fMRI Reveals That Non-Local Processing in Ventral Retinotopic Cortex Underlies Perceptual Grouping by Temporal Synchrony

Gideon P. Caplovitz,* Diego J. Barroso, Po-Jang Hsieh, and Peter U. Tse

Department of Psychological and Brain Sciences, Moore Hall, Dartmouth College, Hanover, New Hampshire

Abstract: When spatially separated objects appear and disappear in a synchronous manner, they perceptually group into a single global object that itself appears and disappears. We employed functional magnetic resonance imaging (fMRI) to identify brain regions involved in this type of perceptual grouping. Subjects viewed four chromatically-defined disks (one per visual quadrant) that flashed on and off. We contrasted %BOLD signal changes between blocks of synchronously flashing disks (Grouping) with blocks of asynchronously flashing disks (no-Grouping). Results: A region of interest analysis revealed %BOLD signal change in the Grouping condition was significantly greater than in the no-Grouping condition within retinotopic areas V2, V3, and V4v. Within a single quadrant of the visual field, the spatio-temporal information present in the image was identical across the two stimulus conditions. As such, the two conditions could not be distinguished from each other on the basis of the rate or pattern of flashing within a single visual quadrant. The observed results must therefore arise through nonlocal interactions between or within these retinotopic areas, or arise from outside these retinotopic areas. Furthermore, when V2 and V3 were split into ventral and dorsal sub-ROIs, ventral retinotopic areas V2v and V3v preferentially differentiated between the two conditions whereas the corresponding dorsal areas V2d and V3d did not. In contrast, within hMT+, %BOLD signal was significantly greater in the no-Grouping condition. Conclusion: Nonlocal processing within, between, or to ventral retinotopic cortex at least as early as V2v, and including V3v, and V4v, underlies perceptual grouping via temporal synchrony. *Hum Brain Mapp 29:651–661, 2008.*

Key words: top-down processing; form; feedback; retinotopic cortex; fMRI

INTRODUCTION

A fundamental characteristic of human vision is the tendency to perceive spatially separated items in the visual scene as belonging to the same object. Understanding how local image features on the retina are integrated into perceived global shapes is a central problem in the field of vision research. Decades of psychophysical research have suggested that particular patterns of feature relationships (such as common color or spatial proximity) mediate how individual image features will be "perceptually grouped" to create an emergent visual percept of a global shape [Koffka, 1935]. However, despite this wealth of psychophysical data, the neural mechanisms that underlie the various grouping processes remain largely unknown.

Contract grant sponsor: NIH; Contract grant numbers: R03 MH0609660-01, F31 EY016386; Contract grant sponsor: NSF; Contract grant number: 2005031192.

*Correspondence to: G.P. Caplovitz, Department of Psychological and Brain Sciences, Moore Hall, Dartmouth College, Hinman Box 6207, Hanover, New Hampshire 03755.
E-mail: gideon.p.caplovitz@dartmouth.edu.

Received for publication 7 November 2006; Revision 13 April 2007; Accepted 26 April 2007

DOI: 10.1002/hbm.20429
Published online 27 June 2007 in Wiley InterScience (www.interscience.wiley.com).

Recent work using neuroimaging techniques has begun to shed light on the neural mechanisms underlying perceptual grouping [Altmann et al., 2003; Harrison et al., 2007; Murray et al., 2002]. These studies have found that in addition to the lateral occipital complex (LOC), early retinotopically organized cortical areas subserve processes that group individual image features together. Huk and Heeger [2001] found that the motion selective area hMT+ also plays a role in perceptual grouping because %BOLD signal in hMT+ was modulated by coherently bound plaid stimuli in a way that could not be explained by the motion of the individual plaid components. However, because the rules by which perceptual grouping are mediated are so numerous and diverse, there is much regarding these processes that remains unknown.

Here we use functional magnetic resonance imaging (fMRI) to investigate the neural basis of temporal synchrony, which has been hypothesized to play a role in governing whether or not individual image features get grouped together into a coherent whole [Alais et al., 1998; Fahle, 1993; Guttman et al., 2005; Kojima, 1998; Parton et al., 2006; Ramachandran and Ramachandran, 1998; Usher and Donnelly, 1998; But see: Farid, 2002; Farid and Adelson, 2001]. Temporal synchrony is similar to the Gestalt grouping principle known as "common fate" [Koffka, 1935], according to which features that move coherently relative to each other have a tendency to be grouped together. Unlike common fate, however, temporal synchrony does not require the coherent motion of individual elements. An animation showing grouping by temporal synchrony similar to that tested in this article is available as online supplementary material.

Temporal synchrony, as investigated here, is defined by the simultaneous onset and offset of individual features within the visual scene. Elements that appear or disappear simultaneously have a tendency to be grouped into a common object. As such, grouping is driven as much by temporal relationships as by spatial relationships between individual elements in the image. We employ a region of interest (ROI) analysis based on the localization, in individual subjects, of early retinotopic areas, and the localization of the motion selective area hMT+, and the LOC, a nonretinotopically organized region that is selective for the form features that define objects.

MATERIALS AND METHODS

Subjects

All subjects had normal or corrected-to-normal vision and, prior to participating, gave written, informed consent according to the guidelines of the Department of Psychological and Brain Sciences, and the IRB of Dartmouth College. Subjects received $20 for their participation in each MRI scanning session. A total of 16 subjects participated in this experiment.

Stimuli

Stimuli were projected from a digital data projector (refresh rate 60 Hz) onto a frosted-Plexiglas screen outside the bore of the magnet, and viewed via a tangent mirror inside the magnet that permitted a maximum of $22^\circ \times 16^\circ$ visible area. The projected image was smaller than this and subtended approximately $17^\circ \times 12^\circ$ of visual angle. In each experiment there were two groups of stimuli used. Each group consisted of four solid circular green disks on a gray background (details below). Each disk had a radius of 1° of visual angle. Exactly one disk was positioned in each of the four quadrants of the visual field in the formation of a virtual square, centered at fixation, whose sides were 12° of visual angle in length. During each stimulation period, the four disks appeared and disappeared at randomly determined intervals. Each disk would appear at an average rate of approximately 2.5 Hz. Each time a disk appeared, it would remain on for 116.7 ms (seven frames at 60 Hz). In the Grouping condition, each of the four disks flashed on and off at the same time (i.e., using a common set of temporal intervals for each disk) in an example of temporal synchrony. In the no-Grouping condition an independent set of randomly determined intervals was applied to each of the four disks, causing them to flicker out of phase with each other, without introducing additional temporally periodic relationships. Viewed within a quadrant, the behavior of disks was not distinguishable between the Grouping and no-Grouping conditions (Fig. 1). The two conditions differed only in the temporal relationships between the individual disk onsets and offsets in the Grouping (simultaneous) and no-Grouping (typically nonsimultaneous) conditions. Perceptually, in the Grouping condition, the four disks were grouped together, such that a virtual square with one disk situated at each of its corners was perceived to flash on and off as a single unit. In the no-Grouping condition, the four disks were perceived to flash on and off independently of each other.

