



Neural correlates of transformational apparent motion

P.U. Tse*

Department of Psychological and Brain Sciences, H. B. 6207, Moore Hall, Dartmouth College, Hanover NH 03755, Germany

Received 26 February 2005; revised 29 November 2005; accepted 20 December 2005

When a figure discretely and instantaneously changes its shape, observers typically do not perceive the abrupt transition between shapes that in fact occurs. Rather, a continuous shape change is perceived. Although this illusory “transformational apparent motion” (TAM) is a faulty construction of the visual system, it is not arbitrary. From the many possible shape changes that could have been inferred, usually just one is perceived because only one is consistent with the shape-based rules that the visual system uses to (1) segment figures from one another within a scene and (2) match figures to themselves across successive scenes. TAM requires an interaction between neuronal circuits that process form relationships with circuits that compute motion trajectories. In particular, this form–motion interaction must happen before TAM is perceived because the direction of perceived motion is dictated by form relationships among figures in successive images. The present fMRI study ($n = 19$) provides the first evidence that both form (LOC, posterior fusiform gyrus) and motion (hMT+) processing areas are more active when TAM is perceived than in a control stimulus where it is not. Retinotopic areas ($n = 10$), hMT+ ($n = 7$), and LOC ($n = 7$) were mapped in a subset of subjects. **Results:** There is greater BOLD response to TAM than to the control condition in V1 and all subsequent retinotopic areas, as well as in hMT+ and the LOC, suggesting that areas that process form interact with hMT+ to construct the perception of moving figures.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Transformational apparent motion; Form; Motion; MRI; Visual cortex; Retinotopy; MT; LOC; Line motion

Introduction

Over the past several years, researchers have investigated a new type of motion phenomenon that occurs when two spatially overlapping shapes presented discretely in time appear to transform smoothly from the first shape into the second as if the sequence

was animated using a succession of intermediate shapes, as shown in Fig. 1. The illusion of apparently smooth and continuous shape change was termed ‘transformational apparent motion’ (TAM; Tse et al., 1998) in order to contrast it with the standard ‘translational apparent motion’ described by Gestalt psychologists (e.g., Wertheimer, 1912; Kenkel, 1913; first described by Plateau, 1829). TAM occurs when a figure is abruptly flashed on next to an abutting static figure, causing the new figure to appear to smoothly extend from the static figure. Upon offset, it appears to smoothly retract into the static figure.

Tse and colleagues (Tse et al., 1998; Tse and Logothetis, 2002) showed that a set of parsing and matching principles that aids in determining figural identity within and between scenes holds for TAM and that this set approximately reduces to the nearest-neighbor principle for cases of translational apparent motion. Their behavioral research to date demonstrates that a stage of figural parsing and matching precedes the perception of motion in TAM displays. This stage of segmentation takes place on the basis of Gestalt-like grouping principles that process the spatiotemporal relationships among figures. In particular, the visual system appears to infer which figures at time 2 are derived from which figures at time 1 on the basis of contour and surface relationships. If a given figure has a different shape at times 1 and 2, a continuous deformation between those shapes is perceived, presumably because the new figure is inferred to be a change in the shape of an already existing figure. Figure formation and matching appear to primarily take place on the basis of good contour continuity (indicating figural identity between scenes) and discontinuity (indicating possible figural difference). The percept of motion is therefore in part the perception of how figures have translated and transformed their shape. The key point here is that form processing must temporally precede or at the very least accompany the motion processing that subserves the perception of TAM.

Whereas translational apparent motion is generally insensitive to shape and color constraints so long as the two stimuli presented remain within the optimal range of spatiotemporal offsets (see, e.g., Cavanagh et al., 1989), TAM is sensitive to such shape and color constraints because these can be used by parsing mechanisms to disambiguate figures in scenes that can only be ambiguously parsed otherwise. Moreover, TAM can occur over figures that have

* Fax: +603 646 1419.

E-mail address: Peter.Tse@dartmouth.edu.

URL: <http://www.dartmouth.edu/~psych/people/faculty/tse.html>.

Available online on ScienceDirect (www.sciencedirect.com).

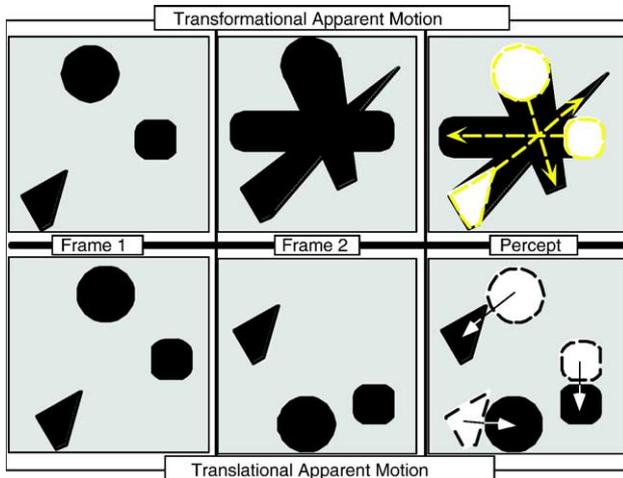


Fig. 1. Transformatational apparent motion (top row) is contrasted with translational apparent motion (bottom row). Note that this is just an example of TAM and is not the stimulus used in the experiment (see Fig. 2). When the figures in frame 1 (leftmost column) are replaced instantaneously with the figures in frame 2 (middle column), a percept (rightmost column) of TAM (i.e. smooth deformations, top row) results when figures spatiotemporally abut, but of rigid translations (bottom row) results when they do not abut. Note that even though frame 2 may consist of a single contiguous shape, as in the top middle frame, multiple motions can be perceived to occur simultaneously, as depicted by the yellow dashed lines on the top right. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

been defined purely by disparity (unpublished data) and by changes in successive subregions of a field of random dots (Hsieh and Tse, *in press*). Because there are no known motion-energy detectors tuned to motion defined by input across the eyes, we can conclude that the stage of form analysis that underlies TAM is not driven solely by an analysis of motion-energy. TAM reveals that parsed figures are matched to parsed figures from scene to scene in high-level motion processing, in some cases, violating nearest-neighbor principles.

