Representation of Shapes, Edges, and Surfaces Across Multiple Cues in the Human Visual Cortex.

Joakim Vinberg¹ and Kalanit Grill-Spector¹,²

¹Psychology Department, Stanford University, Stanford, CA 94305

²Neurosciences Institute, Stanford University, Stanford, CA 94305

Corresponding author:
Kalanit Grill-Spector
Department of Psychology
Stanford University
Stanford, CA 94305
Phone: 650 725 2457
Fax: 650 725 5699
e-mail: kalanit@psych.stanford.edu

Running head: Cue-invariant processing of shapes, edges and surfaces

Manuscript includes: Abstract, Introduction, Materials and Methods, Results, Discussion, Acknowledgements, Grants, References, Figure legends, 2 Tables, 7 Figures.

Supplemental Materials: Supplemental Movie 1: Stereo stimuli; Supplemental Movie 2: Motion stimuli; Supporting Figure 1: LOC responses during gray-level localizer experiment.
ABSTRACT

The lateral occipital complex (LOC) responds preferentially to objects compared to random stimuli or textures independent of the visual cue. However, it is unknown whether the LOC (or other cortical regions) are involved in the processing of edges or global surfaces without shape information. Here, we examined processing of (1) global shape, (2) disconnected edges without a global shape, and (3) global surfaces without edges vs. random stimuli across motion and stereo cues. The LOC responded more strongly to global shapes than edges, surfaces or random stimuli, for both motion and stereo cues, but its responses to local edges or global surfaces were not different than random stimuli. This suggests that the LOC processes shapes, not edges or surfaces. LOC also responded more strongly to objects than holes with the same shape, suggesting sensitivity to border ownership. V7 responded more strongly to edges than surfaces or random stimuli for both motion and stereo cues, while V3a and V4 preferred motion-edges. Finally, a region in the caudal intra-parietal sulcus (cIPS) responded more strongly to stereo versus motion, and to stereo-surfaces versus random stereo, (but not to motion-surfaces versus random-motion). Thus, we found evidence for cue-specific responses to surfaces in the cIPS, both cue-specific and cue-independent responses to edges in intermediate visual areas and shape-selective responses across multiple cues in the LOC. Overall, these data suggest that integration of visual information across multiple cues is mainly achieved at the level of shape and underscore LOC’s role in shape computations.

Keywords: Shape perception, lateral occipital complex (LOC), border ownership, figure-ground segmentation, cue invariance, surface segmentation
INTRODUCTION

The human visual cortex is composed of numerous visual areas whose retinotopy and functional characteristics have been extensively studied (for a review see (Grill-Spector and Malach 2004)). However, the organizational principles that underlie the specialization of visual areas remain mysterious. In particular, it is unclear to what extent processing of primary visual cues such as motion (Albright et al. 1984; Castelo-Branco et al. 2002; Heeger et al. 1999; Huk and Heeger 2002; Newsome et al. 1989; Sack et al. 2006; Tootell et al. 1995; Zeki 2004), stereo (Sakata et al. 1999; Taira et al. 2000; Tsao et al. 2003), or luminance are segregated into distinct cortical regions, or to what extent visual areas integrate different visual cues toward performing a specific visual task, such as object recognition (Goodale and Milner 1992; Mishkin et al. 1983). Previous research has provided evidence for both segregation and convergence of visual cue processing in the human and primate visual cortex (DeAngelis et al. 1998; Flytche et al. 1995; Grill-Spector and Malach 2004; Palanca and DeAngelis 2003; Seidemann et al. 1999; Shikata et al. 2001; Tsutsui et al. 2002; Tsutsui et al. 2005).

Several studies provide evidence for cue-invariant processing of objects in the human and primate ventral stream. Neurons in macaque inferotemporal cortex respond more strongly to shapes than simple features, and their shape selectivity is maintained across different visual cues (Sary et al. 1993; Tanaka et al. 2001; Vogels and Orban 1996). The human lateral occipital complex (LOC (Grill-Spector et al. 2001; Malach et al. 1995)), a region thought to be involved in object recognition (Grill-Spector et al. 2001; Grill-Spector et al. 2000), responds more strongly to objects defined from either luminance, texture, illusory contours, motion, stereo, or color than random stimuli or uniform textures generated from the same primary cues (Appelbaum et al. 2006; Gilaie-Dotan et al. 2002; Grill-Spector et al. 1998; Kastner et al. 2000; Kourtzi and Kanwisher 2000; Mendola et al. 1999). These studies provide evidence for convergence of low-level visual cues in generating object selective responses in the LOC.
Visual processing en-route to object recognition must be able to extract shape\(^1\) independently of the visual cues that define it. The recognition process likely entails a sequence of computations across visual cortex, starting from local computations in early visual cortex related to low level properties of the visual stimulus, such as disparity, motion, or orientation and convey little sense of the global object shape, then proceeding to more global computations in higher levels of the hierarchy of visual processing. However, it is unknown whether the LOC (or other visual regions) are involved in the processing of intermediate visual information such as edge or surface information that may be extracted during processing en-route to object recognition (Marr 1982; Nakayama 1995; Nakayama et al. 1989). In particular, Nakayama and colleagues (Nakayama 1995; Nakayama et al. 1989) suggested that surfaces are represented in intermediate visual areas of the visual processing stream. This surface representation may provide a means to segment the visual input into relevant regions, which in turn may enhance the efficiency of recognition processes because they would operate only on subset of meaningful surfaces, rather than the entire visual input (Nakayama et al. 1989).

Previous studies have examined the processing of edge information or surface information en route to recognition only from luminance cues. Several studies have shown that the LOC activates more strongly for global shapes defined by luminance edges than misaligned, incomplete or scrambled luminance edges (Altmann et al. 2003; Doniger et al. 2000; Lerner et al. 2004; Lerner et al. 2002), suggesting that LOC responds to global shapes defined by luminance edges rather than local luminance edges. However, it is unknown: (1) whether this pattern of results is specific for luminance defined stimuli or extends to other cues and (2) whether the LOC also processes local edge information as compared to random stimuli with no edges whatsoever.

\(^1\) Our working definition of an edge is a discontinuity in the image. Edges can be real or illusory. Our working definition of a shape is a distinct region enclosed by an edge. Shapes, by definition, are bounded by an edge. However, edges do not always define a shape, e.g., disconnected edges (Fig. 1c) do not define a global shape.
Others have suggested that the LOC processes global surface information because its response to enclosed regions defined by Kanizsa inducers was higher than its response to rotated inducers, even when the illusory contours were not perceived (Stanley and Rubin 2003). However, it is unknown: (1) whether this pattern of response is specific to luminance-defined stimuli and (2) whether the LOC responds to global surfaces without shape or edge information.

To address these questions, we investigated the involvement of visual regions in processing global shapes, global surfaces or edges by examining differential cortical activation across these stimuli. Critically, we examined whether preferential responses for shapes, surfaces or edges were cue-specific, or occurred across both motion and stereo cues. We conducted two fMRI experiments in which subjects viewed dot stimuli generated from one visual cue (motion or stereo) that created either: (1) a percept of a global shape (Fig. 1a,b), or (2) disconnected edges that did not form a global shape or a global surface (Fig. 1c), or (3) two global surfaces without edges (Fig 1d), or (4) were completely random (Fig. 1e). Random stimuli had the same relative disparity (or motion) range as other conditions, but contained no coherent structure, edges, or global surfaces.

We reasoned that regions involved in processing global surfaces would respond more strongly to surfaces than random stimuli that are otherwise identical expect that they contain no coherent structure (Fig. 1d vs. 1e). The surfaces condition created a percept of two semitransparent global surfaces in two depth planes (generated via disparity or motion parallax) without any edges. We posited that regions involved in processing edges would respond more strongly to stimuli containing edges (Fig. 1c) than either surfaces or random stimuli. The edges condition contained disconnected motion or stereo-edges generated from scrambling object contours without forming a global shape. Finally, we reasoned that regions involved in global shape processing would activate more strongly to shape (Fig. 1a, 1b) than edges, surfaces or
random stimuli. To better characterize the involvement of LOC in shape processing we created two types of shapes: objects and holes, which had the same shape and contour but differed in their border ownership (Fig. 1a vs. 1b).