It is important to emphasize that because each disk was positioned within only one of the four quadrants of the visual field, in theory, the local information between the two conditions was identical with respect to sensory neurons whose receptive fields are retinotopically organized. That is to say, if only one of the four disks were viewed at any time, the two experimental conditions would be indistinguishable from each other. By isolating the disks to individual quadrants of the visual field, we can identify retinotopically organized brain regions that are sensitive to the nonlocal interactions that necessarily underlie perceptual grouping.

There are populations of neurons starting at the level of the retina (parasol ganglion cells or P-cells) that have transient response profiles and are thus highly sensitive to overall changes in luminance. One could argue that residual luminance energy from the disparate disks could "leak" across the vertical and horizontal meridians in an

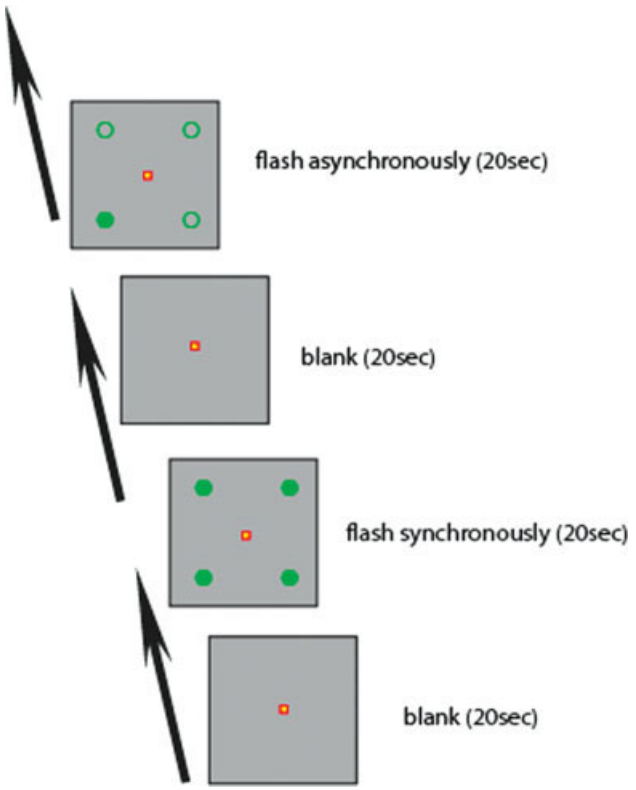

Figure 1.
Experimental Paradigm. Stimuli were presented in 20-s blocks, separated by 20-s blocks in which only the fixation task was presented (baseline). Each run consisted of 4 stimulus blocks: two Grouping, and two no-Grouping. In the Grouping condition, each of the four disks flickered on and off synchronously. In the no-Grouping condition, the four disks flickered on and off asynchronously. Observation of a single disk in isolation could not reveal whether the disk was a member of the Grouping or no-Grouping conditions. Subjects were instructed to maintain and perform a color-change detection task at fixation. (Open circles were not present during the actual experiment.) [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

asymmetrical fashion between the two stimulus types, creating differential temporal patterns of overall luminance levels in the image between the two conditions. To minimize this potential confound, rather than using luminance-defined stimuli, the stimuli used here were chromatically defined (green) and individually set to subjective isoluminance with the grey background for each subject. Unlike P-cells, the retinal ganglion cells (midget ganglion cells or M-cells) that are most sensitive to chromatic stimuli have sustained responses and are thus much less sensitive to the temporal dynamics in overall chrominance levels. Thus, the use of chromatic stimuli significantly reduces the likelihood that changes in BOLD signal between the two conditions will result from asymmetrical patterns in overall luminance levels in the stimuli. In addition to these primary stimuli, a central, stationary fixation spot was present at all times.

The Determination of Subjective Isoluminance

The luminance of the green stimuli used in the experiment was adjusted to become subjectively equal to the grey background for each subject using the minimal flicker technique [Anstis and Cavanagh, 1983]. Before the imaging data were collected, we presented a green flashing (30 Hz) square that subtended 1.5° visual angle in the center of a grey background and let the subjects adjust the green component of the square's color until minimal subjective flicker was reported. The color of the square was then fixed and applied to the stimuli for the duration of the experiment. This was carried out in the scanner to ensure that isoluminance was optimized for the actual experimental conditions that followed.

Fixation Task

Eye-movements, wakefulness, and attention to the fovea were controlled for by requiring subjects to perform a reaction-time task in which subjects had to press a button as fast as possible to a pseudorandomly occurring change in fixation point color. The fixation point subtended 0.25° × 0.25° visual angle, located at the center of the screen, which changed color from red to blue, or blue to red, approximately every 7.3 s. This color change occurred at an equal number of times during each block. This task could only be carried out successfully if subjects fixated during both condition and fixation-only blocks, and attended to the fixation point carefully.

Stimulus Presentation fMRI

Each subject participated in multiple runs (min = 5, max = 7) in the scanner. Each run lasted 180 s (72 volumes, TR = 2.5 s). In each run, there were four 20-s (8 volumes) stimulation blocks. These blocks were interleaved with five 20-s blank periods during which only the fixation spot was present. Thus each run began and finished with blank periods. In addition, 10 s (4 TRs) of dummy images were collected at the beginning of each run to allow for T1 equilibration effects.

In each of the stimulation blocks, one of the two stimulus groups was presented. In each run, two blocks of each condition were presented. The order in which the stimulus groups were presented was pseudorandomly determined for each run. Figure 1 illustrates an example of a single experimental run. The objects in each group were green and isoluminant to the background, which was always a constant and uniform grey, even during the blank period.

Data Acquisition

T1-weighted anatomical images were acquired using a high-resolution 3D spoiled gradient recovery sequence (SPGR; 124 sagital slices, TE = 6 ms, TR = 25 ms, flip angle = 25°, 1 × 1 × 1.2 mm^3 voxels) as well as a T1-weighted coplanar anatomical image with the same slice orientation as the EPI data, which was used for coregistration. Continuous whole-brain BOLD signal was acquired at the Dartmouth Brain Imaging Center on a GE 1.5T signa scanner using a standard head coil. Standard T2* gradient-weighted echoplanar functional images sensitive to BOLD contrast were collected using 25 slices (4.5 mm thickness and 3.75 × 3.75 mm^2 in-plane voxel resolution, interslice distance 1 mm, TR = 2500 ms, T2* echo time = 35 ms, flip angle = 90°, FOV = 240 × 240 × 256 mm^3, descending interleaved slice acquisition, matrix size = 64 × 64) oriented approximately along the anterior- and posterior-commissure plane. These slices were sufficient to encompass the entire brain of each subject.