The importance of figural parsing could not have been revealed by research into translational apparent motion because, in translational apparent motion displays, the parsing of each successive scene is generally given unambiguously in that figures are spatially distinct. In most translational apparent motion displays, a figure seems to disappear at one location and reappear at a different, nonoverlapping location some time later. While Kolers and Pomerantz (1971) did delineate various smooth transformations within the domain of translational AM, the problem in translational apparent motion experiments has generally been the match between figures, not the parsing of figures. However, in TAM displays, there is usually ambiguity in determining which figure in one scene has become which figure in the following scene because of the spatiotemporal overlap of succeeding figures. That is, in the case of the apparent shape transformations of figures, new image data generally appear without the disappearance of the figure(s) that existed in the previous scene. In TAM, the parsing problem has to be solved before the problem of matching or maintaining figural identity across successive scenes can be attempted. But, the visual system faces a problem of ambiguity in its efforts to correctly parse the image so as to coincide with the actual segmentation of the world

into independent, but abutting or overlapping, figures. Since many possible parsings are consistent with a single image, the visual system appears to have learned or evolved default processes for solving the parsing problem, realized, in part, by contour-based completion and segmentation of figures.

Goal of the present study

The goal of the present study is to localize the circuitry underlying the form–motion interactions that underlie TAM. Human neuroimaging studies have localized brain regions that process motion (area V5 or human hMT+; Zeki et al., 1991; Tootell et al., 1995) and have localized other areas that process form (lateral occipital complex, LOC, 1996; Malach et al., 1995). More recently, researchers may have found form-sensitive areas in hMT+ (Kourtzi et al., 2002), suggesting that some areas that process motion may also process object shape. Thus, to date, there is compelling evidence for areas that process motion (primarily hMT+) and areas that process form (primarily LOC), with the possibility that there are areas spanning the boundary of hMT+ and LOC that process both form and motion (Liu et al., 2004). However, there is no clear evidence to date that the two areas interact when solving a motion problem that can only be solved on the basis of form information. While it is suggestive that hMT+ and LOC are typically within a few millimeters of one another in the human brain, this by no means establishes that these areas interact.

The present study is the first to use fMRI to determine the neural basis of TAM in humans. A central goal of the present research is to determine the earliest neural correlates of TAM.

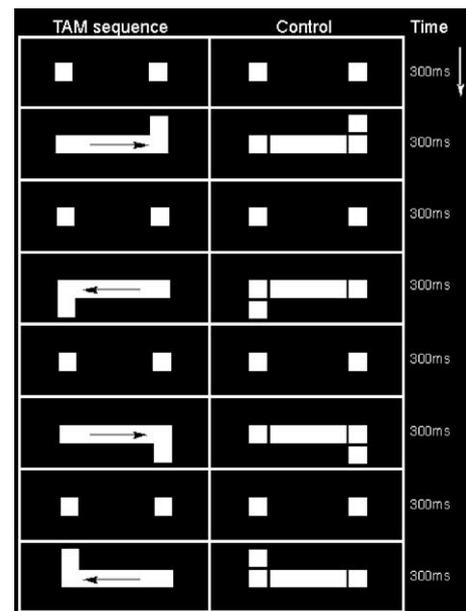


Fig. 2. Two conditions were tested, one where TAM was seen and the other where no TAM or other motion was seen. In both cases, the stimuli consisted of identically sized white squares ($1.68 \times 1.68^\circ$) and bars ($5.40 \times 1.68^\circ$) on a black background. In the TAM case, the bars and squares abutted. In the control case, they were separated by a small gap (0.32°). The direction of perceived motion is indicated with black arrows (no arrows were actually shown to observers). A small fixation point ($0.2 \times 0.2^\circ$, not shown) was located just below (0.30°) the lower contour of the bar.

The present study contrasted BOLD activation for a TAM stimulus and a no-TAM control. The stimuli for the TAM and control conditions are depicted in Fig. 2 and described in detail in the Materials and methods section. Each frame appeared instantaneously and consisted of identically sized squares and bars. The only difference between TAM and control stimuli was that the squares in the TAM condition abutted the bar, whereas they were shifted slightly (0.32°) away from the bar in the control condition. This introduced gaps (black lines) in the control condition that were not present in the TAM condition. Because of abutment in the TAM condition, the bar could be interpreted as a change in the state of the already existing squares. It is important to emphasize that, while smooth motion was perceived in the TAM case and not in the control case, both stimuli in fact consisted of the same squares and rectangles, each of which appeared or disappeared all at once. In the TAM case, they abutted one another, and, in the control case, they did not. A computational estimate of first-order motion generated by the image sequences (not shown here) indicated that there was no low-level difference in motion-energy between the two stimuli, even though motion was perceived in the TAM case and not in the control case. Furthermore, the stimuli were shown to each subject outside of the magnet, and all subjects spontaneously reported seeing smooth shape changes in the TAM condition and flicker in the control condition. The perceived direction of motion in the bar depended on the location of the vertically abutting square. If this square appeared on the right (left) side, motion was perceived to the right (left) across the entire length of the bar. This type of TAM stimulus was chosen because it required a new analysis of form relationships among successive images every 300 ms, evidenced by a new perceived direction of TAM on every image replacement. Because of non-abutment in the control condition, the bar could not be interpreted as a change in state of the squares. Instead, it could only be interpreted as a new object. It therefore appeared to just flash on, with no perceived translational or transformational apparent motion in any direction. This control was chosen because it consisted of the same geometric components but lacked any perceived motion, TAM or otherwise. Further rationales for this control are given in Materials and methods.

