarchy

process perceived shape differences irrespective of motion direction. Grill-Spector, Kourtzi and their colleagues have shown that LOC responds to familiar and novel objects that have large variations in 2-dimensional and 3-dimensional shapes (e.g., Malach et al. 1995; Grill-Spector et al. 1998, 1999, 2000; Kourtzi and Kanwisher 2000; Kourtzi et al. 2003). Recently, Hayworth and Biederman (2006) showed that LOC processes parts defined by nonaccidental image properties rather than local image features. These parts also had large shape variations across objects. At the same time, all these researchers have shown a degree of invariance in LOC with respect to object parts, image size, position, and viewpoint and with respect to the visual cues that define shape (e.g., luminance, texture, or motion). Our results imply that lateral occipital regions are not invariant to subtle shape changes as perceived by the observers, but these regions are invariant to motion. Thus, in contrast to these previous studies, our parametric manipulation of shape revealed subtle metric shape processing in LOC. This perceptual sensitivity to shape is important as subtle changes to the shape of an object could imply a change in object identity.

The fine-grain analysis of shape reported here has been demonstrated further downstream in the fusiform gyrus. For example, Jiang et al. (2006) recently found that responses in this region were correlated with the objective similarity between pairs of morphed faces. This region also seems to be recruited in recognizing visually similar exemplars of the same category such as faces, birds, dogs, and cars (e.g., Gauthier et al. 1997). Interestingly, Rotshtein et al. (2005) also found that the fusiform face area responded to perceptual differences between morphed famous faces (e.g., Margaret Thatcher and Marilyn Monroe), whereas earlier occipital face areas responded to physical differences. Recent fMRI findings also suggest that fine-grain shape analysis by regions along occipitotemporal cortex requires some degree of training (e.g., Gauthier and Tarr 2002; Op de Beeck et al. 2006; Jiang et al. 2007). Importantly, in contrast to previous studies, we find metric shape discrimination early in the visual hierarchy without explicit training (each observer received a total of 336 discrimination trials with no feedback). Along a related line, our findings extend earlier fMRI work, which looked at perceptual similarity and categorization. Edelman et al. (1998), for example, found a correlation between the clustering of categories (e.g., car and fish) based on brain activity in LOC and clustering based on human similarity ratings.

Second, we found that parietal and frontal regions are also engaged in processing metric shape differences. This finding provides evidence that parietal regions play a role in object recognition beyond recovering 3-dimensional shape information from retinal motion signals as found in previous studies (Paradis et al. 2000; Kriegeskorte et al. 2003; Murray et al. 2003; Peuskens et al. 2004). If parietal regions only or predominantly recovered shape from motion, then we would not have expected this region to respond systematically to the shape difference between objects as the information to recover shape from motion was constant (i.e., every object rotated by the same amount). Similarly, our results suggest that frontal regions may also be involved in metric shape discrimination for recognition purposes. Consistent with this claim, other fMRI studies have further shown that parietal and frontal regions are involved in both mental rotation and recognition of static images (Gauthier et al. 2002; Ganis et al. 2007; Jiang et al. 2007; Schendan and Stern 2007). It is important to note, however,

that both parietal and frontal regions do not show the same specificity to perceived shape differences as occipitotemporal regions (see Fig. 6), suggesting that clusters in ventral and nonventral streams may have different, but potentially complementary, functional roles in the recognition process. Note again that task difficulty cannot explain the BOLD responses to shape differences in these regions as responses in these regions did not correlate with observers' accuracy.

In line with this functional segregation of regions, our analysis of residuals (i.e., BOLD signals that were not explained by our experimental design) revealed an interesting temporal pattern of activation across regions in occipitotemporal, parietal, and frontal cortices. We found that clusters in occipitotemporal cortex had similar residual time courses and both had the same specificity to perceived shape difference. Likewise, clusters in frontal and parietal cortices had similar residual time courses and both responded to objective and perceived shape differences. Thus, regions with correlated residuals may be involved in similar processes, such as encoding perceived shape or estimating shape from motion signals for recognition purposes. This synchronous neural activity between regions may therefore reflect important interactions between regions (Macaluso et al. 2000; Haynes et al. 2005).