Analysis of fMRI Data

Preprocessing

Data were analyzed using BRAIN VOYAGER (BV) 4.9.6 and MATLAB software developed in house. Effects of small head movements during and between runs were removed using BV's motion correction algorithms. Slice scan time correction was carried out to correct for the fact that slices were not collected at the same time and were collected in interleaved and descending order. Functional data were not smoothed in the space domain, but any low-frequency temporal fluctuations whose wavelength was greater than 24 TRs were removed. This did not introduce correlations between a voxel and its neighbors. For each subject, the functional data was coregistered to the high-resolution anatomical image and normalized into the Talairach stereotactic coordinate space, which enables comparisons to be made across subjects.

General linear model

Data from each experiment and each of the localizer scans were analyzed using the General Linear Model with a boxcar waveform convolved with a canonical hemodynamic response function. This returned a beta-weight parameter estimate for each condition for each voxel.

Constructing Region of Interest Masks

Retinotopic mapping

Retinotopy was carried out on each subject who participated in the study using standard phase-encoding techniques (4.5 mm thickness and 3.75 × 3.75 mm^2 in-plane voxel resolution, interslice distance 1 mm, TR = 1600 ms, flip angle = 90°, FOV = 240 × 240 × 256 mm^3, interleaved slice acquisition, matrix size = 64 × 64; 16 slices oriented along the calcarine sulcus) with the modification that two wedges of an 8 Hz flicker black and white polar checkerboard grating were bilaterally opposite like a bowtie, to enhance the signal to noise ratio [Sereno et al., 1995; Slotnick and Yantis, 2003]. Wedges occupied a given location for two TRs (3.2 s) before moving to the adjacent location in a clockwise fashion. Each wedge subtended 22.5° of 360°. Six TRs of dummy scans were discarded before each run to allow for T1 equilibration effects. One hundred sixty-eight volumes were collected on each run. A minimum of seven wedge runs were collected for each subject and then averaged to minimize noise before retinotopic data analysis in BV. A minimum of three runs was collected per subject using expanding 8 Hz flickering concentric rings that each spanned ~1° of visual angle in ring width. Each ring was updated after one TR (1.6 s) after which it was replaced by its outward neighbor, except that the outermost ring was replaced by the innermost ring, whereupon the cycle was repeated. Retinotopic areas (V1, V2d, V2v, V3d, V3v, V4v/VO1, and V3A) were defined as masks on the basis of standard criteria [Sereno et al., 1995], assuming a contralateral quadrant representation for V2d, V2v, V3d, and V3v, and a contralateral hemifield representation for V1, V4v/VO1, and V3A [Tootell et al., 1997]. V4v and the hemifield representation just anterior to it, called VO [Brewer et al., 2005], were combined into a common mask because the border between these regions was not distinct in all subjects.

Because the stimuli were symmetric across the visual quadrants, unified V2 and V3 masks were constructed by combining the voxels from V2v and V2d, and V3v and V3d, respectively, for each hemisphere.

Individual hMT+ mask

The analog of macaque motion processing area MT has been called V5 or human hMT+. Left and right hMT+ were localized in 14 of the 16 subjects who participated in the experiment using a localizer scan comprised of three to six runs of 3 min each. Because the localization scans were conducted on different days than the experimental data were collected, we were unable to collect localizers on all of the subjects who participated in the main experiment. The hMT+ localizer stimuli consisted of a grid of 3 × 3 subgrids of white squares on a black background whose length and height were approximately 1° × 1° visual angle. This was constructed by eliminating the zeroth, ± fourth, and ± eighth rows and columns from a regular grid of squares. Square centers were separated by approximately 3° of visual angle. In baseline blocks, the grid remained stationary for a 20-s epoch, followed by an epoch where the grid rotated clockwise around its center at a speed of 270° of rotational angle per second. Each run contained nine epochs of alternating motion and non-motion stimulation. As in the main experiment, subjects carried out a simple fixation task; pressing a button in the

right hand any time the fixation point changed color. hMT+ was localized as activity in the motion > nonmotion GLM contrast. In each case, a statistical threshold of at least $P < 0.0001$ corrected (fixed effects) was used. In addition, activation had to occupy the inferior occipital gyrus or inferior temporal sulcus to be localized as hMT+. The mean Talairach coordinates of hMT+ in the right hemisphere were $x = 45.0$ (S.E. = 0.9), $y = -66.1$ (0.9), and $z = 1.1$ (1.5), and in the left hemisphere: $x = -41.4$ (1.0), $y = -69.1$ (1.3), and $z = -3.1$ (1.3).

Individual lateral occipital complex masks

An individual LOC mask was also determined individually for 13 of the 14 subjects for whom individual hMT+ masks were made following standard procedures [Kourtzi and Kanwisher, 2000]. Object images ($7^\circ \times 7^\circ$) were placed on a white background, and were embedded within a black grid. Control images were comprised of the same images scrambled within the same grid. Their centroid position was updated randomly every TR within a 1° radius of the fixation point to prevent perceptual fading. The left and right hemisphere LOC masks were created from the fixed effects GLM analysis contrast of unscrambled objects > scrambled objects for each subject. The masks in each case were determined using a statistical threshold of at least $P < 0.001$ uncorrected. The mean left LOC mask location in Talairach coordinates was $x = -41.9$ (S.E. = 1.3), $y = -74.1$ (1.5), and $z = -4.0$ (2.2), and the mean right hemisphere LOC mask location was $x = 45.5$ (1.3), $y = -71.8$ (1.5), and $z = -8.2$ (1.7).[1]

In many subjects, there is an anterior portion of the LOC located in the middle Fusiform gyrus and a posterior portion located just inferior to hMT+ that is activated by this contrast. The present LOC masks were selected as the posterior region, since the two subregions were not abutting in any subject, and could well comprise areas with different functionalities.

Separation of hMT+ and LOC

In some subjects, a small number of voxels that were localized as being in hMT+ were also localized as belonging to LOC. This occurs because, depending on the threshold at which hMT+ and LOC masks are specified, there can be common voxels shared between these masks. To eliminate the possibility that measured responses were driven by this common overlap region in the hMT+ and LOC ROI masks, any overlapping voxels were removed from both of the corresponding hMT+ and LOC masks and were not included in the analysis.