Materials and methods

Subjects

Nineteen healthy right-handed volunteers (of both genders between the ages of 18 and 40) were run in the TAM/no-TAM experiment. All gave informed consent within a protocol passed by the Dartmouth committee for the protection of human subjects and internal review board. All had normal depth perception and normal or corrected-to-normal visual acuity. Subjects were paid twenty dollars per session. All stimuli were presented binocularly.

Stimuli

Stimuli were projected from a digital data projector (refresh rate 60 Hz) onto a 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 $17^\circ \times 12^\circ$. The

stimuli consisted of white squares ($1.68 \times 1.68^\circ$) and a bar ($5.40 \times 1.68^\circ$) on a black background. In the TAM condition, the squares and bar abutted. In the control condition, the bar remained in the same location as in the TAM condition, but the squares were shifted 0.32° away from the bar horizontally and from each other vertically, creating a small black gap between them. There was a thin (one pixel wide, $6.67 \times 14.00^\circ$) yellow rectangle that appeared on the screen during all conditions that was used to help center the stimulus on the Plexiglas screen without having to present the TAM stimuli prior to the start of the experiment.

Experimental design

The experiment had a block design with thirteen conditions (3 TAM condition, 3 control, and 7 fixation) in 20-s blocks. The TAM or control stimuli cycled continuously within a block, as depicted in Fig. 2. Each run began with 10 s of dummy scans (four volumes which were discarded) to bring spins to baseline. Each run lasted a total of 4 min and 30 s. Condition order was either BTBCBTBCBTBCB or BCBTBCBTBCBTB on each run (B = blank plus fixation, T = TAM plus fixation, and C = control plus fixation). Subjects carried out a minimum of ten runs each and a maximum of fifteen. The first and last blocks were always fixation-only, and condition blocks were always separated by a fixation-only block. An entire cortical volume was scanned 8 times (TR = 2.5 s) per 20-s block.

Rationale for control stimulus

While a perfect control may not exist, a good control should be conservative. It should be chosen such that areas will not show greater BOLD activation to the test condition simply because the test condition has more low-level feature energy or is more likely to attract attention. Therefore, a good control for TAM should have the same or more low-level feature energy and should be equally or more likely to attract attention than the TAM condition. The control stimulus used here was judged to be a conservative control, in terms of contour and orientation energy, overall size, and in terms of potential confounds introduced by differential levels of attention in the two conditions. The control stimulus had more edges and oriented bars than the TAM stimulus, and, all else being equal, stimuli with more edges or oriented bars tend to drive early visual areas more than stimuli with fewer edges or oriented bars (e.g., Hubel and Wiesel, 1968; Kastner et al., 2000). If retinotopically defined areas are found to respond more to TAM than to this control, then this will not be due to the fact that the TAM stimulus has fewer edges since activation by the preponderance of edges alone would predict the opposite pattern of response (i.e. control > TAM). Moreover, because the squares were displaced slightly from the bar in the control condition, the overall size of the entire configuration of stimuli was slightly larger in the control case than in the TAM condition. Activation based purely upon configurational area would predict a greater response in the control condition.

The control stimulus had more discrete figures, and these discrete figures had phenomenally sudden onsets. Attentional allocation should be greater for the two or four shapes that appear in the control condition than for the one or two shapes that appear in the TAM sequence (Piazza et al., 2003). Moreover, attention increases when elements appear abruptly rather than gradually (Jonides and Yantis, 1988; Yantis and Jonides, 1990,

1996), so the control condition was also conservative with regard to attention.¹

fMRI data acquisition

Continuous whole-brain BOLD signal was acquired at the Dartmouth Brain Imaging Center on a GE 1.5 T signa scanner using a standard head coil. Standard T2*-weighted echoplanar functional images were collected using 25 slices (4.5 mm thickness and 3.75×3.75 mm in-plane voxel resolution, inter-slice distance 1 mm, TR = 2500 ms, flip angle = 90° , field-of-view = $240 \times 240 \times 256$ mm, descending interleaved slice acquisition, matrix size = 64×64) oriented approximately along the anterior commissure–posterior commissure plane. These slices were sufficient to encompass the entire brain of each subject. Cushions were used to minimize head motion. A T1-weighted anatomical image with the same slice orientation as the EPI was collected for each subject, as was a T2-weighted high resolution anatomical scan.

Fixation task

Eye movements, wakefulness, and attention to the fovea were controlled for by requiring subjects to perform a demanding reaction-time task in which the subject had to respond, within 500 ms (via button press), to a randomly occurring change in fixation point color using a button press. The fixation point was $0.2^\circ \times 0.2^\circ$, located 0.30° below the center of the horizontal bar. The fixation point changed color from blue/yellow to red/green on average about once every 1.5 s. This color change occurred an equal number of times during each block. Subjects were required to perform at 92.5% correct or better during each run or the run was not analyzed further.