Lastly, we found a small modulatory effect of task-irrelevant motion direction on BOLD signals only in left STS when we restricted our analysis to anatomically defined bilateral STS regions. This small modulation is somewhat surprising because previous findings found strong STS responses in the perception and recognition of facial and body motion that requires the integration of shape and motion cues (e.g., Grossman et al. 2000; Puce et al. 2003; for similar integration by neurons in superior temporal polysensory area, the monkey homologue of STS, see Oram and Perrett 1994). This small modulatory effect in our study underscores the fact that shape information often plays the dominant role in object recognition (Tarr and Bülthoff 1998; Vuong and Tarr 2006). However, the significant modulation suggests that STS can integrate shape and motion cues for unfamiliar objects, even if motion cues are not relevant for the task. By comparison, previous studies have found strong STS activation for highly familiar dynamic stimuli (such as faces and bodies) in which the motion was relevant for the task. Importantly, the extent to which this integration occurs may depend on individual observers' sensitivity to these separate cues. Future work using a variety of paradigms and stimuli is needed to characterize the role of STS and possibly other areas involved in the integration of shape and motion cues. For example, work in preparation by Sarkheil, Vuong, Bülthoff and Noppeney (unpublished data) found adaptation effects in hMT+/V5 that depended on shape and motion using an fMRI adaptation paradigm.

Connections with Single-Cell Recordings in Monkey Inferior Temporal Cortex

Our fMRI results in lateral occipital regions have interesting parallels to single-cell recording studies in macaque monkeys. In particular, Kayaert et al. (2003, 2005) found neurons in inferior temporal cortex that responded in a graded fashion to quantitative variations in shape, although these neurons were generally more sensitive to qualitative shape changes (what the researchers referred to as nonaccidental properties; Vogels et al. 2001). The parameters of 2-dimensional and 3-dimensional shapes of Kayaert et al. (2003, 2005), such as curvature, are similar to the shape parameters used here. Importantly, the inferior temporal region in monkeys is a likely homologue of LOC in humans. Thus, our results also help bridge findings in human fMRI and monkey single-cell recordings.

Implications for Theories of Object Recognition

Our finding that several regions, particularly early shape processing regions such as LOC, are involved in the discrimination of metric shape differences between dynamic objects has 2 important implications for theories of object recognition. Consistent with the majority of behavioral data, our findings suggest that the human visual system encodes metric representations as opposed to qualitative shape primitives in the visual processing hierarchy (Tarr and Bülthoff 1998). At the same time, our results show that motion has behavioral and neural consequences on individual observers' performance even though this information was not relevant for the task. This finding implies that theories of object recognition need to explain, at least to some extent, how dynamic information is represented. Neural models that integrate both shape and motion cues have been developed for biological motion perception but can naturally be extended to dynamic objects (Giese and Poggio 2003).

Conclusion

The present results point to a network of occipitotemporal, parietal, and frontal regions that work in tandem for dynamic object recognition (see Fig. 1). In this network, there are further functional segregations of regions into complementary processes hypothesized for object recognition. First, ventral regions encode perceived shape as opposed to objective shape. Second, parietal and frontal regions contribute further to processing objective shape differences (e.g., through estimating structure from motion). Lastly, there are small contributions from STS that reflect individual observers' sensitivity to shape and motion cues. This study therefore provides a promising empirical link between different early visual processes and higher level object recognition. Overall, these results integrate a diverse set of studies that have identified individual regions that process specific cues into a single dynamic object-processing network.

Funding

Supplementary Material

Supplementary material can be found at http://www.cercor.oxfordjournals.org/.

Notes

Address correspondence to Quoc C. Vuong, Division of Psychology, Henry Wellcome Building for Neuroecology, Framlington Place, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK. Email: q.c.vuong@ncl.ac.uk.