[1]By way of comparison, the mean location reported by Kourtzi et al. [2003; left hemisphere Talairach coordinates: −41.9, −64,8, −2.7; right hemisphere: 39.1, −65.6, −12.0], found using an LOC localizer in each of ten subjects, fit well within the bilateral activations found in the present study.

Individual ROI Statistical Analysis

One mean beta-weight per condition was computed for each ROI of each subject. Beta-weights are the regressor weights obtained by separate GLM analyses carried out within each subject's ROI and reflect the degree to which the BOLD signal was modulated away from baseline during the corresponding stimulus condition. The ROI beta-weights were then averaged across hemispheres. To determine whether the stimuli produced a detectable BOLD signal within a given ROI, an initial two-tailed, one-sample *t*-test was conducted on the beta weights for each condition within each ROI. Subsequent analyses contrasting the beta weights across condition were restricted to only those areas in which at least one of the two beta weights was significantly ($P < 0.05$) nonzero. Two-tailed paired *t*-tests were then performed to determine if a statistically significant difference ($P < 0.05$) existed between the beta weights for the two experimental conditions. To compare effect-differences between different regions of interest, paired *t*-tests were performed on the difference scores computed with pairs of ROIs.

Whole-Brain Group-Based General Linear Model Analysis

In addition to performing an analysis based on individually localized ROIs, we performed a whole-brain analysis contrasting the computed beta-weights for the two conditions on a voxel by voxel basis across all 16 subjects. Although this analysis is highly susceptible to morphological and functional differences across subjects, it allows for the rough detection of %BOLD signal modulations within brain regions outside the cortical areas covered by the individually localized ROIs. This analysis was limited to only those voxels that were not included in the ROI analysis.

RESULTS

Fixation Task

Behavioral data were available from 15 of the 16 subjects who participated in the study. Subjects accurately responded to changes in the color of the central fixation spot within 2.5 s, on 82% (stderr = 4.5) of fixation changes with a mean reaction time of 857 ms (stderr = 50). Two subjects were identified as outliers because they responded on only 42 and 52% of fixation changes, respectively. The buttonpress mechanism was found to be malfunctioning when the data for these two subjects were collected, and we attribute their low accuracy to this. When the data from these two subjects are excluded, the remaining 13 subjects responded correctly to color changes in the fixation spot color within 2.5 s on 87.5% (min = 71%, max = 100%, stderr = 3.1) of fixation changes. This establishes that subjects were awake and fixating the fixation spot well.

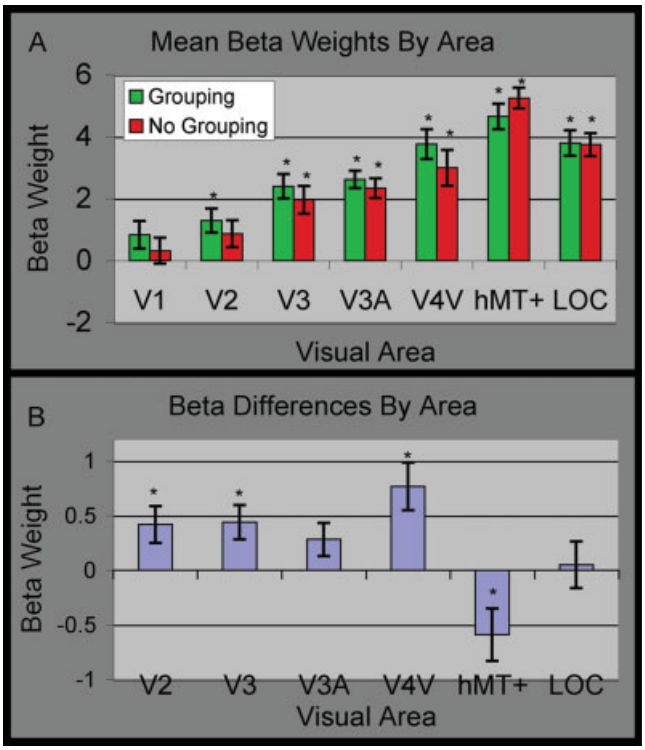

Figure 2.

Retinotopic ROI beta weight analysis. (**A**) Mean beta weights for each condition were computed within each ROI across all subjects. In order to be considered for later analysis, the mean beta weight for at least one of the two conditions needed to be significantly (* = $P < 0.05$) non-zero. All ROIs except for V1 satisfied this criterion. Error bars represent the standard error of the mean across subjects. (**B**) Mean difference values for each ROI were computed by subtracting the mean no-Grouping beta weight from the mean Grouping beta weight for each subject. V2, V3, and V4v all show a significantly greater (* = $P < 0.05$) beta weight in the Grouping condition. hMT+ on the other hand shows a greater beta weight in the no-Grouping condition. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Neuroimaging

Figure 2A illustrates the mean beta weights for each condition within each visual area. The initial beta weight analysis was designed to determine whether or not significant BOLD signal modulations driven by the stimuli themselves could be observed in the data. The one-sample *t*-test analysis indicated that in each ROI except for V1, at least one of the two mean beta weights was significantly nonzero.

Figure 2B illustrates the mean difference scores (Grouping–no-Grouping) across subject between the beta weights within each ROI. The results of the paired *t*-test analysis indicate that the BOLD signal response in the Grouping condition is significantly greater than the no-Grouping condition in V2 ($df = 15$, $t = 2.49$, $P < 0.025$), V3 ($df = 15$, $t = 2.8$, $P < 0.013$) and V4v ($df = 15$, $t = 3.511$, $P < 0.003$). Despite the fact that neither mean beta weight was significantly nonzero within V1, the paired *t*-test analysis showed a just significant difference between the two conditions ($df = 15$, $t = 2.25$, $P < 0.04$). In area hMT+ ($df = 13$, $t = -2.501$, $P < 0.027$), greater activation was observed in the no-Grouping condition in comparison to the Grouping condition. In areas V3a ($df = 15$, $t = 1.799$, $P > 0.09$) and LOC ($df = 12$, $t = 0.229$, $P > 0.82$) no significant differences were observed between the two conditions.