¹ Even though the TAM stimulus also had abrupt onsets at the level of the stimulus and the pattern of retinal activation, phenomenally, there were no perceived abrupt onsets. Research has shown that attention cannot be allocated to the retinal image or other early ‘preattentive’ representations; Instead, attention takes as its inputs mid-level representations of completed surfaces (He and Nakayama, 1992; Rensink and Enns, 1998). This is not to say that attention cannot be directed to early retinotopic areas. Indeed, several studies have shown that attention can modulate the BOLD signal in V1 (Somers et al., 1999; Tootell et al., 1998; Brefczynski and DeYoe, 1999). However, psychophysical data (He and Nakayama, 1992; Rensink and Enns, 1998) nonetheless imply that attention can only be allocated to stimuli that have undergone preliminary processing at a preattentive stage. This is not a contradiction because preattentive processing may happen at a stage of V1 processing that is not accessible to attention. Attention may modulate later-stage representations within V1 or it may be that the attentional modulation observed in V1 occurs because of operations over representations that are different than the early representations shown to be inaccessible to attentional allocation by the psychophysical data. Because attention appears to take completed surfaces or shapes as its input, it is not likely that attention could take as its input the abrupt onset signals that define the TAM stimulus at the level of the retina. Rather, it is more likely that attention operates over the surfaces that are constructed from these stimuli by early visual processing, and there are no abrupt onsets at this level of representation for the TAM case. Thus, areas that respond more to the TAM stimulus than to the control stimulus are not likely to be responding more because the TAM stimulus draws more attention caused by sudden figural onsets. In any case, attention should have been largely taken up by the fixation task, which was identical across conditions.

Thus, subjects were only permitted an average of one miss or delayed response per block. No motor areas were found to be activated differentially between conditions, corroborating that the motor task was equivalent across all conditions. This task could only be carried out successfully if the subject was fixating during both condition and fixation-only blocks and attending to the fixation point carefully.

fMRI data analysis

Data were analyzed offline using BRAIN VOYAGER (BV) 4.9.6 and MATLAB software developed in house. Effects of small head movements were removed using BV’s motion correction algorithm. 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 order and descending order. Slices were corrected to have the same mean intensity. Functional data were not smoothed in the space domain, but low-frequency temporal fluctuations were removed through high-pass filtering. This did not introduce correlations between a voxel and its neighbors. The BOLD signal was quantified as the percent signal change away from the fixation baseline condition.

Retinotopic mapping

Retinotopic ROIs (V1, V2d, V2v, V3d, V3v, V3A/B, V4v) were determined on a different day on a subset ($n = 10$ of 19) of the subjects tested in the TAM experiment. Retinotopic mapping was carried out using standard phase-encoding techniques (Sereno et al., 1995; 4.5 mm thickness and 3.75×3.75 mm in-plane voxel resolution, inter-slice distance 1 mm, TR = 1600 ms, flip angle = 90° , field-of-view = $240 \times 240 \times 256$ mm, 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 signal to noise (Slotnick and Yantis, 2003). The stimuli were viewed on the same $17^\circ \times 12^\circ$ projected area as in all other experiments here. Wedges occupied a given location for 2 TRs (3.2 s) before moving to the adjacent location in a clockwise fashion. Each wedge subtended 18° of 360° . The initial 9.6 s (6 TRs of dummy scans) was discarded before each run to bring spins to baseline. 168 volumes were collected on each run. A minimum of 7 wedge runs were collected for each subject and then averaged to minimize noise before retinotopic data analysis in BV 4.9.6. A minimum of three runs were collected per subject using expanding 8 Hz flickering concentric rings that each spanned approximately 1° of visual angle in ring width. Each ring was updated after one TR (1.6s) 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, and V3A/B) were defined as masks on the basis of standard criteria (Sereno et al., 1995), assuming a contralateral quadrant representation for V1, V2d, V2v, V3d, and V3v, and a contralateral hemifield representation for V4v/VO, and V3A/B (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, as was true for the combination of V3A and V3B into a common V3A/B mask.

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 a subset of subjects tested in the main experiment ($n = 7$) using a localizer scan comprised of three to six runs of 3 min each. These were the same seven subjects in whom LOC was individually mapped. The hMT+ localizer stimuli consisted of a grid of 3×3 subgrids of solid white squares on a black background whose length and height were approximately $1^\circ \times 1^\circ$. This was constructed by eliminating the zeroth, \pm fourth, and \pm eighth rows and columns from a regular grid of squares. Square centers were separated by approximately 3° . 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° 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 to ensure fixation, pressing a button in the right hand any time the fixation point changed color. hMT+ was localized as activity in the motion > non-motion GLM contrast that survived the highly conservative threshold $P < 0.0001$ corrected (fixed effects). In addition, activation had to occupy the inferior occipital gyrus or inferior temporal sulcus in order to be localized as hMT+ (Watson and Ahumada, 1985). The mean Talairach coordinates of hMT+ in the right hemisphere were $x = 44.5$ (4.7), $y = -66.7$ (3.9), and $z = 1.0$ (2.5), and in the left hemisphere: $x = -41.3$ (2.4), $y = -70.8$ (3.9), $z = 0.8$ (2.8). In addition, an “MT-LOC” mask was created for each hemisphere of each subject by removing any voxels shared by the LOC and hMT+ mask from the hMT+ mask. These ‘subtraction’ masks were used because, depending on the threshold at which hMT+ and LOC masks are specified, there can be overlap voxels shared between these masks. In order to control for the possibility that the significant difference between the TAM and control condition seen here is driven solely by this overlap region, the overlap region was removed from the LOC and hMT+ masks for the seven subjects for whom individual LOC and hMT+ masks had been determined.