Appendix

The SPM2 regressors we used to analyze the data were created by convolving delta functions time locked to the start of each trial with the canonical HRF. These single delta
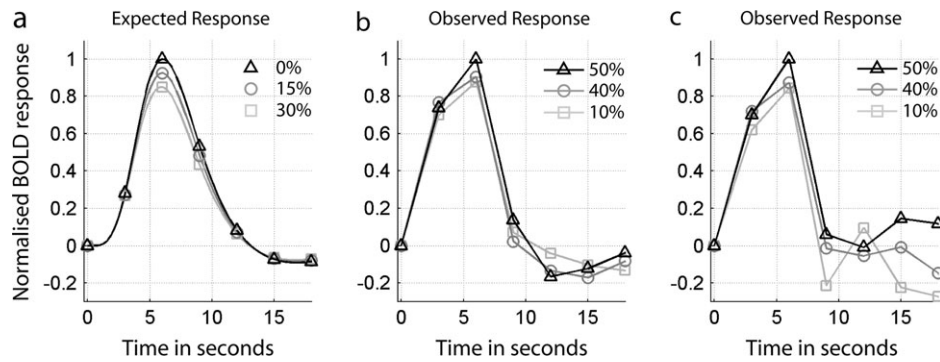
Figure A1. (*a*) The expected event-related time course predicted from a model of adaptation for 3 arbitrary levels of suppressed responses to the second trial (black triangle = 0% suppression; dark gray circle = 15% suppression; and light gray square = 30% suppression). (*b*, *c*) Also plotted is an observer's time course for clusters in the left and right occipitotemporal cortex identified by the perceived shape difference analysis (black triangle = 50% shape difference; dark gray circle = 40% shape difference; and light gray square = 10% shape difference).

functions are a simplification of the neural signal we would expect during a trial composed of 2 stimuli and with varying degrees of neural adaptation to the second stimulus. To show the validity of our analysis, we compared mean time courses extracted from the significant clusters in the lateral occipital cortex with calculations of the expected BOLD response to varying degrees of adaptation.

These calculations were made as follows. We created a time series in which boxcar functions represent periods of neural activity for each trial. The durations of these boxcars were the duration of the trial events (i.e., 700 or 750 ms for each stimulus, with a 500-ms interstimulus interval). The height of these boxcar functions represented the intensity of neural activity. To model the neural events on a given trial, we set the height of the boxcar functions for the first stimulus to a height of 1 arbitrary unit and the height of the boxcar function for the second stimulus to either 1 unit (representing no adaptation) or some fraction of a unit (representing adaptation). We then convolved these boxcars with the canonical HRF to calculate the time course of the expected BOLD signal. The event-related time courses time locked to the onset of the first stimulus are plotted in Figure A1. Two observations are evident in the leftmost plot of the expected response. First, the general form of the response, with or without adaptation, is of the same shape as the canonical HRF. Second, the effect of adaptation is mostly evident as a change in height (smaller with increasing adaptation) with negligible shifts of the peaks in time (earlier with increasing adaptation). These results are compatible with previous work on BOLD signal adaptation (Grill-Spector et al. 2006). Of course, these expected effects are to be taken with caution, as neither the BOLD signal is perfectly modeled by this canonical HRF nor are the responses to 2 stimuli perfectly additive. However, the effect seen in these simulations allowed us to expect findings in an SPM-based GLM analysis.

The middle and rightmost plots of Figure A1 show the observed event-related time course averaged across voxels in occipitotemporal clusters for a single subject. As evident in these plots, we found decreasing response heights for increasingly similar stimulus pairs. As these clusters were identified using the regressors described above, these regressors were effective in identifying regions with BOLD signal corresponding to the adaptation profile predicted from our experimental design. Deviations from the expected effects seen in the middle and rightmost panels could be due to the smaller precision of trial-based averages compared with the weighted least squares fitted GLM analyses performed using SPM2.

References

Beauchamp MS, Lee KE, Haxby JV, Martin A. 2003. fMRI responses to video and point-light displays of moving humans and manipulable objects. J Cogn Neurosci. 15:991–1001.

Biederman I. 1987. Recognition-by-components: a theory of human image understanding. Psychol Rev. 94:115–147.

Biederman I, Bar M. 1999. One-shot viewpoint invariance in matching novel objects. Vision Res. 39:2885–2899.

Biederman I, Gerhardstein PC. 1993. Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. J Exp Psychol Hum Percept Perform. 19:1162–1182.

Biederman I, Gerhardstein PC. 1995. Viewpoint-dependent mechanisms in visual object recognition: reply to Tarr and Bülthoff (1995). J Exp Psychol Hum Percept Perform. 21:1506–1514.

Bülthoff HH. 1991. Shape from X: psychophysics and computation. In: Landy M, Movshon A, editors. Computational models of visual processing. Cambridge: MIT Press. p. 305–330.

Cabeza R, Dolcus F, Prince SE, Rice HJ, Weissman DH, Nyberg L. 2003. Attention-related activity during episodic memory retrieval: a cross-function fMRI study. Neuropsychologia. 41:390–399.

