ELSEVIER

BRAIN
RESEARCH

Research Report

Visual motion and the neural correlates of event perception

Jeffrey M. Zacks[a,b,*], Khena M. Swallow[a], Jean M. Vettel[a], Mark P. McAvoy[b]

[a]Department of Psychology, Washington University, 1 Brookings Drive, Saint Louis, MO 63130, USA
[b]Department of Radiology, Washington University, Saint Louis, MO, USA

ARTICLE INFO

ABSTRACT

People perceive ongoing activity in terms of discrete temporal events. Distinctive changes in the movement of objects or actors may contribute to the perception that one event has ended and another has begun. However, little is known about the quantitative contribution of movement information to the processing of events. This study investigated how movement features are related to the neural processing of events by performing functional magnetic resonance imaging while participants viewed simple animations of moving objects. After the imaging session, participants watched the animations again and segmented them into meaningful events. Movement features were systematically related to viewers' perceptual segmentation and to cortical activity throughout visual processing areas. Activity in the MT complex, which is known to be specialized for processing motion, increased with increases in the objects' speed. The perception of an event boundary was associated with transient changes in the MT complex and in a nearby region in the superior temporal sulcus associated with processing biological motion. Other movement features were associated with changes in activity in occipital, parietal, and frontal cortex. These results indicate a role for movement features in the perceptual processing of meaningful events, and in the neural basis of that processing.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Visual information processing has two striking features. First, it is multidimensional. For example, when a person observes other creatures, those creatures' speed and distance covary in complex ways that are structured by constraints arising from the physics of solids, from biomechanics, and from the neurobiology and psychology of action planning. Second, visual information processing is inherently dynamic. Information is extracted in real time as the visual environment continuously changes. However, for practical reasons of experimental control, systems neuroscience has focused on dramatically simplified situations in which only one or two dimensions are manipulated, and in which the temporal dynamics of the stimulus is restricted to a series of isolated "trials" (Bartels and Zeki, 2004a).

The multidimensional, dynamic structure of visual information processing has been taken up by research in social and cognitive psychology. In particular, a small body of work has focused on the question of how people parse an ongoing stream of activity into meaningful events (for a review, see Zacks and Tversky, 2001). This ability, which we will refer to as

event structure perception, may provide a window into visual information processing in general, particularly into its dynamic nature. At the same time, a better understanding of visual information processing may answer questions about event structure perception.

The perception of event structure can be measured by asking participants to divide an ongoing activity into perceptual units (Newtson, 1973). For example, the activity of making a bed might be segmented into two events: stripping off the dirty sheets, and placing on the clean sheets. Each of these might be broken down in turn into sub-events. Observers show good agreement about the locations of boundaries between events and good test–retest reliability, but also show stable individual differences (Speer et al., 2003). People adaptively modulate the grain at which they segment activity (Cohen and Ebbesen, 1979; Graziano et al., 1988; Newtson, 1973; Wilder, 1978a,b), and the size of the grain at which one segments may have consequences for later memory (Hanson and Hirst, 1989, 1991; Lassiter and Slaw, 1991; Newtson and Engquist, 1976).

Recent neuroimaging studies indicate that a specific neural system undergoes transient changes at event boundaries during ongoing perception. In one experiment, participants passively viewed movies of everyday activities while brain activity was recorded with functional MRI (fMRI) (Zacks et al., 2001a). After the initial passive viewing, participants watched the movies again, and segmented them to identify meaningful event boundaries. A network of regions including extrastriate visual cortex bilaterally and right posterior frontal cortex showed transient increases in activity immediately following event boundaries during the passive viewing of everyday activities. A subsequent study confirmed that the posterior regions included the human MT complex (MT+) (Speer et al., 2003), a brain region that is selectively activated by motion and is thought to be homologous to areas in the monkey that include cells which are selective for the direction and velocity of movement (Tootell et al., 1995).

The observation of MT+ activity at event boundaries is consistent with the hypothesis that distinctive movement features play a role in event structure perception (Newtson et al., 1977). In a recent study (Zacks, 2004), observers watched simple computer-generated animations of two objects moving in two dimensions against a plain background. They were asked to segment the movies into meaningful events by pressing a button to mark the boundaries between events, while the movie continued to play. Observers marked either the smallest units that were meaningful to them (*fine* segmentation) or the largest (*coarse* segmentation). The position, velocity, and acceleration of the objects, as well as their distance, relative velocity, and relative acceleration were calculated from the stimulus movies. Three experiments found that the probability of identifying segment boundaries was related to these movement features, particularly for fine-grained segmentation.