Individual LOC masks

An individual LOC mask was also determined individually for the same 7 subjects for whom individual hMT+ masks were made following standard procedures (Kourtzi and Kanwisher, 2000), and the TAM vs. control BOLD signal was determined in the LOC masks of these subjects and then averaged. Object images ($7^\circ \times 7^\circ$) were placed on a white background and were embedded within a black grid. Their centroid position was updated randomly every TR within a 1° radius of the fixation point in order to present perceptual fading. Control images were comprised of the same images scrambled within the same grid. The left and right hemisphere LOC masks were created from the fixed effects GLM analysis contrast of unscrambled objects > scrambled objects for each of seven subjects, at a P level of 0.001 uncorrected. The mean left LOC mask location in Talairach coordinates was $x = -44.5$ (SD = 4.8), $y = -69.1$ (1.8), $z = -12.8$ (3.8), and the mean right hemisphere LOC mask location was $x = 44.5$ (1.8), $y = -68.6$ (4.3), and $z = -11.1$ (6.9).²

² By way of comparison, the mean location reported by Kourtzi et al. (2003a) (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, fits well within the bilateral activations found in the present study.

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. In addition, an “LOC-MT” mask was created for each hemisphere of each subject by removing any voxels shared by the LOC and hMT+ mask from the LOC mask. This was done to eliminate the possibility that measured responses were driven by this common overlap region in the hMT+ and LOC ROIs.

Results: BOLD signal averaging

The average timecourse of the BOLD signal was determined for each subject’s retinotopic ROI separately, time-locked to the start of a block for a given condition, and averaged across runs for that subject. The BOLD signal for a given ROI was then reduced to a single number for each subject by averaging the BOLD signal from the second TR (5 s after TAM or control stimulus onset) to the tenth TR (5 s after TAM or control stimulus offset) to account for the approximately 5-s lag assumed to be introduced by the hemodynamic response function. Numbers were averaged across corresponding visual areas of the two hemispheres for each subject. These numbers were then averaged across subjects.

Averaged BOLD signal data within all the ROIs specified are summarized in Fig. 3. An asterisk is placed over areas where condition means differed significantly. Paired sample two-tailed t tests for matched samples revealed significantly greater BOLD activation in the TAM than control condition at the alpha equals 0.05 level in areas V1v (comparing left/right hemisphere averaged activity in TAM and control conditions, $P < 0.0244$, $n = 10$), V2v ($P < 0.0157$, $n = 10$), V3v ($P < 0.0188$, $n = 10$), V4v ($P < 0.0068$, $n = 10$), V3A/B ($P < 0.0274$, $n = 10$), hMT+–LOC ($P < 0.0036$; $n = 7$), and LOC–hMT+ ($P < 0.0333$, $n = 7$). If Bonferroni correction is applied for multiple comparisons in the several visual areas tested, only hMT+ reaches significance. In contrast, areas V1d ($P < 0.322$, $n = 10$), V2d ($P < 0.4216$, $n = 10$), and V3d ($P < 0.2332$, $n = 10$) did not reach significance, even with uncorrected statistics. This was almost certainly because these areas process the contralateral lower quadrant, and the stimulus was primarily confined to the upper hemifield. (The stimulus was placed in the upper hemifield because, had the bar component of the stimulus passed through the fixation point, it might have induced eye movements.) All other retinotopic areas process the upper hemifield, either because they only process the contralateral upper quadrant (V1v, V2v, V3v) or because they process the contralateral hemifield (V4v, V3A/B).

Results: Whole-brain general linear model analysis

Bonferroni correction within the $\sim 53,000$ voxels of a cortical mask (defined as the union of each individual subject’s cortical mask in Talairach space) at the $P < 0.05$ level ($n = 19$) left regions of significantly greater activation in the TAM condition than in the control condition bilaterally in area hMT+ only. However, the Bonferroni correction can be regarded as excessively conservative, effectively eliminating the possibility of false alarms at the expense of making excessively many incorrect rejections. A reasonable compromise between erring on the side of type 1 or type 2 errors is to locate voxels whose P values fall below a low uncorrected threshold of $P < 0.0005$, corresponding to a tolerated level for type

1 errors of 0.05%. At this threshold, one would expect only 27 voxels to reach significance by chance among the ~53,000 voxels examined within the cortical mask used, and one would not expect clustering. Areas that showed more BOLD signal in the TAM condition than the control condition at $P < 0.0005$ uncorrected (random effects, $n = 19$) are shown in Fig. 4. These were always differences among positive BOLD responses relative to fixation baseline. Note that the number of voxels in evidence is much greater than the 27 that would be expected to arise by chance. Areas that are evident include hMT+ and lateral occipital complex (LOC). Anatomically, LOC was ventral to hMT+ and located on the lateral bank of the inferior temporal gyrus extending both ventrally and dorsally, near the lateral occipital sulcus and in the ventral occipito-temporal region on the posterior fusiform gyrus near the occipito-temporal sulcus, matching the location of the LOC reported in past studies (e.g., Grill-Spector et al., 2001), except that no mid-fusiform gyrus activation was apparent here. In addition to the LOC, there was extensive posterior fusiform activation that may correspond to area V4.

General discussion

The present study is the first to use fMRI to determine the neural basis of transformational apparent motion (TAM). TAM comprises an important class of motion stimuli because TAM requires an interaction between areas that specify figures based on form cues with areas that compute motion trajectories of those