All stimuli were made of random dots of the same density, contained no luminance edges (except for the screen boundary which was present during the entire experiment, including the blank baseline period) and were controlled for low-level properties (Materials and Methods). We chose to generate stimuli from motion and stereo cues because: (i) they allowed controlling low-level properties across conditions, (ii) perception of edges is possible from these cues alone, (iii) both motion parallax and disparity provide sufficient information for perceiving surfaces and their relative depth, and (iv) processing of motion and stereo cues is thought to involve distinct cortical regions, which allowed us to separate cue processing from content processing.

**MATERIALS AND METHODS**

**Subjects**

Thirteen subjects (7 female, 6 male; age range 21-36) participated in the study. All subjects had normal or corrected-to-normal vision and intact stereoscopic vision. Subjects provided informed consent to participate in the experiment, which was approved by the Stanford Internal Review Board on Human Subjects Research. Subjects viewed all stimuli through red/cyan glasses.

**fMRI Acquisition**

Subjects were scanned on a 3.0T GE scanner located at the Lucas Imaging Center, Stanford University. We used a receive-only surface coil that covered the occipital lobe, the posterior parietal lobe, and most of the temporal lobe. We acquired T1-weighted inplane anatomical images at the same prescription as the functional data using an SPGR pulse sequence (TR = 2000 ms, TE = 1.9 ms, flip angle = 15°, FOV = 20 cm, bandwidth = 15.63). We collected functional data from 32 slices oriented perpendicular to the calcarine, with a voxel size
of 3.125 x 3.125 x 3 mm, and a TR=2000ms using a spiral sequence (Glover 1999). We acquired whole brain high-resolution anatomical images using a head coil in a separate session for 11 out of 13 subjects.

Subjects' behavioral data were collected via a scanner-compatible response box. One subject's behavioral data during scan was incompletely recorded and excluded from analysis.

**General Experimental Design**

Stimuli were presented in a block design in which 16s blocks of visual stimuli alternated with 8s blank periods which consisted of a gray background with a centrally presented fixation. Each trial in a block lasted 2s. The blank baseline was identical across motion and stereo experiments. The only luminance edge was the screen border, which was present in all conditions including the blank condition, see Supplemental Movies online. In each of the stereo and motion experiments, subjects viewed six blocks of each experimental condition, totaling 48 unique trials per condition. Shapes were identical across objects and holes and across stereo and motion experiments, and included silhouettes of animals, cars, common objects, geometric forms and abstract shapes. Object and hole sizes were $11.5° ± 1.8°$. Since there is no significant difference in the level of LOC activation across nameable and non-nameable shapes (Supporting Figure 1), using both types of shapes enabled us to relate our findings to previous experiments which used both namable (Grill-Spector et al. 1998; Kourtzi and Kanwisher 2000) and non-nameable shapes (Kourtzi and Kanwisher 2000; Stanley and Rubin 2003).

**Stereo Experiments**

Stimuli were stationary random dot stereograms with 8% dot density (Fig. 1 and online Supplemental Movie 1). All conditions used the same range of relative disparity ($± 0.24°$ to $0.37°$). There were no luminance edges in any of the conditions. Each stereogram contained two depth-defined surfaces, except the random condition in which dots were assigned all possible disparities in a random assignment. Each trial in a block (lasting 2s) contained a
different pairing of disparities. Conditions: (1) Stereo-object (Fig. 1a): superposition of two random dot patterns, a shaped flat region at a closer disparity and a farther background plane. (2) Stereo-hole (Fig. 1b): a front random dot surface with a shaped hole through which a flat random dot surface at the farther disparity could be seen. (3) Stereo-edges (Fig. 1c): disconnected stereo-edges presented at a single depth in front of a flat background presented at a farther disparity. Edges were generated by: (i) scrambling shape outlines, (ii) expanding the edge 3 pixels on either side, (iii) filling this region with random dots of density 8% and (iv) placing these edges at a nearer plane than the background random dot plane. The width of edges was $0.36^\circ$. There were no luminance edges and stereo-edges did not form a global shape. The edge stimuli had the same total contour length as objects and holes and edges, as in previous studies (Altmann et al. 2003; Lerner et al. 2004; Lerner et al. 2002). (4) Stereo-surfaces (Fig. 1d): two semitransparent planes filling the entire display, at two different depths with no edges, created by the superposition of two random dot patterns (front and back planes) which were assigned different disparities. Semi-transparent surfaces were created by the following procedure: (i) we generated 2 different random dot images, (ii) each of the patterns was assigned a different disparity and colored red or cyan and (iii) we superimposed these two images into one image by taking the minimum color value in each pixel across the two images. That is, if a pixel was colored in at least one image, it was assigned the minimum color, otherwise it was assigned to white. (5) Random (Fig. 1e): dots were assigned a random disparity and appeared as a 3D cloud of dots with no structure.

In each object or hole block, the objects were presented centrally in half of the trials, and displaced horizontally by $6.7^\circ \pm 0.7^\circ$ in the other half of trials. Thus, the centrally presented fixation was on the front surface of the object (or hole) for $\frac{1}{2}$ of the trials in the block and on the farther (background) surface for the other $\frac{1}{2}$ of the trials. The object or hole was displaced in
order to equalize responses in the behavioral task (where subjects judged whether the fixation was on the front or back surface) in object and hole trials to the other experimental conditions.

**Task:** Subjects were asked to fixate and judge in each trial whether the stereoscopically coded fixation point overlaid the nearer or farther surface; during blank trials subjects pressed a key when the fixation cross blinked. Fixation was assigned to the near surface in half the trials and the far surface in the other half, in the random condition the fixation was in front of, or behind the dot cloud. This task requires parsing of the image across all experimental conditions and is orthogonal to the contrasts of interest. Performance was similar across conditions (see *Behavioral data during scan*).

Stereo Experiments 1 and 2 were identical except that Experiment 1 did not include the edges condition. 9 subjects participated in the first experiment, 8 subjects in the second and 4 in both experiments.

**Motion Experiments**

Dot density was 8%, and dot lifetime was 100%. Each 2s trial in a block was composed of twenty frames, each 100ms long, to create a percept of visual motion. There were no disparity and no luminance edges. Surfaces, edges, object and hole conditions contained a stationary background plane made of random dots and a front region (made of another random dot pattern) that moved coherently to the right in half the trials and to the left in the other half of the trials (online Supplemental Movie 2). Speeds (3-3.5 degree/sec) and motion directions were randomly assigned across trials. Conditions: (1) Motion-object: a shaped occluding random dot surface moved across a stationary background of random dots. (2) Motion-hole: an occluding random dot surface with a shaped hole moved over the stationary background surface made of random dots. (3) Motion-edges: object outlines were scrambled to create disconnected edges. These edges were filled with random dots (same as stereo experiment) and moved in the same direction and speed across a stationary random dot background. There were no luminance edges and motion-edges did not form a global shape. (4) Motion-surfaces: a front
semitransparent plane composed of random dots moved across a stationary background plane composed of random dots; motion parallax provide information of which plane was in front. (5) Random: dots moved randomly, motion speed was similar to other conditions. Here in each 2s trial we generated 20 different random images, which created the percept of random motion.

**Task:** Subjects were asked fixate and judge the direction of movement of the moving surface (left or right); during blank trials subjects pressed a key when the fixation cross blinked. This task requires image parsing in all experimental conditions and is orthogonal to the contrasts of interest. During blank trials subjects pressed a button when the fixation blinked. Performance was similar across conditions and across stereo and motion experiments (see *Behavioral data during scan*).

Motion Experiments 1 and 2 were identical except that Experiment 1 did not include the edges condition. The same 9 subjects participated in motion and stereo Experiment 1. The same 8 subjects participated in motion and stereo Experiment 2.

**Localizers and Retinotopy**

All subjects participated in an independent block-design LOC localizer scan in which they viewed gray-level photographs of cars, faces, animals, and novel sculptures, and scrambled versions of these images. Subjects viewed six blocks of each condition and covertly named the stimuli as cars, faces, animals, sculptures and scrambled while fixating. This experiment was used to define LOC regions of interest (ROIs), see Fig. 3.

All subjects participated in an MT localizer in which they viewed blocks of 16s low-contrast expanding and contracting rings that alternated with 16s low-contrast stationary rings. Subjects viewed 12 blocks of each condition. MT was defined in each subject as voxels that responded more strongly to moving stimuli compared to stationary stimuli located at the posterior end of the infero-temporal sulcus (ITS), thresholded at \( p < 10^{-5} \), uncorrected (Tootell et al. 1995, see Fig. 2).
Eleven (out of 13) subjects participated in meridian mapping and eccentricity scans (Grill-Spector and Malach 2004) that were used to delineate retinotopic areas that are defined by visual field maps: V1, V2, V3, V3a, V4 and V7, see Fig. 2.

fMRI Data Analysis

Data were preprocessed and co-registered to a high-resolution anatomical volume using BrainVoyager 2000 (Brain Innovation, Masstricht, Netherlands). All subsequent analyses were performed in BrainVoyager QX and in-house code written in Matlab (Mathworks, MA).