The posterior activations observed during event boundaries in the previous fMRI study (Zacks et al., 2001a) included not only MT+, but also adjacent regions in the posterior superior temporal sulcus. These activations may have included an area we will refer to as pSTS, which is selectively activated by biological motion (Grossman et al., 2000). This region does not show large fMRI signal changes in response to rigid motion, but responds vigorously to nonrigid biological motion (Beauchamp et al., 2003), and may prefer movement indicative of intentional activity to similar movement that does not provide evidence for intentions (Saxe et al., 2004). The possibility that activity in pSTS may increase during event boundaries is consistent with previous observations linking event segmentation to actors' goals (Baldwin and Baird, 2001).

Thus, the available evidence indicates that people are disposed to perceive complex dynamic activity in terms of discrete events, that the perception of event boundaries is correlated with transient responses in extrastriate visual cortex, and that the perception of event boundaries is related to movement information. These facts indicate the importance of visual information processing for understanding event structure perception, and also suggest that the methods used to study event perception may provide a valuable leverage point for understanding the multidimensional and dynamic characteristics of vision. In the following section we describe an experiment that used event perception methods to address two questions. First, how do movement features correlate with brain activity during ongoing events? In particular, we were interested in whether areas known to be specialized for motion processing during simple brief trials respond to motion signals in temporally extended multidimensional stimuli. Second, do areas involved in processing motion, and biological motion in particular, respond to event boundaries during events that possess neither animate actors nor intentional actions?

To address these questions, we asked participants to view simple abstract animations while recording brain activity with fMRI. The animations depicted the movements of two small objects (Fig. 1). Participants were told either that the animations depicted two objects being controlled by two people playing a video game (the *intentional* condition), or that the animations were generated randomly (the *random* condition). In fact, all animations were generated randomly; however, previous research suggests top-down inferences about the animate nature of movement sequences can affect event segmentation (Bartels and Zeki, 2004a) and neural processing (Baldwin and Baird, 2001).

Because the animations showed only two objects moving on a white background, they could be completely characterized by specifying the changes in position of the objects over time. These position data were used to calculate a number of features of the objects' movement over time, including their velocity, acceleration, distance, and relative velocity and relative acceleration. These features were used to estimate the neural response to properties of movement in ongoing events.

Following the initial scanning session, participants segmented the animations to identify event boundaries. These data were used to estimate the neural response to event boundaries in these animations.

Finally, each participant completed functional localizer scans to identify MT+ and pSTS. This allowed both the movement feature analyses and event boundary analyses to
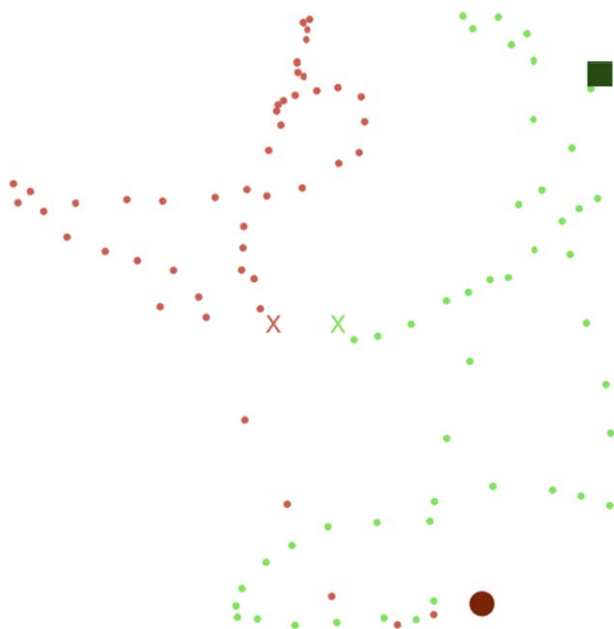
Fig. 1 – Illustration of the movements of two objects, a green square and an orange circle, during the first 5 s of stimulus movie 1. The small dots depict the path of each of the objects, sampled at 10 frames/s, and x's mark the objects' initial positions. (The black frame bordering the canvas was not visible to the participants.)

characterize effects unique to individually defined functional processing areas.