Cutzu F, Edelman S. 1998. Representation of object similarity in human vision: psychophysics and a computational model. Vision Res. 38:2229–2257.

Dukelow SP, DeSouza JF, Culham JC, van den Berg AV, Menon RS, Vilis T. 2001. Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements. J Neurophysiol. 86:1991–2000.

Duvernoy HM. 1999. The human brain. Vienna (Austria): Springer Verlag.

Edelman S, Grill-Spector K, Kushnir T, Malach R. 1998. Towards direct visualization of the internal shape representation space by fMRI. Psychobiology. 26:309–321.

Fang F, Murray SO, He S. 2007. Duration-dependent fMRI adaptation and distributed viewer-centered face representation in human visual cortex. Cereb Cortex. 17:1402–1411.

Friston KJ. 2005. Models of brain function in neuroimaging. Annu Rev Psychol. 56:57–87.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS. 1995. Statistical parametric mapping in functional imaging: a general linear approach. Hum Brain Mapp. 2:189–210.

Ganis G, Schendan HE, Kosslyn SM. 2007. Neuroimaging evidence for object model verification theory: role of prefrontal control in visual object categorization. Neuroimage. 34:384–398.

Gauthier I, Anderson AW, Tarr MJ, Skudlarski P, Gore JC. 1997. Levels of categorization in visual recognition studied using functional magnetic resonance imaging. Curr Biol. 7:645–651.

Gauthier I, Hayward WG, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. 2002. BOLD activity during mental rotation and viewpoint-dependent object recognition. Neuron. 34:161–171.

Gauthier I, Tarr MJ. 2002. Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. J Exp Psychol Hum Percept Perform. 28:431–446.

Giese MA, Poggio T. 2003. Neural mechanisms for the recognition of biological movements. Nat Rev Neurosci. 4:179–192.

Gilaie-Dotan S, Malach R. 2007. Sub-exemplar shape tuning in human face-related areas. Cereb Cortex. 17:325–338.

Goldman-Rakic PS. 1987. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Plum F, editor. Handbook of physiology: the nervous system. Bethesda (MD): Am Physiol Soc. p. 373–417.

Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci. 10:14–23.

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.

Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, Malach R. 1998. Cue-invariant activation in object-related areas of the human occipital lobe. Neuron. 21:191–202.

Grill-Spector K, Kushnir T, Hendler T, Malach R. 2000. The dynamics of object-selective activation correlate with recognition performance in humans. Nat Neurosci. 3:837–843.

Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst). 107:293–321.

Grossman E, Donelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R. 2000. Brain areas involved in the perception of biological motion. J Cogn Neurosci. 12:711–720.

Grossman ED, Blake R, Kim CY. 2004. Learning to see biological motion: brain activity parallels behavior. J Cogn Neurosci. 16:1669–1679.

Haynes JD, Driver J, Rees G. 2005. Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. Neuron. 46:811–821.

Hayward WG, Tarr MJ. 2000. Differing views on views: comments on Biederman and Barr (1999). Vision Res. 40:3895–3899.

Hayworth J, Biederman I. 2006. Neural evidence for intermediate representations in object recognition. Vision Res. 46:4024–4031.

Jiang X, Bradley E, Rini RA, Zeffiro T, VanMeter J, Riesenhuber M. 2007. Categorization training results in shape- and category-selective neural plasticity. Neuron. 53:891–903.

Jiang X, Rosen E, Zeffiro T, Vanmeter J, Blanz V, Riesenhuber M. 2006. Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. Neuron. 50:159–172.

Josephs O, Henson RN. 1999. Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philos Trans R Soc Lond B Biol Sci. 354:1215–1228.

Kanwisher N, Wojciulik E. 2000. Visual attention: insights from brain imaging. Nat Rev Neurosci. 1:91–100.

Kayaert G, Biederman I, Op de Beeck HP, Vogels R. 2005. Tuning for shape dimensions in macaque inferior temporal cortex. Eur J Neurosci. 22:212–224.

Kayaert G, Biederman I, Vogels R. 2003. Shape tuning in macaque inferior temporal cortex. J Neurosci. 23:3016–3027.

Kourtzi Z, Bülthoff HH, Erb M, Grodd W. 2002. Object-selective responses in the human motion area MT/MST. Nat Neurosci. 5:17–18.

Kourtzi Z, Erb M, Grodd W, Bülthoff HH. 2003. Representation of the perceived 3-D object shape in the human lateral occipital complex. Cereb Cortex. 13:911–920.