Our localization of visual areas V2 and V3 permitted us to construct independent V2v, V2d, V3v, and V3d masks for each hemisphere. Figure 3 illustrates the results of a secondary analysis in which we split V2 and V3 into ventral and dorsal ROIs. One-sample *t*-tests demonstrated that our initial criterion for conducting cross-condition comparisons in V2v, V2d, V3v, and V3d was satisfied within each of these sub-ROIs. The paired *t*-test analysis demonstrated that the

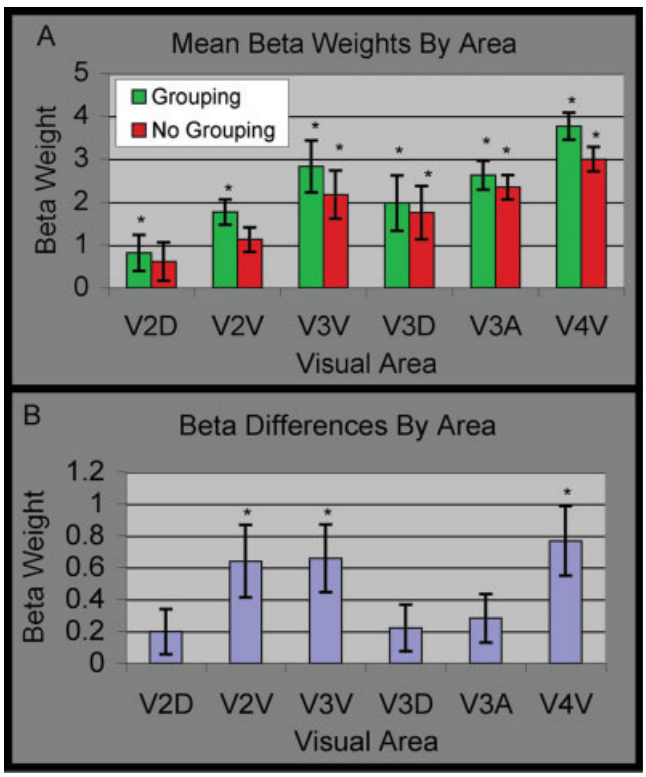

Figure 3.

Ventral/dorsal retinotopic ROI beta weight analysis. The V2 and V3 masks were divided into V2v/V2d and V3v/V3d masks respectively. (**A**) We see that although the BOLD signal modulations in each of these new submasks meet the initial statistical criterion (* = $P < 0.05$), (**B**) only the ventral ROIs V2v, and V3v show significant differences (* = $P < 0.05$) between conditions. This ventral/dorsal asymmetry is carried through to V3a (dorsal, not significant) and V4v (ventral, significant). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

initial differences between conditions observed in areas V2 and V3 were primarily driven by changes in the ventral ROIs V2v ($df = 15$, $t = 2.712$, $P < 0.016$) and V3v ($df = 15$, $t = 3.07$, $P < 0.007$) as compared to the dorsal ROIs V2d ($df = 15$, $t = 1.356$, $P < 0.159$) and V3d ($df = 15$, $t = 1.458$, $P < 0.166$), which did not reach statistical significance. Paired t-tests were performed between the difference scores for V2v/V2d and V3v/V3d. The results of these tests indicate that the differences between the Grouping and no-Grouping conditions were indeed greater in the ventral ROIs than in the dorsal ROIs (V2v/V2d: $df = 15$, $t = 2.185$, $P < 0.045$; V3v/V3d: $df = 15$, $t = 2.262$, $P < 0.039$).

Although the statistical values reported here would not survive Bonferroni correction for multiple comparisons in some ROIs, the modulations observed in areas V3v and V4v remain significant even after this conservative statistical correction. Although our findings in V3v and V4v are, therefore, highly unlikely to have arisen by chance, it would be incorrect to conclude that our findings in other areas necessarily arose by chance. In particular, the uncorrected P-value of less than 0.016 for V2v is suggestive that %BOLD signal is modulated by temporal synchrony in V2v as well as V3v and V4v. Indeed, the increasing significance of P-values from V1, through V2v, V3v, and V4v, suggests an increasing capacity to be influenced by nonlocal events in these retinotopic areas.

The results of the whole-brain GLM did not identify any brain areas outside those covered by the ROI analysis in which %BOLD signal was significantly ($P < 0.006$ uncorrected; minimum 50 mm spatial extent) modulated by the two stimulus conditions. Although clusters of greater activity in the Grouping versus no-Grouping conditions in the whole-brain GLM were indeed identified, none of these clusters were located outside the areas of cortex isolated by the individual localization procedures.

DISCUSSION

Here we have presented data from an experiment designed to determine whether perceptual grouping through temporal synchrony modulates %BOLD signal in key ROIs. We contrasted two conditions as follows: one where four chromatic isoluminant disks, one in each visual quadrant, flashed on and off synchronously, and the other, where they flashed on and off asynchronously. Because these conditions only differed in terms of the temporal relationships between the spatially isolated disks, the two conditions were identical in terms of local spatial and temporal statistics. That is, based on purely local information, a visually responsive neuron in retinotopic cortex would not be able to tell the difference between these two experimental conditions. This constrains the interpretation of any observed differences in BOLD signal across the two conditions and can thus be used to gain insight into the types of neural mechanisms and computations that integrate the events occurring in the four visual quadrants.

Summary of Results

On the basis of the data presented, we report three primary findings. First, BOLD signal at least as early as V2v[2], is modulated by perceptual grouping on the basis of temporal synchrony. We observed greater BOLD signal activity in V2v, V3v, and V4v when stimuli were presented in a synchronous manner (thereby creating a percept of a blinking square) as compared to when the stimuli were presented asynchronously.

hMT+ also reaches significance, but in the opposite direction, responding more to nonsynchronously blinking disks. We believe that it is likely that the increase in activation within hMT+ during the no-Grouping condition results from apparent motion signals that arise in the no-Grouping condition, as disks in neighboring quadrants flash on and off asynchronously. Indeed, recent research [Muckli et al., 2002] has demonstrated that %BOLD signal in hMT+ is activated more by apparent motion stimuli than by those that are perceived to be flashing.

Second, there is no difference in %BOLD signal change between conditions in the LOC. Despite the fact that the BOLD signal within the LOC was strongly activated by the stimuli of both conditions, there was no significant difference in the signal between the Grouping and no-Grouping conditions.

Third, there is an asymmetry in BOLD signal activation between ventral and dorsal retinotopic ROIs. The BOLD signal in ventral retinotopic areas V2v and V3v is modulated to a greater extent by perceptual grouping through temporal synchrony than the BOLD signal in the corresponding dorsal retinotopic areas V2d and V3d, which in fact failed to reach significance. This finding is unexpected, because it is widely believed that the only difference between ventral and dorsal retinotopic V2 or V3 is the region of space processed. Furthermore, the BOLD signal in the contralateral hemifield representation of V4v, a ventral retinotopic region, is also greater in the Grouping condition than in the no-Grouping condition, whereas no difference between conditions was observed in the contralateral hemifield representation of V3a, a dorsally situated retinotopic area.