2. Results

2.1. Relationship between movement features and brain activity

To examine the relationship between movement features and brain activity during the viewing of the animations, we selected seven movement features that had previously been found to be associated with perceptual event segmentation in similar stimuli (Zacks, 2004). These were

* The speed of the circle.
* The speed of the square.
* The magnitude of acceleration (first derivative of speed) of the circle.
* The magnitude of acceleration of the square.
* The distance between the objects.
* The relative speed of the objects (first derivative of distance).
* The relative acceleration of the objects (second derivative of distance).

The movement features were averaged over each scanning acquisition frame for each animation, convolved with a model hemodynamic response function (Boynton et al., 1996), and entered into a set of general linear models to predict blood oxygen level dependent (BOLD) activity in each brain voxel for

each participant. During the fixation intervals that preceded and followed the movies during scanning, the movement features' values were zero. The models also included terms to code scan-to-scan baseline changes and linear trends over the course of each scan.

2.1.1. Region-based analysis

We first conducted a focused analysis to characterize the effect of movement features on activity in MT+ and pSTS. We computed linear contrasts for each of the movement features for each brain voxel in each participant. These were then averaged across voxels within each of the four regions of interest (ROIs) for each participant, using that individual's MT+ and pSTS regions. To test whether each movement feature was related to event segmentation, the resulting mean contrast magnitudes were subjected to one-sample t tests with participants as a random effect. To test whether the relationship between movement and brain activity differed between the two interpretation groups (random, intentional), between-groups t tests with interpretation as the independent measure were also conducted. Given the large number of tests involved, a Bonferroni correction was used to control the overall false positive rate for each region across the seven movement features. As can be seen in Fig. 2, the speed of the objects was significantly related to activity in MT+, but not to activity in pSTS. This relationship was slightly stronger for the square than for the circle in both MT+ and pSTS in both regions, but none of these differences approached statistical significance [largest $t(19) = 0.62$, $P = 0.54$]. Therefore, for further analysis we created a composite speed metric by summing the two objects' speeds.

The pattern shown in Fig. 2 suggests that the relationship between object speed and brain activity is stronger in MT+ than in pSTS. This difference was statistically significant in both hemispheres [left: $t(19) = 4.42$, $P < 0.001$; right: $t(19) = 2.42$, $P = 0.03$]. In MT+, the relationship between object speed and brain activity was significantly stronger in the left hemisphere than in the right [$t(19) = 3.12$, $P = 0.006$]. In pSTS, there was no significant difference between the hemispheres [$t(19) = 0.89$, $P = 0.38$].

None of the other movement features had relationships with MT+ or pSTS activity that approached statistical significance. There were no cases in which the relationship between a movement feature and regional brain activity differed significantly across the two groups.

To summarize, this analysis revealed a selective relationship between movement features and brain activity in MT+ and pSTS. The only movement feature that was related to brain activity in these regions was speed. The objects' speeds were significantly related to activity in MT+. This was stronger in the left hemisphere, and stronger than for pSTS.

2.1.2. Whole brain analysis

We also conducted exploratory whole-brain analyses to measure the relationship between movement features and brain activity without restricting the analysis to MT+ and pSTS. To identify regions whose activity was correlated with each movement feature, the contrast magnitude maps for each feature were submitted to one-sample t tests with participant as a random effect. To test for group differences,
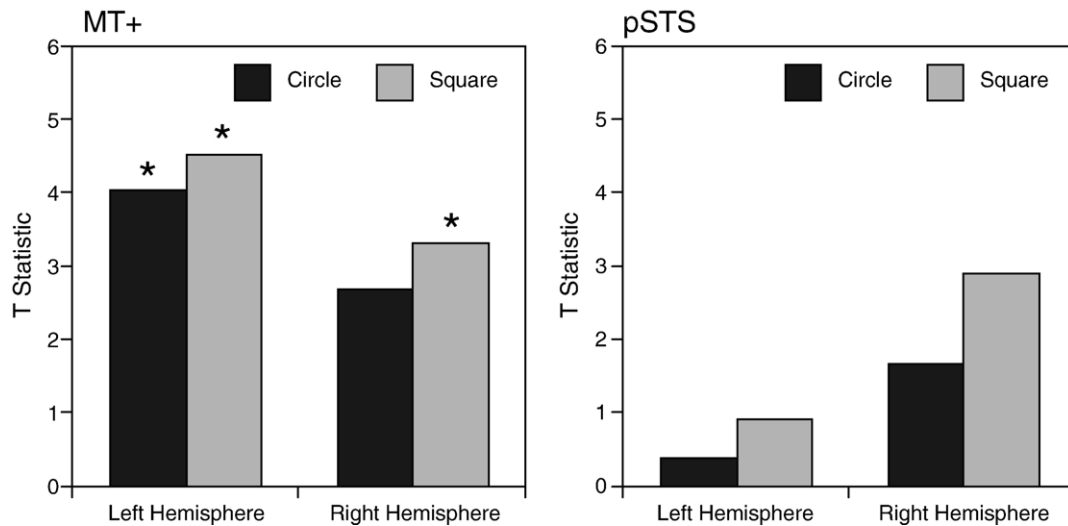
Fig. 2 – Strength of relationship between object speed and activity in MT+ and pSTS. Asterisks denote statistically significant relationships ($P < 0.05$, corrected for multiple comparisons).