Kourtzi Z, Kanwisher N. 2000. Cortical regions involved in perceiving object shape. J Neurosci. 20:3310–3318.

Kriegeskorte N, Sorger B, Naumer M, Schwarzbach J, van den Boogert E, Hussy W, Goebel R. 2003. Human cortical object recognition from a visual motion flowfield. J Neurosci. 23:1451–1463.

Lawson R, Bülthoff HH, Dumbell S. 2003. Interactions between view changes and shape changes in picture-picture matching. Perception. 32:1465–1498.

Liu T, Cooper LA. 2003. Explicit and implicit memory for rotating objects. J Exp Psychol Learn Mem Cogn. 29:554–562.

Macaluso E, Frith CD, Driver J. 2000. Modulation of human visual cortex by crossmodal spatial attention. Science. 289:1206–1208.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA. 92:8135–8139.

Miller EK. 2000. The prefrontal cortex and cognitive control. Nat Rev Neurosci. 1:59–65.

Murray SO, Olshausen BA, Woods DL. 2003. Processing shape, motion and three-dimensional shape-from-motion in the human cortex. Cereb Cortex. 13:508–516.

Op de Beeck HP, Baker CI, DiCarlo JJ, Kanwisher NG. 2006. Discrimination training alters object representations in human extrastriate cortex. J Neurosci. 13:13025–13036.

Oram MW, Perrett DI. 1994. Response of anterior superior temporal polysensory STPa neurons to "biological motion" stimuli. J Cogn Neurosci. 6:99–116.

Paradis AL, Cornilleau-Peres V, Droulez J, Van de Moortele PF, Berthoz A, Le Bihan D, Poline JB. 2000. Visual perception of motion and 3-D structure from motion: an fMRI study. Cereb Cortex. 10:772–783.

Peuskens H, Claeys KG, Todd JT, Norman JF, Van Hecke P, Orban GA. 2004. Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. J Cogn Neurosci. 16:665–682.

Poline JB, Worsley KJ, Evans AC, Friston KJ. 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. Neuroimage. 5:83–96.

Puce A, Syngeniotis A, Thompson JC, Abbott DF, Wheaton KJ, Castiello U. 2003. The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. Neuroimage. 19:861–869.

Rorden C, Brett M. 2000. Stereotaxic display of brain lesions. Behav Neurol. 12:191–200.

Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ. 2005. Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat Neurosci. 8:107–113.

Sack AT, Kohler A, Linden DE, Goebel R, Muckli L. 2006. The temporal characteristics of motion processing in hMT/V5+: combining fMRI and neuronavigated TMS. Neuroimage. 29:1326–1335.

Schendan HE, Stern CE. 2007. Mental rotation and object categorization share a common network of prefrontal and dorsal and ventral regions of posterior cortex. Neuroimage. 35:1264–1277.

Schultz J, Friston KJ, O'Doherty J, Wolpert DM, Frith CD. 2005. Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. Neuron. 45:625–635.

Stone JV. 1998. Object recognition using spatiotemporal signatures. Vision Res. 38:947–951.

Stone JV, Buckley D, Moger FA. 2000. Determinants of object recognition. Vision Res. 40:2723–2736.

Tarr MJ, Bülthoff HH. 1995. Is human recognition better described by geon-structural descriptions or by multiple-views? Comments on Biederman and Gerhardstein 1993. J Exp Psychol Hum Percept Perform. 21:1494–1505.

Tarr MJ, Bülthoff HH. 1998. Object recognition in man, monkey, and machine. Cognition. 67:1–20.

Tootell RB, Reppas JB, Dale AM, Look RB, Sereno MI, Malach R, Brady TJ, Rosen BR. 1995. Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. Nature. 375:139–141.

Ullman S. 1979. The interpretation of visual motion. Cambridge: MIT Press.

Vogels R, Biederman I, Bar M, Lorincz A. 2001. Inferior temporal neurons show greater sensitivity to nonaccidental properties than to metric shape differences. J Cogn Neurosci. 13:444–453.

Vuong QC, Tarr MJ. 2006. Structural similarity and spatiotemporal noise effects on learning dynamic novel objects. Perception. 35:497–510.

Wichmann FA, Hill NJ. 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys. 63:1293–1313.

Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS. 1991. A direct demonstration of functional specialization in human visual cortex. J Neurosci. 11:641–649.