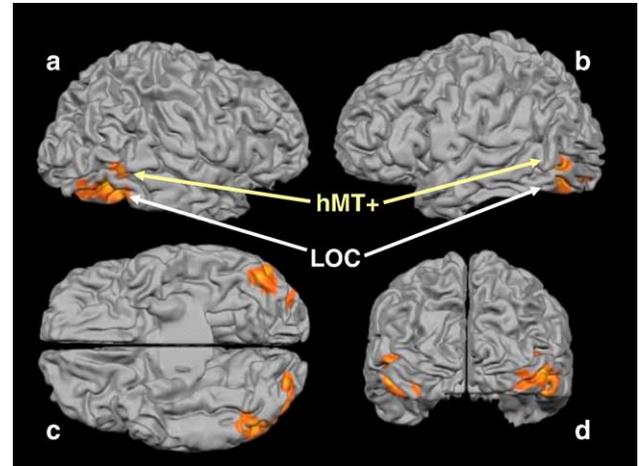


Fig. 4. Areas demonstrating greater BOLD response relative to the baseline of the fixation-only condition under the general linear model contrast TAM > Control for all subjects (random effects analysis, $n = 19$) at $P < 0.0005$ uncorrected are shown here superimposed upon one subject's cortical mesh (a = right hemisphere, b = left hemisphere, c = ventral, d = posterior). Areas survived Bonferroni correction within a cortex mask of approximately 53,000 voxels, but areas of significance were smaller, centered on hMT+ bilaterally. No areas responded more to the control than the TAM condition due to differences of positive BOLD activation (i.e. rises of the BOLD signal above baseline).³

figures. In particular, the form–motion interaction underlying TAM must happen before TAM is perceived because the perceived direction of TAM is dictated by form relationships among successive stimuli. The present fMRI study provides the first evidence that both form (LOC, posterior fusiform gyrus) and motion processing areas (hMT+) are more active when TAM is perceived than in a control stimulus where it is not. Moreover, there is greater BOLD response to TAM than to the control condition in V1 and all subsequent retinotopic areas. We can therefore conclude with confidence that the neural basis of TAM resides in these and other areas, such as the posterior fusiform revealed by the whole-brain GLM. Because TAM is thought to invoke high-level motion processing mechanisms, it is likely that high-level motion processing occurs in at least these areas.

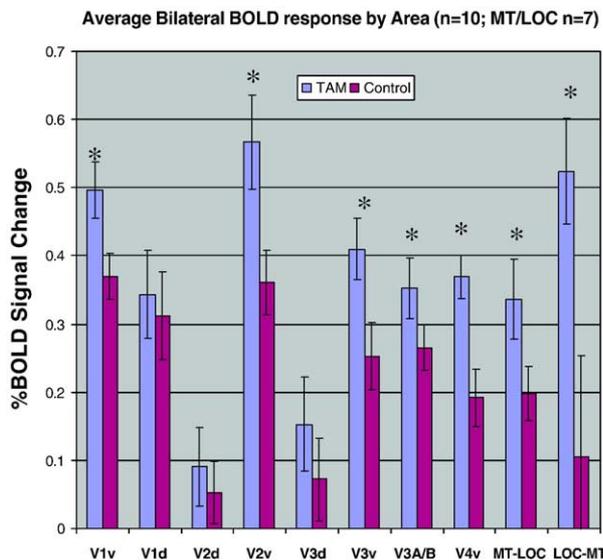


Fig. 3. Average BOLD signal as a function of area averaged across hemispheres and subjects. Retinotopic ROIs were determined separately for each subject. A separate individual MT localizer was carried out for seven of the subjects. An LOC mask was computed for the same seven subjects. An event-related average BOLD timecourse was calculated for each retinotopic area of each subject ($n = 10$) then averaged across hemispheres by area. Average BOLD activation for a given area and given subject was calculated by averaging the BOLD signal between the second and tenth TR positions in order to account for the lag introduced by the hemodynamic response function. Asterisks indicate areas where the average BOLD response in the TAM condition was significantly greater than that in the control condition at an alpha of 0.05. Error bars indicate standard errors of the mean.

³ There were no areas where the control condition revealed greater BOLD activation than the TAM condition, where both timecourses involved positive BOLD signal activation above the baseline level found in the fixation-only condition. Surprisingly, there was bilateral insula (Brodmann area 13) gray matter (peak activity: sublobar insula, Talairach coordinates: 40, -8, -3; claustrum: -37, -10, -3; -32, 3, 5) activation that was greater in the control than the TAM condition ($P < 0.0005$ uncorrected). However, this was a difference of negative BOLD activations (decreases from baseline relative to the BOLD activation level of the fixation-only condition). There is no established interpretation for such decreases from baseline. While it seems likely that neurons are capable of drawing more blood to themselves to replenish their needs for oxygen and glucose, it seems unlikely that they are able to push blood away. Instead, some authors suggest that BOLD signal deactivations may be due to shunting of blood supply from one region of the brain to another, where the limited supply of oxygenated blood is needed (Shmuel et al., 2002). Others suggest that BOLD deactivations are due to the active inhibition of neuronal processing (Smith et al., 2004). Interpretation of the present findings in the insula must wait until the causes and meaning of differences of BOLD signal deactivations during a task are better understood.

The LOC has been implicated in the processing of form (Malach et al., 1995; Kanwisher et al., 1996; Grill-Spector et al., 2001; Haxby et al., 2001). Evidence is emerging that LOC processes global 3D object shape, rather than local 2D shape features (Grill-Spector et al., 1998, 1999; Malach et al., 1998; Kourtzi and Kanwisher, 2000, 2001; Avidan et al., 2002; Mendola et al., 1999; Gilaie-Dotan et al., 2001; Moore and Engel, 2001; Kourtzi et al., 2003a,b), and may even mediate aspects of object recognition (Grill-Spector et al., 2000). Indeed, TAM itself occurs over 3D representations of form (Tse and Logothetis, 2002). The present fMRI data suggest that the LOC plays a greater role in TAM than the control. While both the TAM condition and the control condition involve global forms, the TAM condition may place a greater workload upon the LOC because the global form or global figural relationships that are believed to be computed there are presumably fed into motion-processing areas in the TAM case, whereas this output is not required for the control case. Another possibility with which the present data are consistent is that the motion processing that subserves the perception of TAM takes place in the LOC itself. To date, there have been no reports of motion processing in the LOC, making this a less likely alternative.