the contrast magnitude maps were submitted to between-groups t tests. The resulting t statistics were converted to z statistics and corrected for multiple comparison to control the overall map-wise error rate at $P = 0.05$, based on Monte Carlo simulation (McAvoy et al., 2001). A cluster size of 5 and a z threshold of 4.0 were used.

The regions that showed a reliable overall relationship between movement features and brain activity are shown in Fig. 3. Three regions whose activity was reliably related to the speed of the square were identified. All three were in extrastriate visual cortex [Brodmann's area (BA) 18], two in the left hemisphere and one in the right hemisphere (Fig. 3, top row). A region-wise follow-up test was performed to test whether these regions were also sensitive to the speed of the circle. For all three regions, activity was positively related with the speed of the circle, but this was statistically significant only for the left occipital region [left temporal: $t(23) = 1.86$, $P = 0.08$; right occipital: $t(23) = 1.04$, $P = 0.31$; left occipital: $t(23) = 2.99$, $P = 0.007$]. No region showed a significantly different relationship between the speed of the circle and the speed of the square, though this approached significance in right occipital cortex [left temporal: $t(23) = 0.641$, $P = 0.53$; right occipital: $t(23) = 1.78$, $P = 0.09$; left occipital: $t(23) = 1.20$, $P = 0.24$].

Three regions whose activity was reliably related to the distance between the two objects were identified (Fig. 3, middle row). The first, in the right lateral occipital sulcus (BA 18), was more active when the objects were close together. The other two, in the inferior parietal lobule (BA 40) and precentral sulcus (BA 6), were more active when the objects were farther apart. Finally, one region in the cuneus (BA 18/19) was identified that was more active when the objects were moving toward each other, i.e., when they had a negative relative speed (Fig. 3, bottom row). We observed no regions whose activity was related to the relative acceleration of the objects.

The whole-brain analyses revealed only one region that showed a statistically significant group difference on any movement feature. This region, in the precuneus (center of mass coordinates –14, 57, 33; BA 31), had a significantly more positive relationship between the square's speed and brain activity for the random group than the intentional group. Follow-up tests were conducted to characterize this region's behavior. These tests revealed that for the random group, this region's activity was significantly positively correlated with the speed of the square [$t(11) = 4.22$, $P = 0.001$], but for the intentional group its activity was negatively correlated [$t(11) = -3.42$, $P = 0.006$]. Overall, there was no significant relationship between the square's speed and activity in this area [$t(23) = 0.96$, $P = 0.83$]. There was no significant relationship between the speed of the circle and activity in this area [$t(23) = -0.33$, $P = 0.75$], and no significant difference between the groups for this relationship [$t(23) = -0.92$, $P = 0.37$].

2.2. Event segmentation

As can be seen in Fig. 4, participants identified larger (fewer) events when asked to identify the largest units that were meaningful to them (coarse segmentation), and smaller (more) events when asked to identify the smallest units that were meaningful to them (fine segmentation). This led to a significant main effect of segmentation grain on event unit size [$F(1, 22) = 72.1$, $P < 0.001$]. There were also differences in event unit size across movies. As can be seen in the figure, one of the movies produced larger coarse-grained units than the others, leading to a statistically significant main effect of movie [$F(2, 44) = 13.4$, $P < 0.001$] and a significant movie-by-grain interaction [$F(2, 44) = 9.60$, $P < 0.001$]. None of the other effects were statistically significant (largest $F = 1.17$). In particular, there was no indication that the two interpretation groups differed in their grain of segmentation.