Functional data was temporally filtered with a linear trend removal and high pass filtering with a cutoff of 3 cycles per run. For each subject, we registered their functional data to their high-resolution anatomical volume. Subjects’ high-resolution anatomies were transformed to Talairach coordinates, segmented to gray and white matter, and their cortical surface was reconstructed.

Voxel-by-Voxel Analysis

Voxel based analyses used a General Linear Model (GLM) implemented in BrainVoyager QX. Statistical maps were generated from GLM analyses that were done on a subject-by-subject basis.

ROI Analyses

ROIs were defined for each subject on their inflated cortex (to restrict ROIs to the gray matter) and then projected to the 3D volume from which time course data was extracted. Fig. 2 shows the location of all ROIs on one representative subject’s inflated cortical surface. Talairach coordinates and ROI volumes are presented in Table 1.

| Table 1 approximately here |

We extracted the average time course of each ROI and percent signal change was calculated relative to the fixation baseline. Amplitudes were calculated by averaging the percent
signal change 4-18 seconds following the onset of a block across the six repetitions of each block.

**LOC ROIs**

LOC was defined as lateral and ventral occipital regions that activated more strongly to gray-level images of objects (animals, cars and abstract sculptures) than scrambled versions of these objects (Fig. 3a). This contrast activates a large cortical extent in lateral and ventral occipito-temporal cortex. Typically, the LOC is separated into two regions LO and pFUS/OTS using a combination of anatomical and functional criteria (Grill-Spector et al. 1999; Grill-Spector et al. 2000; Sayres and Grill-Spector 2006).

**LO:** a region in lateral occipital cortex posterior and adjacent to MT, that responded significantly to gray level objects than scrambled objects, P < 0.0001, voxel level, uncorrected). We excluded voxels that overlapped MT. We divided LO into 2 sub-ROIs (Fig. 2). The first ROI (marked (a) in Figs. 2, 3d) included the dorsal aspect of LO and overlapped a retinotopic representation. This ROI extended into cortex defined as LO2 (Larsson and Heeger 2006). However, we did not find a consistent upper visual meridian, therefore we could not define the ROI using the same retinotopic criteria as Larsson and colleagues. The more ventral part of LO that extended beyond the retinotopic region was taken as a second ROI marked (b) in Figs. 2 and 3d. In (2/13) subjects without a retinotopic map, LO was subdivided using anatomical cues.

Our definition of LO is unlikely to overlap with KO (Dupont et al. 1997; Tyler et al. 2006; Van Oostende et al. 1997) because their functional definition and spatial location is different. KO refers to a region between V3a and MT that responds more strongly to kinetic gratings than uniform motion (Dupont et al. 1997; Van Oostende et al. 1997) and to stereo gratings vs. a single plane (Tyler et al. 2006), but it does not activate to luminance defined gratings (Van Oostende et al. 1997). KO overlaps with LO1, a retinotopic region adjacent and inferior to V3a (Larsson and Heeger 2006), but LO is ventral and lateral to LO1.
PFUS/OTS: a ventral region overlapping the fusiform gyrus and occipito-temporal sulcus that responded more strongly to gray level objects than scrambled objects (P < 0.0001, voxel level, uncorrected, Fig. 2 and 3d). In previous studies we referred to this region as pFUS (Grill-Spector et al. 1999; Grill-Spector et al. 2000; Sayres and Grill-Spector 2006).

Dorsal ROIs

We defined three ROIs along the IPS on a subject-by-subject basis from responses during motion and stereo experiments. (1) The most posterior ROI (Fig. 2 and Fig. 6c) was defined as a region in the posterior IPS that responded more strongly to both motion-edges versus random-motion and to stereo-edges versus random-stereo (each contrast, P < 0.01, voxel level). This ROI was adjacent to V3a and was located at the posterior end of the IPS. Superposition of this ROI on retinotopic maps showed that this region overlapped with the retinotopic representation of V7 (Press et al. 2001). Because this ROI overlapped with the visual field map of V7 defined from retinotopic scans, we will use the term V7 when referring to this ROI to link between our results and the existing literature (e.g. Tsao et al. 2003, Tyler et al. 2006). (2) A region in caudal IPS (cIPS) was defined based on its higher responses to stereo-edges than random-stereograms (P < 0.01, voxel level, uncorrected, Fig. 2 and Fig. 6c) and was detected in all eight subjects who participated in Experiment 2. In five subjects the same region responded more strongly for motion-edges than random motion (P < 0.01, voxel level, uncorrected). We labeled this ROI cIPS, based on its definition in previous studies (Tsao et al. 2003). (3) In six subjects we found an IPS region anterior to V7 and posterior to cIPS that responded more strongly to stereo-surfaces vs. random surfaces (P < 0.01, voxel level, uncorrected) and we analyzed responses of this region separately (DF2, Fig 2 and Fig. 6d).
Measuring the Overlap between Activation to Objects from Motion and Objects from Stereo

To quantify the overlap between regions responding to objects from motion and objects from stereo, we ran a voxel-by-voxel GLM detecting voxels that responded more strongly for stereo objects than random stereo (P < .005, voxel level, Fig. 3b) and a second GLM detecting voxels that responded more strongly for motion objects than random motion (P < .005, voxel level, Fig. 3c). Activations overlapped the LOC, extending anteriorly along the lateral bank of the brain and to the posterior portion of the fusiform gyrus and occipito-temporal sulcus. For each subject we quantified the size of the region that activated for stereo objects, motion objects and the overlapping region that was activated by both contrasts. We excluded voxels identified as MT. We measured in each subject the fraction of the overlapping region relative to the region activated by stereo objects or motion objects and calculated the mean and standard deviation across subjects.

Psychophysics of the Perceived Shape of Objects and Holes

All subjects participated in a psychophysics study conducted outside the scanner approximately four months after the fMRI scans using identical stimuli. We measured accuracy and response times to 40 trials of each of the following conditions: stereo-objects, stereo-holes, motion-objects, and motion-holes.

Task: Subjects were asked to fixate and report whether the fixation was inside or outside the shape on screen. This task tests whether subjects could determine a shaped region in the hole and object conditions. We chose to use this task because it was used previously to evaluate the perceived saliency of shaped regions (Stanley and Rubin 2003) and because many of our stimuli contained abstract shapes that did not allow measurement of explicit object recognition.
RESULTS

Behavioral Responses during Motion and Stereo Experiments

Performance during fMRI scans was high and similar across conditions (Table 2). In both stereo Experiments 1 and 2 there were no significant differences in accuracy or response time (RT) across experimental conditions (t–test, Ps > 0.06). During the first motion experiment there were no significant differences in accuracy or response times across experimental conditions (t–tests, P > 0.08) except for higher accuracy and longer RTs for motion–surfaces than motion–objects (t–test, P < 0.05). These differences were not significant after Bonferroni correction. In the second motion experiment there were no significant differences in RT or accuracy across experimental conditions (t–test, Ps > 0.2), except that accuracy was higher for motion–holes than motion–surfaces, Bonferroni corrected (Table 2). Overall, performance was similar across conditions and there was no condition in which performance was consistently better or worse.