[2]A significant difference between conditions was also observed in V1. However, even if this result is not due to chance (V1 does not reach significance after statistical correction for multiple comparisons) this result is difficult to interpret in light of the fact that the mean beta weight was not significantly different from zero for either condition alone relative to baseline. Inspection of the raw data within V1 revealed that in over a third of subjects (6 of 16), negative beta weights corresponding to BOLD deactivation away from baseline were found for both conditions. Thus in some subjects, differences observed between condition relates to differences in deactivation and to differences in activation in other subjects. Such inconsistencies were not observed in other areas. For this reason, we remain skeptical of the slightly significant, uncorrected difference between conditions observed in V1.

Implications of Primary Findings: V2 Modulated by Perceptual Grouping

We observed modulations of BOLD signal as a function of perceptual grouping through temporal synchrony within retinotopic cortex at least as early as V2. This result is surprising given the retinotopic organization of V2, and the local nature of receptive fields in this area. Each stimulus consisted of a single disk presented within each of the four visual quadrants. Individual disks always remained sufficiently far from the vertical or horizontal meridians to minimize the likelihood that neurons whose receptive fields straddled those meridians would respond to the stimuli under conditions of fixation. Moreover, a fixation task helped to ensure that subjects maintained central fixation. Because of the retinotopic organization of V2, no V2 neuron has a receptive field that could receive local stimulation from more than one of the disks. The statistics of the stimulus presentation across the two experimental conditions were identical within a given visual quadrant. As a result, the local information for each V2 neuron is identical between the two experimental conditions. This suggests that although the four disks are separated in space by a distance that would exceed the size of any V2 neuron receptive field, the temporal synchrony of the activity in separate quadrants of retinotopic areas, which would presumably be evoked by the temporal synchrony of the stimulus, can result in an enhancement of activity within extrastriate visual areas.

This observation leads to the conclusion that the modulations observed in the data are not due to local, feedforward processes (e.g. those neuronal projections which generally terminate in Layer 4 of V2, when cell bodies of projecting cells reside in Layers 2 and 3 of V1). This constrains the types of neural connectivity that could underlie perceptual grouping on the basis of synchrony to two nonmutually exclusive neuroanatomically defined categories of nonlocal connections, namely, lateral and/or feedback projections. Lateral axonal projections can originate in all cortical layers except Layers 1 and 4, can terminate in any cortical layer and generally connect corresponding layers of contiguous cortex. Feedback axonal projections, on the other hand, typically originate in Layers 5 or 6 and usually terminate in Layer 1 (and sometimes Layer 6, but avoiding Layer 4) and do not necessarily connect to adjacent cortical tissue [Johnson and Burkhalter, 1996, 1997; review: Salin and Bullier, 1995].

These distinct types of nonlocal connections in turn raise two nonmutually exclusive hypotheses for the neural basis of perceptual grouping through temporal synchrony. The first is that long-range lateral processing at least as early as V2, and perhaps even V1, can integrate information across the four visual quadrants and that this processing underlies the formation of early global shape representations that are then sent in a feed-forward manner to later visual areas (e.g., V4v, and occipital temporal areas) that are known to process complex feature configurations [Gallant et al., 2000; Kastner et al., 2000; Tanaka, 1996]. Supporting this hypothesis is the result that lesions of V2 disrupt perceptual grouping [Merigan et al., 1993]. Although the grouped objects were not separated in the Merigan et al. [1993] experiment by the large visual angle seen in the current experiment, there is evidence that there are V2 neurons that are affected by visual stimulation separated by as much as 12° of eccentricity [Zhang et al., 2005]. In this context, the information represented by the V2 neurons would include information about global form and perceived visual shape in addition to local stimulus-driven information, and thus mediate the early stages of object recognition [Bar et al., 2001; Gilaie-Dotan et al., 2002; Grill-Specter et al., 1998, 2000; Logothetis and Sheinberg, 1996]. The second hypothesis is that later visual areas with larger receptive fields that can pool information across the visual quadrants send feedback information [Bullier, 2001; Lamme and Roelfsema, 2000; Lamme et al., 1998a,b, 1999; Zipser et al., 1996] to retinotopic cortex as early as V2, or perhaps even V1. Because of the low temporal resolution of the BOLD signal, the present data cannot distinguish between these two possibilities. However, because the individual quadrants of the visual field are not represented by adjacent regions of cortex, it is unlikely that lateral connectivity, as defined above, within a contiguous portion of retinotopic cortex, is the sole basis of the observed differential activations in the Grouping and no-Grouping conditions. These observed differences in %BOLD signal activity therefore most likely arise because of feedback connectivity.

Our findings, that extrastriate retinotopic areas are modulated by perceptual grouping, are consistent with several neurophysiological studies [Allman et al., 1985; Gilbert, 1992, 1998; Lamme et al., 1998a,b] and neuroimaging studies [Altman et al., 2003; Murray et al., 2002] which suggest that recurrent mechanisms of visual processing, based either on local or feedback interactions, mediate global processing. Because the stimuli were presented in a fashion in which the local sources of information were segmented across the visual quadrants, we believe that it is less likely that long range lateral connections like those reported to exist within V1 [Gilbert, 1992, 1998; Kapadia et al., 1995, 1999] primarily drive the modulations that we observe. Rather, we believe that feedback mechanisms from later visual areas containing neurons with larger receptive fields most likely drive the modulations that we observe in our data. Such feedback connections have been previously hypothesized [Bullier, 2001; Lamme et al., 2000; Pascual-Leone and Walsh, 2001] to play a critical role in visual perception in general and perceptual grouping [Murray et al., 2004] in particular. Bullier [2001], for example, argues that local lateral connections are simply too slow to mediate rapid grouping effects over large visual angles, but that very rapid signal transfer from higher to lower areas could account for grouping effects, because neurons in higher areas typically have larger receptive fields than lower areas.

Implication of Primary Findings: No Difference in LOC

Despite large activations within the LOC in both experimental conditions, no difference in BOLD signal was observed between stimulus blocks in which the four disks were temporally grouped as compared to when they were asynchronous with each other. This result is somewhat surprising in light of neuroimaging data that has consistently shown the LOC to subserve both object recognition [Kanwisher et al., 1996; Kourtzi and Kanwisher, 2000, 2001; Kourtzi et al., 2003; Malach et al., 1995] and the processing of spatial form [Altman et al., 2003; Murray et al., 2004].

The reason why these two stimulus conditions are not distinguished by BOLD signal activity in the LOC remains unknown. One hypothesis is that form and positional information processed within the LOC can be temporally integrated to represent spatial relationships that reveal themselves over time. Indeed, the human visual system is fully capable of object recognition under circumstances where the spatial relationships between features of the object are revealed over time. Two classic examples of such object recognition are anorthoscopic perception [Helmholtz, 1867, 1925; Parks, 1965; Zöllner, 1862] in which an object moving behind a narrow aperture can be recognized, and the kinetic depth effect [Wallach and O'Connell, 1953] in which the 3D structure of an object is revealed as the viewpoint perspective is changed over time. Neuroimaging investigations of anorthoscopic perception [Yin et al., 2002], transformational apparent motion [Tse, 2006], and form-based trackable features [Caplovitz and Tse, 2007] have found the LOC to be responsive to motion percepts that depend on the temporal relationships between spatially defined form features of objects, suggesting that neural populations within the LOC can temporally integrate spatial and form information. On this account, no differential activation in LOC was observed because in both conditions the spatial relationships between the individual disks representing a global shape of a square would have been represented in the LOC.