These findings contrast with recent fMRI findings that used translational apparent motion as a probe. Liu et al. (2004) found no difference between apparent motion and flicker conditions in any retinotopic area, although they did see greater activation for apparent motion than flicker in hMT+. They conclude from this that there is no evidence for the filling-in of features along the path of perceived motion in early retinotopic areas. Other fMRI studies have found greater activation for apparent motion than flicker in hMT+, but not in V1 (Goebel et al., 1998; Muckli et al., 2002). While both hMT+ and V1 contain motion-sensitive cells, the types of motion to which these cells respond need not be the same, perhaps accounting for this difference. For example, Mikami et al. (1986) found that neurons in MT but not in V1 responded to long-range apparent motion. Since the present data reveal differences between TAM and flicker in early retinotopic areas, it is possible that TAM, unlike translational apparent motion, does involve filling-in of features in early retinotopic areas.

In light of these past data, finding both form and motion processing areas more active in the TAM condition than the control in the present fMRI data suggests that the rapid analysis of form that must precede the perception of TAM most likely takes place in retinotopic areas, posterior fusiform areas that may not be retinotopic, the LOC, and perhaps parts of the hMT+ complex itself (Kourtzi et al., 2003a⁴). While the present data cannot specify the temporal dynamics of interactions among these areas, a reasonable model would place contour-based form analysis in the LOC and perhaps also V4v. The results of parsing would then be sent to hMT+ where motion trajectories would be computed in light of both form and motion-energy cues. Activation in early retinotopic areas may reflect top-down feedback involved in filling-in the figural trajectories that are experienced as the TAM illusion.

These findings suggest that hMT+ should be thought of as part of a form/motion processing stream of analysis rather than as an area just dedicated to the processing of motion and, in particular,

just motion-energy. Similarly, these findings suggest that the LOC, or at least the posterior LOC portion just inferior to hMT+ for which masks were made (see Materials and methods), should be thought of as part of a form/motion processing stream of analysis rather than as an area just dedicated to form processing. Indeed, a number of recent papers have come to the conclusion that there is potential anatomical and functional overlap between hMT+ and the LOC (Kourtzi et al., 2003a,b; Ferber et al., 2003; Murray et al., 2003; Zhuo et al., 2003; Liu et al., 2004) supported by behavioral data as well (Liu and Cooper, 2003; Stone, 1999). Because the form processing that underlies TAM must be extremely rapid, it would appear that LOC, hMT+, and other areas involved in form-motion processing operate in conjunction to solve the problem of what went where in TAM before TAM is perceived.

Acknowledgment

This project was funded by NIH R03 MH0609660-01.

References

- Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., Malach, R., 2002. Contrast sensitivity in human visual areas and its relationship to object recognition. *J. Neurophysiol.* 87, 3102–3116.
- Brewer, A.A., Liu, J., Wade, A.R., Wandell, B.A., 2005. Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat. Neurosci.* 8 (8), 1102–1109.
- Brefczynski, J.A., DeYoe, E.A., 1999. A physiological correlate of the ‘spotlight’ of visual attention. *Nat. Neurosci.* 2 (4), 370–374.
- Cavanagh, P., Arguin, M., von Grünau, M., 1989. Interattribute apparent motion. *Vision Res.* 29 (9), 1197–1204.
- Ferber, S., Humphrey, G.K., Vilis, T., 2003. The lateral occipital complex subserves the perceptual persistence of motion-defined groupings. *Cereb. Cortex* 13, 716–721.
- Gilaie-Dotan, S., Ullman, S., Kushnir, T., Malach, R., 2001. Shape-selective stereo processing in human object-related visual areas. *Hum. Brain Mapp.* 15, 67–79.
- Goebel, R., Khorrám-Sefat, D., Muckli, L., Hacker, H., Singer, W., 1998. The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur. J. Neurosci.* 10, 1563–1573.
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y., Malach, R., 1998. Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* 21, 191–202.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., Malach, R., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187–203.
- Grill-Spector, K., Kushnir, T., Hendler, T., Malach, R., 2000. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.* 3 (8), 837–843.
- Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. *Vision Res.* 41, 1422.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2330–2425.
- He, Z.J., Nakayama, K., 1992. Surfaces versus features in visual search. *Nature* 359 (6392), 231–233.
- Hsieh, P.-J., Tse, P.U., in press. Stimulus factors affecting Illusory Rebound Motion. *Vis. Res.*
- Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243.
- Jonides, J., Yantis, S., 1988. Uniqueness of abrupt visual onset in capturing attention. *Percept. Psychophys.* 43 (4), 346–354.

⁴ Note, however, that it seems worth treating the Kourtzi et al. result with caution since they found form-specific activation only along the inferior fringe of hMT+ and did not evaluate the degree of partial voluming that could have artifactually generated this overlap.