Table 2 approximately here

LOC Responds Preferentially to Objects Across Multiple Visual Cues

The lateral occipital complex (LOC, Grill-Spector et al. 2001; Malach et al. 1995) consists of a constellation of regions that respond more strongly to objects than scrambled objects or random stimuli. We used an independent localizer experiment to find in each subject regions that activate more strongly to gray–level images of objects than scrambled versions of these objects (Materials and Methods). Object–selective activations occurred in the lateral bank of the occipital and temporal cortex, extending anteriorly into the fusiform gyrus. In all subjects, regions that responded preferentially to gray–level images of objects (animals, cars and novel) versus scrambled objects (P < 0.005, voxel level, Fig. 3a) also responded more strongly to stereo–objects than random disparity stereograms (P < 0.005, voxel level, Fig. 3b) and more strongly to motion–objects than random motion (P < 0.005, voxel level, Fig. 3c). These data concur with prior reports of object-selective responses in the LOC across multiple visual cues.
On average, a larger extent of lateral occipital cortex was activated by stereo-objects than motion-objects (right hemisphere: stereo $3812 \pm 2378$ mm$^3$ (mean $\pm$ std); motion: $1957 \pm 1266$ mm$^3$. Left hemisphere, stereo $3959 \pm 2418$ mm$^3$; motion: $1688 \pm 1333$ mm$^3$). We found substantial overlap in the spatial extent of regions activated more strongly to stereo objects vs. random stereo and motion objects vs. random motion. The extent of overlap was limited by the smaller size of the region that showed higher responses for motion objects (%overlap, right hemisphere: 

\[
\frac{\text{motion objects} \& \text{stereo objects}}{\text{motion objects}} = 70 \pm 27\% , \quad \frac{\text{motion objects} \& \text{stereo objects}}{\text{stereo objects}} = 39 \pm 23\% ; \quad \text{Left hemisphere:} \\
\frac{\text{motion objects} \& \text{stereo objects}}{\text{motion objects}} = 64 \pm 31\% ; \quad \frac{\text{motion objects} \& \text{stereo objects}}{\text{stereo objects}} = 37 \pm 25\% .
\]

Figure 3 approximately here

**LOC Processes Shape, not Edges or Surfaces**

We extracted responses from LOC subregions defined by the gray level localizer experiment (Fig. 2, 3 and Materials and Methods) during motion and stereo experiments to examine whether the LOC is involved in processing of shape, local edges or global surfaces across multiple cues. Using a four-way analysis of variance (ANOVA) we examined the effects of stimulus type (Experiment 1: object/hole/surfaces/random, Experiment 2: object/hole/edges/surfaces/random), cue (stereo/motion), ROI (LO a, b, pFUS/OTS) and hemisphere (left/right) on fMRI responses. Each cell in the ANOVA contained the mean response for a condition (across 6 blocks) for each subject, and subjects were treated as a random variable.

We found a significant effect of stimulus type (Experiment 1: $F_{3,519} = 48.06$, $P < 10^{-6}$; Experiment 2: $F_{4,389} = 18.24$, $P < 10^{-6}$) indicating that LOC responses differ across object, hole,
edges, surfaces and random. There was no significant effect of cue in Experiment 2 ($F_{1,389} = .57$, $P = 0.45$), and no significant effect of cue when the ANOVA was performed across both experiments ($F_{1,509} = 1.01, P = 0.39$), but there was a small effect of cue in Experiment 1 ($F_{1,519} = 7.54, P = 0.01$). Moreover, there were no significant interactions between cue and other factors in either Experiment 1 or 2 ($Fs < 2.24; Ps > 0.07$), suggesting that LOC responses are similar across motion and stereo stimuli. We found a main effect of ROI (Experiment 1: $F_{4,519} = 40.84, P < 10^{-6}$; Experiment 2: $F_{2,389} = 20.37; P < 10^{-6}$). Responses were not different between hemispheres (Experiment 1: $F_{4,519} = 4.44, P = 0.0356$; Experiment 2: $F_{2,389} = .86; P = 0.35$, except for LO (b) in Experiment 1) and there were no significant interactions between any of the factors and hemisphere. Therefore, subsequent analyses were performed separately for each ROI, averaged across hemispheres.

We next examined whether processing of either surfaces or edges alone contribute to LOC responses. Across all LOC ROIs we found no significant difference between the responses to surfaces and responses to the random condition in either motion or stereo experiments (paired t-test, $t(12) < 0.88, p > 0.2$, Fig. 4). This suggests that global surfaces without edges (generated from motion or disparity) do not differentially activate LOC compared to random stimuli. For both motion and stereo cues disconnected edges did not activate LOC ROIs significantly more than surfaces (Fig. 4b, $ts(6) < 1.7, Ps > 0.06$) or more than random (Fig. 4b, $ts(6) < 0.58, Ps > 0.29$).

However, for both motion and stereo cues, LOC subregions responded to objects and holes significantly more than random stimuli (Fig. 4c, objects: $ts(12) > 5.14, Ps < 0.0002$, holes: $ts(12) > 2.1, Ps < 0.03$) and significantly more than global surfaces (Fig. 4c, objects: $ts(12) > 4.2, Ps < 10^{-3}$, holes: $ts(12) > 2.6, Ps < 0.02$, except for pFUS/OTS, stereo holes vs. stereo...
surfaces, $t(12) = 1.6, p = 0.07$). Activation to objects was also significantly higher than edges (Fig. 4c, $t(6) > 3.16, P < 0.01$) in both motion and stereo experiments.

These data indicate that the LOC responds more strongly to objects than global surfaces, disconnected edges or random stimuli for both motion and stereo cues. In contrast, global segmentable surfaces without edges do not activate LOC more than random stimuli. Finally, edges that did not enclose a global shape did not activate LOC regions more than random stimuli or global surfaces.

**LOC Responds More Strongly to Objects than Holes**

Notably, LOC's response to objects was greater than holes for both motion and stereo cues (Fig. 4c, paired $t$-test, $t(11) > 2.73; P < 0.01$) even though the contour that defined the shape of objects and holes was identical. Since the only difference between objects and holes is the border ownership of the contour defining the shaped region, this suggests that LOC is sensitive to border ownership. Alternatively, higher LOC responses for objects than holes may be a consequence of lower responses to holes vs. objects in early visual areas or may relate to differences in perceiving a shaped region during hole vs. object conditions. We examined each of these possibilities in turn.

**Processing of Surfaces, Edges, and Shape in Early and Intermediate Visual Areas**

We extracted responses from retinotopic visual regions (Methods and Fig. 2) to examine whether differential responses in LOC reflect downstream responses propagated from early visual cortex and to test whether retinotopic regions show cue-specific or cue-independent preferential responses to edges or global surfaces compared to random stimuli. Single unit recordings suggest that luminance-edges and motion-edges are processed in early visual areas (Hubel 1995) and that responses of V1, V2 and V4 neurons are modulated by the border ownership of edges (Qiu and von der Heydt 2005; Zhou et al. 2000).
V1-V3 activated robustly to all experimental conditions including the random condition (Fig. 5a-c). There were no significant differences in V1 responses across stimulus types (object/hole/edges/surfaces/random, F_{4,69} = 0.86, P = 0.49) or across motion and stereo stimuli (F_{1,69} = 3.56, P = 0.06, same for Experiment 1, data not shown). V2 and V3 responses were not significantly different across experimental conditions (F_{4,69} < .48; P > 0.75), however, responses to motion stimuli were higher than stereo stimuli (F_{1,69} > 6.3, P < 0.02). The lack of differential response across experimental conditions in early visual cortex validates that our experimental conditions were controlled for low-level visual properties. There were no significant differences in responses of early retinotopic regions between objects and holes, for either motion or stereo cues (t_{s}(10) < 1.7 Ps > 0.06). This suggests that higher responses in LOC to objects than holes were not propagated from early visual cortex.

Intermediate visual areas V3a and V4 showed no significant effects of cue or stimulus type (Fs < 2.79, Ps > 0.1). Notably, V3a and V4 activations to stimuli containing motion-edges (edges, hole and object conditions) were significantly higher than random motion (each comparison, paired t-test, P < 0.03, Fig. 5d,e) and motion-surfaces (Fig. 5d,e; paired t-test, P < 0.03). Responses were not different across motion-objects, motion-holes and motion-edges (Ps > 0.35). Activation to stereo-edges was not significantly higher than random-disparity stereograms or stereo-surfaces (Ps > 0.38).

MT responses were significantly higher for motion than stereo stimuli (F_{1,79} = 39.51; P <10^{-6}). While the main effect of stimulus type was not significant (F_{4,79} = 1.21; P = 0.32), MT responded significantly more strongly to motion-objects and motion-holes than random motion or motion-surfaces, and more to stereo-objects than random stereo or stereo-surfaces, (t_{s}(11) > 2.34, Ps < 0.02, Fig. 5f). Thus, MT showed a strong preference for motion stimuli, with a slight preference for objects (see also Kourtzi et al. 2002).

Overall, analyses of retinotopic visual areas found no difference in their responses for objects vs. holes, indicating that the preference to objects observed in LOC was not propagated
from lower visual areas. We found evidence for the involvement of V3a and V4 in motion-edge processing, evidence for general processing of visual motion in MT and a weak preference for motion in V2 and V3.