As Fig. 5 illustrates, the two groups identified similar event boundaries. We conducted a set of statistical tests to characterize any difference between the two groups' segmentation. First, the number of participants in each group who identified an event boundary during each scanning frame of the movies was computed, separately for coarse and fine
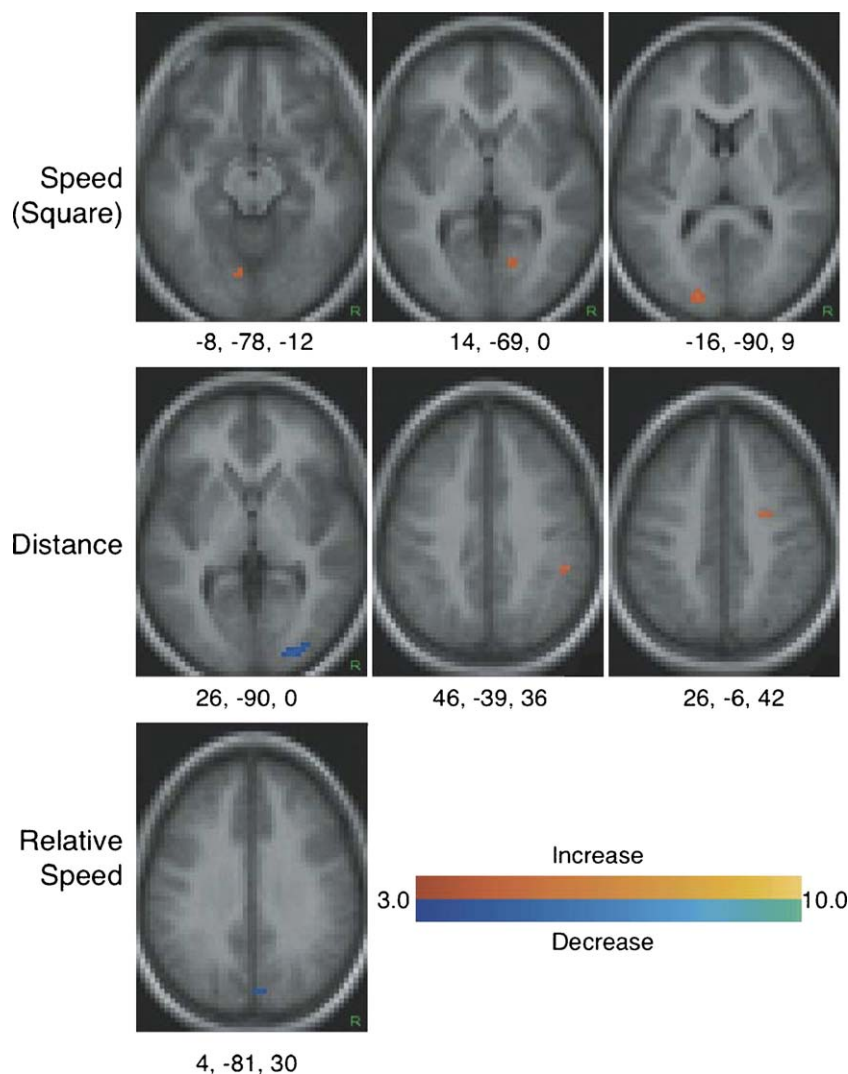
Fig. 3 – Brain regions whose activity was related to movement features. The first row shows regions whose activity was significantly related to the speed of the green square. The second row shows regions whose activity was significantly related to the distance between the objects. The third row shows regions whose activity was significantly related to the relative speed of the objects. Below each image is the Talairach coordinate of the center of mass of the region. The color scale plots the z statistic value, with positive values in red/yellow, and negative values in blue/green.

segmentation. We then computed the correlation between the two groups for the number of boundaries identified. For fine-grained segmentation, the correlation was 0.76; for coarse-grained segmentation, it was 0.56. Thus, the two groups' segmentation patterns were quite similar. To test whether participants within a group agreed better about event boundary locations than participants from different groups, we computed the mean correlation in event boundary location between all possible pairs of participants, separately for fine-grained and coarse-grained segmentation. For fine-grained segmentation, the mean correlation for pairs of participants taken from the intentional group was 0.17. For pairs taken from the random group, the mean correlation was also 0.17. For pairs chosen such that one participant came from each group, the mean correlation was 0.19. For coarse-grained segmentation, the mean correlation for pairs of participants taken from the intentional group was 0.05. For pairs taken from the random group, the mean correlation was 0.13. For