- Kanwisher, N., Chun, M.M., McDermott, J., Ledden, P.J., 1996. Functional imaging of human visual recognition. *Brain Res. Cogn. Brain Res.* 5 (1–2), 55–67.
- Kastner, S., De Weerd, P., Ungerleider, L.G., 2000. Texture segregation in the human visual cortex: a functional MRI study. *J. Neurophysiol.* 83 (4), 2453–2457.
- Kenkel, F., 1913. Untersuchungen über den Zusammenhang zwischen Erscheinungsgröße und Erscheinungsbewegung bei einigen sogenannten optischen Täuschungen. *Z. Psychol.* 67, 358–449.
- Kolers, P.A., Pomerantz, J.R., 1971. Figural change in apparent motion. *J. Exp. Psychol.* 87, 99–108.
- 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., Bühlhoff, H.H., Erb, M., Grodd, W., 2002. Object-selective responses in the human motion area MT/MST. *Nat. Neurosci.* 5 (1), 17–18.
- Kourtzi, Z., Erb, M., Grodd, W., Bühlhoff, H.H., 2003a. Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cereb. Cortex* 13 (9), 911–920.
- Kourtzi, Z., Tolia, A.S., Altmann, C.F., Augath, M., Logothetis, N.K., 2003b. Integration of local features into global shapes. Monkey and human fMRI studies. *Neuron* 37 (2), 333–346.
- Liu, T., Cooper, L.A., 2003. Explicit and implicit memory for rotating objects. *J. Exp. Psychol., Learn. Mem. Cogn.* 29, 554–562.
- Liu, T., Slotnick, S.D., Yantis, S., 2004. Human MT+ mediates perceptual filling-in during apparent motion. *NeuroImage* 21 (4), 1772–1780.
- Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci.* 92 (18), 8135–8139.
- Malach, R., Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y., 1998. Rapid shape adaptation reveals position and size invariance. *NeuroImage* 7, S43.
- Mendola, J.D., Dale, A.M., Fischl, B., Liu, A.K., Tootell, R.B.H., 1999. The representation of real and illusory contours in human cortical visual areas revealed by fMRI. *J. Neurosci.* 19, 8560–8572.
- Mikami, A., Newsome, W.T., Wurtz, R.H., 1986. Motion selectivity in macaque visual cortex: II. Spatiotemporal range of directional interactions in MT and V1. *J. Neurophysiol.* 55, 1328–1339.
- Moore, C., Engel, S.A., 2001. Neural response to perception of volume in the lateral occipital complex. *Neuron* 29, 277–286.
- Muckli, L., Kriegeskorte, N., Lanfermann, H., Zanella, F.E., Singer, W., Goebel, R., 2002. Apparent motion: event-related functional magnetic resonance imaging of perceptual switches and states. *J. Neurosci.* 22, RC219.
- Murray, S.O., Olshausen, B.A., Woods, D.L., 2003. Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cereb. Cortex* 13, 508–516.
- Piazza, M., Giacomini, E., Le Bihan, D., Dehaene, S., 2003. Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proc. Biol. Sci.* 270 (1521), 1237–1245.
- Plateau, J., 1829. Sur un nouveau genre d'illusion d'optique. *Correspondance Mathématique et Physique*. Bruxelles.
- Rensink, R.A., Enns, J.T., 1998. Early completion of occluded objects. *Vision Res.* 38 (15–16), 2489–2505.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268, 889–893.
- Shmuel, A., Yacoub, E., Pfeuffer, J., Van De Moortele, P.F., Adriany, G., Hu, X., Ugurbil, K., 2002. Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron* 36, 1195–1210.
- Slotnick, S.D., Yantis, S., 2003. Efficient acquisition of human retinotopic maps. *Hum. Brain Mapp.* 18 (1), 22–29.
- Smith, A.T., Williams, A.L., Singh, K.D., 2004. Negative BOLD in the visual cortex: evidence against blood stealing. *Hum. Brain Mapp.* 21 (4), 213–220.
- Somers, D.C., Dale, A.M., Seiffert, A.E., Tootell, R.B., 1999. Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl. Acad. Sci.* 96 (4), 1663–1668.
- Stone, J.V., 1999. Object recognition: view-specificity and motion-specificity. *Vision Res.* 39, 4032–4044.
- Tootell, R.B., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., Belliveau, J.W., 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* 15 (4), 3215–3230.
- Tootell, R.B., Mendola, J.D., Hadjikhani, N.K., Ledden, P.J., Liu, A.K., Reppas, J.B., Sereno, M.I., Dale, A.M., 1997. Functional analysis of V3A and related areas in human visual cortex. *J. Neurosci.* 17 (18), 7060–7078.
- Tootell, R.B., Hadjikhani, N., Hall, E.K., Marrett, S., Vanduffel, W., Vaughan, J.T., Dale, A.M., 1998. The retinotopy of visual spatial attention. *Neuron* 21 (6), 1409–1422.
- Tse, P.U., Logothetis, N.K., 2002. The duration of 3-D form analysis in transformational apparent motion. *Percept. Psychophys.* 64 (2), 244–265.
- Tse, P.U., Cavanagh, P., Nakayama, K., 1998. The role of parsing in high-level motion processing. In: Watanabe, T. (Ed.), *High-level Motion Processing: Computational, Neurobiological, and Psychophysical Perspectives*. MIT Press, Cambridge, MA, pp. 249–266.
- Watson, A.B., Ahumada Jr., A.J., 1985. Model of human visual-motion sensing. *J. Opt. Soc. Am. A* 2, 232–242.
- Wertheimer, M., 1912. Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.* 61, 161–265; 1961. English translation: “Experimental studies on the seeing of motion”. In: Shipley, T. (Ed.), *Classics in Psychology*. Philosophical Library, New York, pp. 1032–1088.
- Yantis, S., Jonides, J., 1990. Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *J. Exp. Psychol. Hum. Percept. Perform.* 16 (1), 121–134.
- Yantis, S., Jonides, J., 1996. Attentional capture by abrupt onsets: new perceptual objects or visual masking? *J. Exp. Psychol. Hum. Percept. Perform.* 22 (6), 1505–1513.
- Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C., Frackowiak, R.S., 1991. A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* 11 (3), 641–649.
- Zhuo, Y., Zhou, T.G., Rao, H.Y., Wang, J.J., Meng, M., Chen, M., Zhou, C., Chen, L., 2003. Contributions of the visual ventral pathway to long-range apparent motion. *Science* 299, 417–420.