**Subjects perceive a shaped region in both object and hole conditions.**

We next examined whether subjects were able to perceive a shaped region during hole and object conditions (see Materials and Methods) to determine whether differential LOC responses for objects and holes related to differential perception of the shaped region. Subjects participated in a behavioral experiment in which they viewed the same objects and holes as during fMRI scans, generated from either motion or stereo cues and were instructed to indicate whether the fixation point was inside or outside the shape. This task examines whether subjects can determine a shaped region in the image (Stanley and Rubin 2003). Lower accuracy on this task for holes would indicate that subjects were worse at determining the shaped region for holes vs. objects.

Subjects’ accuracy was high for both objects and holes, indicating that subjects were able to perceive a shaped region for both object and holes (motion-objects: accuracy: 97±3%, motion-holes: 98±3%; stereo-objects: 97±2%, stereo-holes: 93±8%). There were no significant differences in accuracy (paired t-tests, all Ps > .07, uncorrected). Subjects responded more quickly to objects versus holes (motion-objects: 967±108ms, motion-holes: 1025±115ms; stereo-objects: 902±128ms, stereo-holes: 1056±240ms, Ps < 0.005). The longer RTs for holes are unlikely to explain the lower LOC responses for holes, since during scanning there were no RT differences (Table 2) and conditions with longer response times are usually associated with higher LOC signals (Sayres and Grill-Spector 2006), not lower responses. Since subjects were able to perceive the shaped region for both objects and holes, the lower LOC activation for holes is unlikely to be due to subjects’ inability to determine the shaped region for hole stimuli.
Voxel-by-Voxel Analyses of Selective Responses to Shape, Edges and Global Surfaces.

To complement ROI analyses we performed a voxel-by-voxel search for any voxels that activated more strongly to objects, holes, edges or surfaces than random (Fig. 6). These analyses were done on a subject-by-subject basis. The goal of these analyses was to detect any region in visual cortex that showed preferential responses to edges, global surfaces or global shapes vs. random stimuli containing no structure for each of the visual cues.

Voxel-based analyses revealed that higher activation to shape vs. random (object vs. random and hole vs. random, P < 0.01, voxel level) in both motion and stereo experiments occurred in higher order visual regions including LOC, but excluded early visual cortex (Fig. 6a,b). These activations overlapped with LOC ROIs in all subjects. Ventral LOC activation to holes vs. random was less robust than objects vs. random for both motion and stereo cues in all subjects: Fewer voxels in anterior ventral regions were significantly activated by the former vs. latter contrast (Fig. 6b). Higher activation for holes than random (P < 0.01, voxel level) overlapped most subjects’ LO (ROI a: stereo: LH: 12/13, RH: 10/13; motion LH: 12/13, RH 9/13; ROI b: stereo LH: 12/13, RH 12/13; motion LH: 9/13, RH: 12/13), but in fewer subjects in pFUS/OTS (stereo: LH 7/13, RH: 7/13; motion LH: 3/13; RH: 3/13).

Voxel-by-voxel activations for edges vs. random (P < 0.01, voxel-level, Fig. 6c) revealed a dorsal region adjacent to V3a on the lateral bank of the intraparietal sulcus and/or middle occipital gyrus that activated more strongly both to motion-edges vs. random motion and also to stereo-edges vs. random disparity (Fig. 6c, Talairach coordinates in Table 1b). This pattern of response occurred bilaterally in 6/8 subjects and provides evidence for cue-invariant processing of edges. Examination of retinotopic maps indicates that these preferential responses to edges for both motion and stereo stimuli overlapped the retinotopic representation of V7. We will therefore refer to this ROI as V7. Another activation focus that responded more strongly to
edges vs. random was detected in the cIPS: cIPS responded more strongly for stereo--edges than random stereo in all subjects (P < 0.01, voxel--level, right: 8/8 subjects; left: 6/8 subjects) and in half the subjects also to motion-edges vs. random motion (P < 0.01, voxel level, right: 5/8; left 4/8 subjects).

We analyzed the mean response across V7 (Fig. 7a, see Materials and Methods). Analysis of variance of V7 responses with factors of stimulus type (object/hole/edges/surfaces/random), cue (motion/stereo) and hemisphere (right/left) revealed a significant effect of stimulus type (F$_{4,119}$ = 7.1, P < 0.00005), but no effect of cue or interactions (F$s < 1.27$ P > 0.26). Responses in the right hemisphere were stronger than the left (F$_{1,119}$ = 9.23, P < 0.003). Since there was no significant interaction between hemisphere and condition, responses were averaged across hemispheres. V7 responded significantly more strongly to edges vs. random, holes vs. random and objects vs. random for both motion and stereo stimuli (paired t-test, Ps < 0.01, Fig. 7a). In contrast, responses to either motion-surfaces or stereo-surfaces were not higher than random (Ps > 0.49). Responses to random disconnected edges were not different than responses to holes or objects that contained a shape and responses to objects were not different than holes (Ps > 0.22). This suggests that V7 participates in processing of edges from both motion and stereo cues, whether or not they define a global shape.

We also found evidence for cue-specific processing of edges: Higher activation to motion-edges vs. random motion occurred in V3a (5/7 subjects), V4 (6/7 subjects) and LO (a, 5/8 subjects), consistent with ROI analyses (Fig. 5).

**Cue-Specific Processing of Surfaces in IPS regions**

Voxel-by-voxel analyses of global surfaces vs. random stimuli (P < 0.01, voxel level) showed that voxels in the intraparietal sulcus activated more strongly to stereo-surfaces than random disparity stereograms. Activations occurred in two foci: in caudal IPS (right: 9/13, left: 8/13 subjects) and in an IPS region between V7 and cIPS (right: 7/13, left: 6/13 subjects, Fig. 6d-left). In contrast, few voxels responded significantly more strongly to motion-surfaces than random-
motion (Fig. 6d-right). Even when we reduced our statistical threshold to $P < 0.05$ (voxel level, uncorrected) we found little activations for motion-surfaces vs. random motion. When they occurred, they partially overlapped MT (5/13 subjects), dorsal V3 (2/11) or V3a (4/11 subjects), rather than IPS regions. However, these differential responses were not significant across subjects (all $t$s < 0.2; $P > 0.45$, Fig. 5). Thus, we did not find evidence for any visual region that shows preferential responses to global surfaces relative to the random condition across motion and stereo cues.

Figure 7 approximately here

We extracted responses from IPS regions (Materials and Methods). Responses in the IPS region between V7 and cIPS (Fig. 7b) showed no main effect of cue, stimulus type or hemisphere ($F$s < 2.03, $P > 0.09$) or any interactions ($F$s < 0.45, $P > .77$). Responses were higher for stereo-surfaces than random stereo ($t(6) > 2.1$ $P < 0.04$), but were not higher for motion-surfaces than random motion ($t(6) = -.69$, $P = 0.74$).

cIPS responses (Fig. 7c) revealed a significant effect of stimulus type ($F_{4,139} = 3.93$, $P < 0.005$), cue ($F_{1,139} = 4.9$, $P < 0.03$) and hemisphere ($F_{1,139} = 4.8$, $P < 0.03$), but there were no significant interactions ($P > 0.23$). Responses were higher for stereo than motion stimuli and higher in the right than left hemisphere. During the stereo experiment responses were significantly higher for surfaces vs. random, edges vs. random, holes vs. random and objects vs. random disparity stereograms ($t$s(7) > 2.9, $P < 0.01$, Fig. 7c). However, responses for motion-surfaces were not significantly higher than random motion ($t(7) = 0.6$, $P = 0.37$). There were no significant differences across stereo-objects, holes, edges or surfaces ($t$s(7) < 1.9, $P > 0.06$). Thus, the higher responses for global surfaces compared to random occurred only for stereo stimuli. Overall, cIPS responded more strongly to stereo than motion stimuli and both stereo-surfaces and stereo-edges activated this region more than random disparity stereograms.
DISCUSSION

Shape Selectivity Across Multiple Cues in the LOC

Our data provide strong evidence for selectivity for shape-from-luminance, shape-from-motion and shape-from-stereo in the LOC: the LOC responded more strongly to object and holes than random dot stimuli, disconnected edges and global surfaces without edges across motion and stereo cues. However, LOC’s responses to edges or global surfaces were not different than random stimuli across both stereo and motion cues. These data provide strong evidence that LOC does not represent global surfaces or local edges. LOC’s similar activation for shapes defined from either motion or stereo cues may reflect either invariance at the single neuron level or invariance across the population of LOC neurons. That is, cue-invariance revealed by standard fMRI may reflect a homogeneous cue-invariant neural population or a mixture of heterogeneous neural populations selective to shape from different cues within fMRI voxels. Findings of cue-invariant responses to shapes in macaque IT neurons (Sary et al. 1993; Tanaka et al. 2001; Vogels and Orban 1996) and results of fMRI-adaptation studies in humans (Kourtzi and Kanwisher 2000) favor the former interpretation. Future fMRI-adaptation studies in humans are necessary for determining if the higher responses to motion objects and stereo objects vs. control stimuli involve the same neural population or distinct neural populations in the LOC. Overall, our data underscore the involvement of LOC in processing shape across multiple visual cues.