Implications of Primary Findings: Ventral/Dorsal Asymmetries

In the data presented here, we observed greater differences in the BOLD signal between the two experimental conditions in ventral ROIs (V2v, V3v, V4v) than in dorsal ROIs (V2d, V3d, V3a). This result in part replicates previous findings [Altmann et al., 2003] in which ventral visual areas V1, V2, Vp (identical to our V3v), and V4v showed greater activation during stimulus conditions in which image elements were perceptually grouped. However, by explicitly separating V2 and V3 into ventral and dorsal ROIs, as well as examining the activations within the dorsal region V3a, we are able to extend upon the previous findings by directly examining the relationship between the ventral and dorsal retinotopic areas, and thus identify the asymmetry observed in the present data.

The observation that ventral ROIs show greater differentiation between the two experimental conditions is suggestive of a differential role in perceptual grouping between the "what" and "where/how" visual pathways [Goodale and Milner, 1992; Ungerleider and Mishkin, 1982]. However, there is, to our knowledge, no neurophysiological evidence linking functional specialization along the lines of ventral and dorsal processing streams and the ventral and dorsal regions of V2 and V3. All of the neurophysiological evidence that we are aware of suggests that the functional specifications of these regions are solely limited to differences in the quadrant of the visual field that they represent. However, recent neuroimaging studies using fMRI have reported asymmetries in ventral and dorsal activations in V1 and V2 [Mendola et al., 2006] and V3 [Altmann et al., 2003]. If such a functional asymmetry exists across these subregions, one might predict that they would lead to asymmetries in behavior between the upper and lower visual hemifields. There has been recent debate in the literature with respect to whether such hemifield-dependent behavioral asymmetries exist [Danckert and Goodale, 2001; Hagenbeek, 2002; He et al., 1996; Kahn and Lawrence, 2005; Previc, 1995; Rubin et al., 1996; von Grünau, 1994; Whishaw, 1994; but see: Binsted and Heath, 2005; Krigolsen and Heath, 2006].

It should be noted that even if there are not any differences in the local processing mechanisms of V2d and V2v, or V3d and V3v, there may be differences in the degree or type of feedback that the dorsal and ventral subregions receive. For example, there may be more attentional or other types of top–down feedback to the ventral subareas than to dorsal subareas. For example, if feedback projections predominantly originate in the inferior temporal and occipitotemporal regions, it could be that V2v, V3v, and V4v receive more feedback than corresponding dorsal areas simply because of greater proximity.

Care must be taken when trying to create a link between ventral/dorsal pathways of visual processing and the ventral and dorsal regions of early retinotopic cortex. Asymmetries such as those found in the data reported here could manifest themselves for reasons having nothing to do with ventral/dorsal stream processing. For example, increased BOLD signal can arise due to an increase in the spatial extent over which neural activity occurs or through increases in the magnitude of neural activity that occurs within a fixed spatial extent. Asymmetries in either the spatial distribution or relative magnitude of activity between the ventral and dorsal regions of early visual cortex could account for the ventral/dorsal asymmetry reported here.

CONCLUSIONS

We conclude that perceptual grouping through temporal synchrony is at least in part mediated by neural processing

in ventral retinotopic cortex at least as early as V2v, V3v, and V4v. Because of the relationship between the statistical properties of the stimuli and retinotopic organization of these visual areas, we conclude that the BOLD signal modulations we observe are likely due to feedback projections from higher visual areas, and that these feedback mechanisms may play a critical role in perceptual grouping and object recognition.

REFERENCES

Alais D, Blake R, Lee S-H (1998): Visual features that vary together over time group together over space. Nat Neurosci 1:160–164.

Allman J, Miezin F, McGuinness E (1985): Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. Annu Rev Neurosci 8:407–430; Review.

Altmann CF, Bülthoff HH, Kourtzi Z (2003): Perceptual organization of local elements into global shapes in the human visual cortex. Curr Biol 13:342–349.

Anstis SM, Cavanagh P (1983): A minimum motion technique for judging equiluminance. In: Mollon JD, Sharpe LT, editors. Colour Vision: Psychophysics and Physiology. London: Academic Press. pp 155–166.

Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM (2001): Cortical mechanisms specific to explicit visual object recognition. Neuron 29:529–535; Erratum in Neuron 2001; 30:299.

Binsted G, Heath M (2005): No evidence of a lower visual field specialization for visuomotor control. Exp Brain Res 162:89–94.

Brewer AA, Liu J, Wade AR, Wandell BA (2005): Visual field maps and stimulus selectivity in human ventral occipital cortex. Nat Neurosci 8:1102–1109.

Bullier J. (2001): Integrated model of visual processing. Brain Res Brain Res Rev 36:96–107; Review.

Caplovitz GP, Tse PU (2007): V3A processes contour curvature as a trackable feature for the perception of rotational motion. Cereb Cortex 17(5):1179–1189.

Danckert J, Goodale MA (2001): Superior performance for visually guided pointing in the lower visual field. Exp Brain Res 137:303–308.

Fahle M (1993): Figure-ground discrimination from temporal information. Proc R Soc Lond B Biol Sci 254:199–203.

Farid H (2002): Temporal synchrony in perceptual grouping: A critique. Trends Cogn Sci 6:284–288.

Farid H, Adelson EH (2001): Synchrony does not promote grouping in temporally structured displays. Nat Neurosci 4:875–876.

Gallant JL, Shoup RE, Mazer JA (2000): A human extrastriate area functionally homologous to macaque V4. Neuron 27:227–235.

Gilaie-Dotan S, Ullman S, Kushnir T, Malach R (2002): Shape-selective stereo processing in human object-related visual areas. Hum Brain Mapp 15:67–79.

Gilbert CD (1992): Horizontal integration and cortical dynamics. Neuron 9:1–13; Review.

Gilbert CD (1998): Adult cortical dynamics. Physiol Rev 78:467–485; Review.

Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzchak Y, Malach R (1998): A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum Brain Mapp 64:316–328.

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.

Guttman SE, Gilroy LA, Blake R (2005): Mixed messengers, unified message: Spatial grouping from temporal structure. Vision Res 45:1021–1030.