Processing of Edges Across Multiple Cues in the Human Visual Cortex

We found evidence for both cue-specific and cue-independent processing of edges. Selective responses to edges across motion and stereo stimuli overlapped the visual field map named V7. V7 responded more strongly to edges than random stimuli and surfaces for both motion and stereo cues. Activation to edges, objects and holes was similar, suggesting that disconnected edges, which do not form a global shape, are sufficient to robustly activate this region. This extends previous findings (Gilaie-Dotan et al. 2002; Grill-Spector et al. 1998) that
implicated this dorsal region in processing objects by suggesting that local edges alone are sufficient to activate it more than random stimuli. These selective responses to edges across multiple cues may be useful for detecting boundaries between regions in the visual input.

Cue-specific processing of edges occurred in V3a and V4: these regions responded more strongly to motion-edges than both random-motion and motion-surfaces, but they did not respond more strongly to stereo-edges than random stereo. Further, these regions responded more strongly to any stimulus that contained a motion edge (objects or hole) than random stimuli. Since the higher responses for edges occurred only for one cue, it suggests a cue specific response to motion edges.

**Cue Specific Representation of Global Surfaces**

We did not find any region that showed selective responses to global surfaces vs. random across motion and stereo cues. However, we found evidence for stereo-specific processing of global surfaces along the IPS. Notably, cIPS responded more strongly to stereo-surfaces than random disparity stereograms and also more strongly to stereo than motion stimuli. These results are consistent with previous studies suggesting the cIPS's role in processing stereo cues in humans and monkeys (Shikata et al. 2001; Taira et al. 2000; Tsao et al. 2003; Tsutsui et al. 2005). Our data suggests that cIPS is involved in processing stereo information and it also prefers stereo-surfaces and stereo-edges compared to random disparity stereograms.

It is possible that cIPS responses to stereo-surfaces were higher than motion-surfaces because of the greater perceived depth during the stereo experiments. However, this explanation is unlikely because random disparity stereograms and stereo-surfaces had similar depth, yet responses to stereo-surfaces were higher than responses to random stereograms. Another possibility is that small differences in eye movements across conditions may have contributed to differential responses in parietal activations (Schluppeck et al. 2005). However, it is not obvious why differential eye movements would occur only for stereo experiments. Future
experiments examining responses to surfaces defined from other cues that create a powerful depth perception such as texture gradients (Tsutsui et al. 2002), and measurements of eye movements will determine whether surface processing in cIPS is specific to stereo information.

A question that remains is why we did not find robust responses to global surfaces than random stimuli across visual cortex in general? One possibility is that regions involved in surface representation are involved in both local and global surface computations and therefore activate similarly for global surfaces and the random conditions (that may have local surface information). Another possibility is that the mean response across a visual region may not be higher for global surfaces than random stimuli or may not differ across various kinds of surfaces, but within these regions there is a subpopulation of neurons that is sensitive to surface information. fMRI-adaptation, for example, may reveal sensitivity to surface orientation or depth even when the mean response to surfaces of different orientations or depths is similar. Future experiments are required to distinguish between these possibilities and for a more comprehensive understanding of surface processing in visual cortex.

**Comparison of fMRI Data to Single Unit Studies of Early Visual Cortex and MT**

Analyses of low–level visual areas revealed that there were no significant differences between conditions validating that experimental conditions were matched for low–level properties. Our finding of shape selective activations in LOC across multiple cues, but not early visual cortex concurs with several neuroimaging studies (Appelbaum et al. 2006; Gilaiie-Dotan et al. 2002; Grill-Spector et al. 1998; Schira et al. 2004) and an electrophysiology study (Rossi et al. 2001). However, these results differ from electrophysiology studies that reported that V1 and V2 neurons respond more vigorously when a cell's receptive field overlaps an object boundary or lies within a shaped figure region (Lamme 1995; Qiu and von der Heydt 2005; Zhou et al. 2000; Zipser et al. 1996). One possibility for the discrepancy between neuroimaging and electrophysiology data is that modulations of neural responses during figure-ground segmentation are small (relative to the large initial nonspecific response) and occur in a minority
of neurons. Therefore, these modulations cannot be observed by fMRI that measures the pooled activity across a neural population. Alternatively, our random stimuli had more local variance in the range disparity or motion than the object or hole conditions and therefore produced more activation in V1 neurons that have small receptive fields. A third explanation, based on recurrent models of figure–ground segmentation (Jehee et al. 2007; Roelfsema et al. 2002), suggests that processing of the figure region occurs across multiple visual areas. According to this model, the area that will show the largest modulations during figure–ground computations is the region which neurons’ receptive fields match the size of the figure region. Since we used shapes which were ~11°, the largest effects would be observed in higher order areas (such as LOC) that contain neurons with relatively large receptive fields (Yoshor et al. 2007). Supporting this idea, electrophysiology data show that figure–ground effects in V1 occur only when the figure region is small (< 4°) and its edge is close to the cell’s receptive field (Rossi et al. 2001; Zipser et al. 1996).

We also found evidence for preferential responses to motion in MT, V2, and V3. These regions responded more strongly to all motion stimuli than stereo stimuli. Of the regions we measured, MT showed the largest cue specific responses. To what extent can we attribute higher MT responses to moving than stationary stereo stimuli to a cue effect? While our stimuli were matched for dot density, contrast, contour length and performance was similar across stereo and motion experiments (Table 2), it not possible to equate X units of motion to Y units of stereo. Nevertheless, higher MT responses for moving stimuli than stationary stereo may be explained by the properties of MT neurons. 96% of MT neurons are sensitive to the direction of motion, 99% to the speed of motion, and 93% to horizontal disparity (DeAngelis et al. 1998; DeAngelis and Uka 2003; Palanca and DeAngelis 2003). Most experiments that investigated MT responses to stereo used moving stimuli, except for the Palanca 2003 study. When Palanca and colleagues examined neural responses to moving stereo stimuli vs. stationary stereo stimuli
they found that 66% of MT neurons showed a significant reduction in their response to stationary stereo stimuli compared to moving stereo stimuli (to a level less than 1/3 of the response to moving stimuli). Since fMRI signals likely reflect the overall neural responses across space and time, lower firing rates for stationary than moving stimuli are likely to underlie the lower MT responses to stationary stereo stimuli. Nevertheless, a future quantitative study relating neural firing and fMRI responses for moving and stationary stereo stimuli is necessarily for validating this explanation.

The Role of LOC in Shape Computations

We found higher LOC responses for objects than random, surfaces and edges across motion and stereo cues. Importantly, we found no difference in LOC responses to global surfaces compared to random stimuli containing no structure for either motion or stereo cues. In a previous study, Stanley and colleagues (Stanley and Rubin 2003) created modified Kanizsa shapes via ‘pacmen-like’ inducers in which there was a clear percept of an enclosed region even when the bounding illusory contours were not perceived. Because they found that the LOC responded more strongly to enclosed regions than similar stimuli (created by rotating the pacmen), which did not enclose a region they concluded that the presence of a global surface is sufficient to activate LOC even when its illusory edges were not perceived. However, our data shows that global surfaces that are not enclosed do not activate LOC more than random stimuli, and that global shapes activate LOC more than global surfaces. Therefore, our data suggest that the pattern of LOC responses reported by Stanley and colleagues may have been driven by the processing of the global shape of the enclosed region rather than the processing of its global surface.

Our findings of higher LOC responses for stereo-objects vs. stereo-edges and higher responses for motion-objects vs. motion-edges (created by randomly scrambling object contours) extend previous results that reported higher LOC responses to luminance defined
global shapes vs. luminance defined random contours, misaligned contours or partial object contours (Altmann et al. 2003; Doniger et al. 2000; Lerner et al. 2004; Lerner et al. 2002). While the amount of dots that moved (or were in the near disparity) in the edges condition was lower than the object, hole or surfaces conditions, it is unlikely that this is factor explains lower LOC responses to edges. There were more dots that moved (or had near disparity) in the random or surfaces conditions than the object or hole conditions, but this did not produce higher LOC responses. Overall, our data provide evidence that LOC processes global shape information (rather than local edges) also for motion and stereo cues.