Hagenbeek RE, Van Strien JW (2002): Left-right and upper-lower visual field asymmetries for face matching, letter naming, and lexical decision. Brain Cogn 49:34–44.

Harrison LM, Stephan KE, Rees G, Friston KJ (2007): Extra-classical receptive field effects measured in striate cortex with fMRI. Neuroimage 34:1199–1208.

He S, Cavanagh P, Intriligator J (1996): Attentional resolution and the locus of visual awareness. Nature 383:334–337.

Helmholtz H von (1867/1925): Treatise on Physiological Optics, 3rd ed., Vol 3. New York: Dover Press.

Huk AC, Heeger DJ (2002): Pattern-motion responses in human visual cortex. Nat Neurosci 5:72–75.

Johnson RR, Burkhalter A (1996): Microcircuitry of forward and feedback connections within rat visual cortex. J Comp Neurol 368:383–398.

Johnson RR, Burkhalter A (1997): A polysynaptic feedback circuit in rat visual cortex. J Neurosci 17:7129–7140.

Kanwisher N, Chun MM, McDermott J, Ledden PJ (1996): Functional imagining of human visual recognition. Brain Res Cogn Brain Res 5:55–67; Review.

Kapadia MK, Ito M, Gilbert CD, Westheimer G (1995): Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. Neuron 15:843–856.

Kapadia MK, Westheimer G, Gilbert CD (1999): Dynamics of spatial summation in primary visual cortex of alert monkeys. Proc Natl Acad Sci USA 96:12073–12078.

Kastner S, De Weerd P, Ungerleider LG (2000): Texture segregation in the human visual cortex: A functional MRI study. J Neurophysiol 83:2453–2457.

Khan MA, Lawrence GP (2005): Differences in visuomotor control between the upper and lower visual fields. Exp Brain Res 164:395–398.

Koffka K (1935): Principles of Gestalt Psychology. New York: Harcourt, Brace & Co.

Kojima H (1998): Figure-ground segregation from temporal delay is best at high spatial frequencies. Vision Res 38:3729–3734.

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

Kourtzi Z, Kanwisher N (2001): Representation of perceived object shape by the human lateral occipital complex. Science 293:1506–1509.

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.

Krigolson O, Heath M (2006): A lower visual field advantage for endpoint stability but no advantage for online movement precision. Exp Brain Res 170:127–135.

Lamme VA, Roelfsema PR (2000): The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 23:571–579; Review.

Lamme VA, Super H, Spekreijse H (1998a): Feedforward, horizontal, and feedback processing in the visual cortex. Curr Opin Neurobiol 8:529–535; Review.

Lamme VA, Zipser K, Spekreijse H (1998b): Figure-ground activity in primary visual cortex is suppressed by anesthesia. Proc Natl Acad Sci USA 95:3263–3268.

Lamme VA, Rodriguez-Rodriguez V, Spekreijse H (1999): Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. Cereb Cortex 9:406–413.

Logothetis NK, Sheinberg DL (1996): Visual object recognition. Annu Rev Neurosci 19:577–621; Review.

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.

Mendola JD, Conner IP, Sharma S, Bahekar A, Lemieux S (2006): fMRI measures of perceptual filling-in in the human visual cortex. J Cogn Neurosci 18:363–375.

Merigan WH, Nealey TA, Maunsell JH (1993): Visual effects of lesions of cortical area V2 in macaques. J Neurosci 13:3180–3191.

Muckli L, Kriegeskorte N, Lanfermann H, Zanella FE, Singer W, Goebel R (2002). Apparent motion: Event-related functional magnetic resonance imaging of perceptual switches and States. J Neurosci 22:RC219.

Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002): Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci USA 99:15164–15169.

Murray SO, Schrater P, Kersten D (2004): Perceptual grouping and the interactions between visual cortical areas. Neural Netw 17:695–705; Review.

Parks TE (1965): Post-retinal visual storage. Am J Psychol 78:145–147.

Parton A, Donner TH, Donnelly N, Usher M (2006): Perceptual grouping based on temporal structure: Impact of subliminal flicker and visual transients. Visual Cogn 13:481–502.

Pascual-Leone A, Walsh V (2001): Fast backprojections from the motion to the primary visual area necessary for visual awareness. Science 292:510–512.

Previc FH, Breitmeyer BG, Weinstein LF (1995): Discriminability of random-dot stereograms in threedimensional space. Int J Neurosci 80:247–253.

Rogers-Ramachandran DC, Ramachandran VS (1998): Psychophysical evidence for boundary and surface systems in human vision. Vision Res 38:71–77.

Rubin N, Nakayama, K, Shapley R (1996): Enhanced perception of illusory countours in the lower versus upper visual hemifields. Science 271:651–653.

Salin PA, Bullier J (1995): Corticocortical connections in the visual system: Structure and function. Physiol Rev 75:107–154; Review.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995): Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Slotnick SD, Yantis S (2003): Efficient acquisition of human retinotopic maps. Hum Brain Mapp 18:22–29.

Tanaka K (1996): Inferotemporal cortex and object vision. Annu Rev Neurosci 19:109–139; Review.

Tootell RB, Mendola JD, Hadjikhani NK, Ledden PJ, Liu AK, Reppas JB, Sereno MI, Dale AM (1997): Functional analysis of V3A and related areas in human visual cortex. J Neurosci 17:7060–7078.

Tse PU (2006): Neural correlates of transformational apparent motion. Neuroimage 31:766–773.

Ungerleider LG, Mishkin M (1982): Two cortical visual systems. In Ingle DJ, Goodale MA, Mansfield RJ, editors. Analysis of Visual Behavior, Cambridge, MA: MIT Press. pp 549–586.

Usher M, Donnelly N (1998): Visual synchrony affects binding and segmentation in perception. Nature 394:179–182.

von Grünau M, Dube S (1994): Visual search asymmetry for viewing direction. Percept Psychophys 56:211–220.

Wallach H, O'Connell DN (1953): The kinetic depth effect. J Exp Psychol 45:205–217.

Whishaw IQ (1994): How do primates reach? Behav Brain Sci 17:173–174.

Yin C, Shimojo S, Moore C, Engel SA (2002): Dynamic shape integration in extrastriate cortex. Curr Biol 12:1379–1385.

Zhang B, Zheng J, Watanabe I, Maruko I, Bi H, Smith EL 3rd, Chino Y (2005): Delayed maturation of receptive field center/surround mechanisms in V2. Proc Natl Acad Sci USA 102:5862–5867.

Zipser K, Lamme VA, Schiller PH (1996): Contextual modulation in primary visual cortex. J Neurosci 16:7376–7389.

Zöllner (1862): Ueber eine neue Art anorthoskopischer Zerrbilder. Poggendorff's Annalen CXVII:477–484.