Notably, we found that LOC activated more strongly to objects than holes across both motion and stereo cues. This higher response to objects than holes occurred specifically in the LOC and more prominently in ventral subregions, which responded robustly to objects vs. random, but did not respond robustly to holes vs. random (Fig. 6b). We did not find differential responses between holes and objects in other visual regions. What is the source of this differential response? One possibility is that this pattern of responses reflects a bias for near disparity. However, this explanation is unlikely because both object and hole conditions had the same relative disparity and the contour defining the shape was in the nearer surface in both conditions. Alternatively, lower LOC responses to holes may reflect worse recognition of the shape of holes than the shape of objects because reduced LOC responses are correlated with lower accuracy during object recognition tasks (Bar et al. 2001; Grill-Spector et al. 2000; James et al. 2000). While we did not find performance differences between objects and holes in our experiments, differences may become apparent during tasks that require finer discriminations (e.g. distinguishing between a cow and a deer). A third possibility is that the difference between responses to objects and holes reflects the difference in the assignment of the border ownership of the contour defining the shape. In the object condition, the border is an intrinsic boundary and “belongs to the object”, while the border defining the hole is an extrinsic contour which “belongs” to the front occluding surface. This explanation suggests that information about border
ownership may be processed in the LOC. Nevertheless, explanations based on border ownership and recognition are not mutually exclusive. It is possible that the LOC participates in border ownership computations, and is optimized to process shapes defined by intrinsic contours (objects) rather than shapes defined by extrinsic contours (holes). Therefore, LOC responses are higher for objects than holes, and as a consequence the perception of objects may be more salient than the perception of holes. Further research is necessary to distinguish between these alternatives.

Implications for Theories of Figure-Ground Segmentation

Our results have important implications for theories of figure-background segmentation. Figure-ground segmentation is the process in which the figure region (a shaped front region) is segmented from the rest of the image (the shapeless “background”). Traditional theories of figure-ground segmentation propose a bottom-up computation, in which determining surfaces and their relative depth occurs before shape processing and this process also determines the border ownership of edges (Nakayama 1995; Nakayama et al. 1989). Nakayama and colleagues suggested that representation of global surfaces occurs in a visual region between V1 and inferotemporal cortex. Other theories suggest that figure-ground segmentation includes a top-down computation in which shape information is used during figure-ground processing (Peterson and Gibson 1994, 1993; Roelfsema et al. 2002; Vecera and O'Reilly 1998).

Several aspects of our data challenge traditional bottom-up theories of figure-ground segmentation. First, bottom-up theories suggest an intermediate region in which global surfaces are represented. However, we did not find evidence for cue-invariant representation of global surfaces. Second, for both motion and stereo cues, LOC activation to objects was higher than holes, but there was no difference in activation to objects and holes in other visual regions. This suggests that computations of border ownership (or computations of the figure region) may occur in higher stages of the visual hierarchy, such as regions involved in object and shape
perception, contrary to predictions of bottom-up theories. Therefore, our data suggest the involvement of object-selective cortex during figure-ground segmentation (Peterson and Gibson 1994, 1993; Roelfsema et al. 2002; Vecera and O'Reilly 1998).

In sum, we found evidence for selective responses to global shape across multiple visual cues in the LOC, evidence for both cue-specific and cue-independent responses to edges in intermediate visual areas and stereo-specific responses to global surfaces in IPS regions. Further, our results suggest that some types of information, specifically, motion-edges and stereo-surfaces, may be represented in a cue-specific manner in distinct cortical regions. Future experiments testing additional visual cues (e.g. color or texture) will advance our understanding of the nature of cue-specific processing in these regions. Overall, these data indicate that integration across multiple visual cues is mainly achieved at the level of shape and underscore LOC’s role in shape computations.
REFERENCES


Qiu FT and von der Heydt R. Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. Neuron 47: 155-166, 2005.

Roelfsema PR, Lamme VA, Spekreijse H, and Bosch H. Figure-ground segregation in a recurrent network architecture. J Cogn Neurosci 14: 525-537, 2002.


ACKNOWLEDGEMENTS

We would like to thank David Andersen, Bob Dougherty, David Remus, Rory Sayres, Nathan Witthoft & Jon Winawer for fruitful discussions and for their comments on earlier versions of this manuscript.

GRANTS

This research was funded by Whitehall Foundation 2005-05-111-RES grant to KGS and Stanford UROP grant to JV.
FIGURE LEGENDS

Figure 1: Experimental conditions. *Top:* schematic illustration of experimental conditions, darker regions indicate front surfaces, stimuli had no luminance edges (except for the screen border which was present during the entire experiment including blank baseline blocks). *Bottom:* examples of stereo conditions (use red/cyan glasses, red on left eye). Noise during printing may incorrectly render some stereo-surfaces to appear nonflat. To view full screen stereo and motion stimuli watch online Supplemental Movies 1 & 2. (a) Object on the front surface in front of a flat background plane. (b) Hole on the front surface in front of a flat background. (c) Disconnected edges in front of a flat background. (d) Two semitransparent flat surfaces at different depths. (e) Random stimuli with no coherent structure, edges, global surfaces or global shape. Random stimuli had the same relative disparity or depth range as other conditions.

Figure 2: Regions of interest on an inflated cortical surface of a representative subject. LOC ROIs were defined from the gray-level localizer experiment. MT was defined from the motion localizer experiment. Retinotopic visual areas were defined from meridian mapping experiments and dorsal ROIs were defined from the motion and stereo experiments.

Figure 3: Selective responses to objects across multiple visual cues in the LOC. *Top:* Statistical maps of objects > random (P < 0.005, voxel level) from gray-level (a), motion (b), and stereo experiments (c), on the inflated left hemisphere of a representative subject. LOC activated more to objects than random from all three cues. Visual meridians are represented by the red, blue and green lines: *red:* upper; *blue:* horizontal; *green:* lower. White contour: MT. (d) LOC ROIs defined from higher activation to objects than random in the gray-level localizer experiment.
Figure 4: LOC responses during stereo and motion experiments. (a) Average BOLD response in LOC ROIs (defined from the gray-level localizer experiment) during motion experiment 1 (dark gray) and stereo experiment 1 (light gray) across nine subjects. % Signal is plotted relative to the blank (fixation) baseline. (b) Average responses from LOC ROIs (defined from the gray-level localizer experiment) in motion and stereo experiment 2 averaged across eight subjects. In a & b, diamonds reflect significantly higher responses in a condition than the random condition, paired t-test across subjects. (c) Differential LOC responses between objects and other conditions, averaged across all 13 subjects (except for objects-edges that include 8 subjects from Experiment 2). Diamonds reflect differences that are significantly greater than zero, paired t-test across subjects. Error bars: standard error of the mean (SEM) across subjects. Acronyms: Surf: surfaces; Rand: random.

Figure 5: Activation of retinotopic visual areas and MT. Mean BOLD response amplitudes relative to a blank (fixation) baseline during motion and stereo Experiment 2 for retinotopic visual areas (7 subject average) and MT (8 subject average). Error bars indicate SEM across subjects.

Figure 6: Voxel-by-Voxel Statistical maps of activations compared to random. Voxel-by-voxel activation maps of each condition vs. random in each of the stereo and motion experiments were generated for each subject and displayed on flattened hemispheres, with a representative subject shown here. Statistical threshold is identical across all maps. Black: LOC ROIs: LO a, b, pFUS/OTS from the gray-level localizer experiment; White: MT; Purple: dorsal ROIs along the IPS, from posterior to anterior: V7, DF2: a region between V7 and cIPS, and cIPS; Visual meridians: red: upper, blue: horizontal and green: lower; Abbreviations: IPS: intraparietal sulcus; STS: superior temporal sulcus; OTS: occipito-temporal sulcus; COS: collateral sulcus.
**Figure 7: Mean BOLD responses in dorsal ROIs.** Mean BOLD responses are plotted relative to the blank (fixation) baseline for motion and stereo Experiment 2. (a) Average V7 responses across 6 subjects. (b) Average responses for an IPS region between V7 and cIPS across 6 subjects. (c) Average cIPS responses across 8 subjects. Error bars indicated SEM across subjects.

**Supporting Figure 1: Average LOC BOLD responses during the gray level localizer experiment.** Responses are averaged across all 13 subjects and plotted per category. Error bars: standard error of the mean (SEM) across subjects. Novel objects were abstract sculptures. Animals included photographs of 4-legged animals. All LOC ROIs respond significantly to all objects (including abstract objects) vs. scrambled objects.
Table 1: Talairach Coordinates and Size of ROIs. Data represent mean ± standard deviation across subjects.

(a) LOC ROIs

<table>
<thead>
<tr>
<th>ROI</th>
<th>Left Hemisphere</th>
<th>Right Hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>a</td>
<td>-39 ± 3mm</td>
<td>-79 ± 4mm</td>
</tr>
<tr>
<td>LO</td>
<td>-43 ± 5mm</td>
<td>-68 ± 4mm</td>
</tr>
<tr>
<td>pFUS/OTS</td>
<td>-40 ± 4mm</td>
<td>-56 ± 5mm</td>
</tr>
</tbody>
</table>

(b) Dorsal ROIs

<table>
<thead>
<tr>
<th>ROI</th>
<th>Left Hemisphere</th>
<th>Right Hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>V7</td>
<td>-27 ± 6mm</td>
<td>-79 ± 3mm</td>
</tr>
<tr>
<td>DF2</td>
<td>-19 ± 4mm</td>
<td>-79 ± 4mm</td>
</tr>
<tr>
<td>cIPS</td>
<td>-21 ± 6mm</td>
<td>-64 ± 10mm</td>
</tr>
</tbody>
</table>
Table 2: Behavioral Responses During Scan

Values reflect mean ± standard deviation. Asterisks indicate significantly different than other conditions, P < 0.05. Accuracy during Experiment 2 was higher for motion holes vs. other conditions, however, only motion-holes vs. motion-surfaces difference remained significant after Bonferroni correction.

**Experiment 1**

(a) stereo

<table>
<thead>
<tr>
<th></th>
<th>Object</th>
<th>Hole</th>
<th>Surfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>% correct</td>
<td>92 ± 10%</td>
<td>90 ± 11%</td>
<td>90 ± 13%</td>
</tr>
<tr>
<td>RT [ms]</td>
<td>753 ± 73</td>
<td>752 ± 82</td>
<td>761 ± 80</td>
</tr>
</tbody>
</table>

(b) motion

<table>
<thead>
<tr>
<th></th>
<th>Object</th>
<th>Hole</th>
<th>Surfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>% correct</td>
<td>96 ± 5%</td>
<td>96 ± 6%</td>
<td>98 ± 4%</td>
</tr>
<tr>
<td>RT [ms]</td>
<td>733 ± 113</td>
<td>782 ± 127</td>
<td>785 ± 141</td>
</tr>
</tbody>
</table>

**Experiment 2**

(a) stereo

<table>
<thead>
<tr>
<th></th>
<th>Object</th>
<th>Hole</th>
<th>Edges</th>
<th>Surfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>% correct</td>
<td>96 ± 3%</td>
<td>92 ± 5%</td>
<td>93 ± 5%</td>
<td>94 ± 4%</td>
</tr>
<tr>
<td>RT [ms]</td>
<td>797 ± 51</td>
<td>771 ± 102</td>
<td>754 ± 62</td>
<td>749 ± 85</td>
</tr>
</tbody>
</table>

(b) motion

<table>
<thead>
<tr>
<th></th>
<th>Object</th>
<th>Hole</th>
<th>Edges</th>
<th>Surfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>% correct</td>
<td>90 ± 6%</td>
<td>95 ± 3% *</td>
<td>89 ± 6%</td>
<td>83 ± 6%</td>
</tr>
<tr>
<td>RT [ms]</td>
<td>807 ± 141</td>
<td>808 ± 127</td>
<td>784 ± 95</td>
<td>788 ± 134</td>
</tr>
</tbody>
</table>
Figure 1: Experimental conditions. Top: schematic illustration of experimental conditions, darker regions indicate front surfaces, stimuli had no luminance edges (except for the screen border which was present during the entire experiment including blank baseline blocks). Bottom: examples of stereo conditions (use red/cyan glasses, red on left eye). Noise during printing may incorrectly render some stereo-surfaces to appear nonflat. To view full screen stereo and motion stimuli watch online Supplemental Movies 1 & 2. (a) Object on the front surface in front of a flat background plane. (b) Hole on the front surface in front of a flat background. (c) Disconnected edges in front of a flat background. (d) Two semitransparent flat surfaces at different depths. (e) Random stimuli with no coherent structure, edges, global surfaces or global shape. Random stimuli had the same relative disparity or depth range as other conditions.

127x49mm (420 x 420 DPI)
Figure 2: Regions of interest on an inflated cortical surface of a representative subject. LOC ROIs were defined from the gray-level localizer experiment. MT was defined from the motion localizer experiment. Retinotopic visual areas were defined from meridian mapping experiments and dorsal ROIs were defined from the motion and stereo experiments.
Figure 3: Selective responses to objects across multiple visual cues in the LOC. Top: Statistical maps of objects > random (P < 0.005, voxel level) from gray level (a), motion (b), and stereo experiments (c), on the inflated left hemisphere of a representative subject. LOC activated more to objects than random from all three cues. Visual meridians are represented by the red, blue and green lines: red: upper; blue: horizontal; green: lower. White contour: MT. (d) LOC ROIs defined from higher activation to objects than random in the gray-level localizer experiment.
Figure 4: LOC responses during stereo and motion experiments. (a) Average BOLD response in LOC ROIs (defined from the gray-level localizer experiment) during motion experiment 1 (dark gray) and stereo experiment 1 (light gray) across nine subjects. Signal is plotted relative to the blank (fixation) baseline. (b) Average responses from LOC ROIs (defined from the gray-level localizer experiment) in motion and stereo experiment 2 averaged across eight subjects. In a & b, diamonds reflect significantly higher responses in a condition than the random condition, paired t-test across subjects. (c) Differential LOC responses between objects and other conditions, averaged across all 13 subjects (except for objects-edges that include 8 subjects from Experiment 2). Diamonds reflect differences that are significantly greater than zero, paired t-test across subjects. Error bars: standard error of the mean (SEM) across subjects. Acronyms: Surf: surfaces; Rand: random.

Figure 4: LOC responses during stereo and motion experiments. (a) Average BOLD response in LOC ROIs (defined from the gray-level localizer experiment) during motion experiment 1 (dark gray) and stereo experiment 1 (light gray) across nine subjects. Signal is plotted relative to the blank (fixation) baseline. (b) Average responses from LOC ROIs (defined from the gray-level localizer experiment) in motion and stereo experiment 2 averaged across eight subjects. In a & b, diamonds reflect significantly higher responses in a condition than the random condition, paired t-test across subjects. (c) Differential LOC responses between objects and other conditions, averaged across all 13 subjects (except for objects-edges that include 8 subjects from Experiment 2). Diamonds reflect differences that are significantly greater than zero, paired t-test across subjects. Error bars: standard error of the mean (SEM) across subjects. Acronyms: Surf: surfaces; Rand: random.
Figure 5: Activation of retinotopic visual areas and MT. Mean BOLD response amplitudes relative to a blank (fixation) baseline during motion and stereo Experiment 2 for retinotopic visual areas (7 subject average) and MT (8 subject average). Error bars indicate SEM across subjects.
Figure 6: Voxel-by-Voxel Statistical maps of activations compared to random. Voxel-by-voxel activation maps of each condition vs. random in each of the stereo and motion experiments were generated for each subject and displayed on flattened hemispheres, with a representative subject shown here. Statistical threshold is identical across all maps. Black: LOC ROIs: LO a, b, pFUS/OTS from the gray-level localizer experiment; White: MT; Purple: dorsal ROIs along the IPS, from posterior to anterior: V7, DF2: a region between V7 and cIPS, and cIPS; Visual meridians: red: upper, blue: horizontal and green: lower; Abbreviations: IPS: intra-parietal sulcus; STS: superior temporal sulcus; OTS: occipito-temporal sulcus; COS: collateral sulcus.
Figure 7: Mean BOLD responses in dorsal ROIs. Mean BOLD responses are plotted relative to the blank (fixation) baseline for motion and stereo Experiment 2. (a) Average V7 responses across 6 subjects. (b) Average responses for an IPS region between V7 and cIPS across 6 subjects. (c) Average cIPS responses across 8 subjects. Error bars indicated SEM across subjects.

116x44mm (300 x 300 DPI)