pairs chosen such that one participant came from each group, the mean correlation was 0.08. To test whether these levels of agreement were greater than would be expected by chance, we constructed bootstrap confidence intervals (Mooney and Duval, 1993) for each mean correlation. These indicated that agreement across participants was greater than expected by chance ($P < 0.05$) for all correlations except for those among the intentional participants segmenting at a coarse grain. To test whether pairs of participants from the same group chose segment boundaries that were more similar than pairs of participants from different groups, we constructed bootstrap confidence intervals for the difference between correlations taken from pairs from the same group and pairs from different groups. For neither fine nor coarse segmentation were these differences statistically significant, indicating that agreement across groups was not statistically different from agreement within each group. To summarize the behavioral segmentation results: Participants identified larger or smaller event
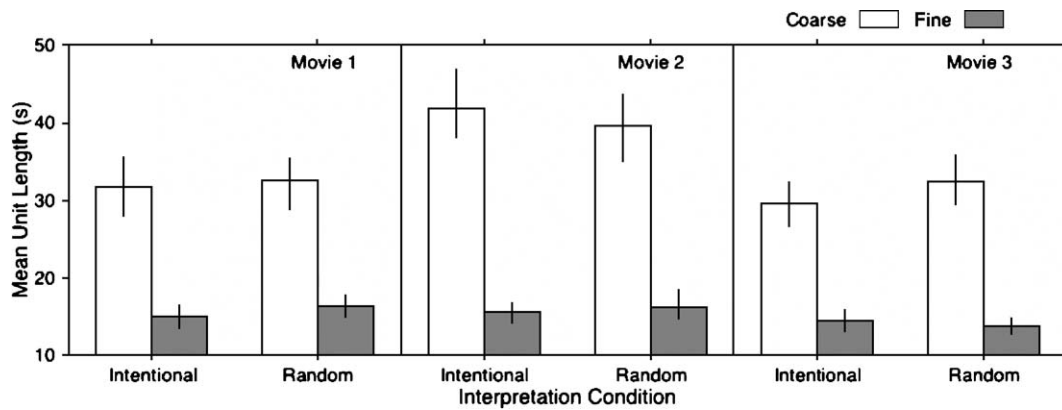
Fig. 4 – Event unit length as a function of grain of segmentation and stimulus movie. Error bars indicate standard errors.

units as instructed, and the intentional and random groups identified similar boundaries.

2.3. Neural correlates of event boundaries

2.3.1. Region-based analysis
To characterize the relationship between event boundaries and brain activity in MT+ and pSTS, we first coded each scanner acquisition frame during each movie as either an event boundary or a non-boundary. For each frame, the number of participants who identified an event boundary during that frame was counted. This was done separately for fine-grained and coarse-grained segmentation. Fine-grained and coarse-grained event boundaries were designated as those frames whose counts were more than two standard deviations above the mean.[1] Because the two interpretation groups' segmentation patterns did not differ significantly, segmentation data were collapsed across them. For each participant, a general linear model of the BOLD signal was calculated for each brain voxel. A set of 16 delta function predictor variables were included to estimate the time course of brain activity over the 34.56 s window surrounding the event boundary, without assuming any response shape

(Liu and Frank, 2004). Predictor variables were also included to code scan-to-scan baseline changes, linear drift in the MRI signal, and low-frequency changes over the course of the movies (using 60 s boxcar functions covering the duration of each movie). Time course estimates were averaged over each participant's MT+ and pSTS regions, and were then entered into mixed effect ANOVAs, with timepoint (1–16) and segmentation grain (fine, coarse) as repeated measures and interpretation (random, intentional) as a between-participants variable. Participant was the random factor.

In this analysis, an overall evoked response at event boundaries is identified as a main effect of timepoint. We observed statistically significant effects of timepoint in right MT+ [$F(15, 270) = 2.06$, $P = 0.01$] and pSTS [$F(15, 270) = 2.52$, $P = 0.001$]. The main effect of timepoint in left pSTS was marginally significant [$F(15, 270) = 1.70$, $P = 0.05$], and in left MT+ it did not approach significance [$F(15, 270) = 1.07$, $P = 0.39$]. These overall effects were increases in BOLD signal following event boundaries. However, the main effects that were observed were qualified by timepoint-by-grain interactions. As can be seen in Fig. 6, the interactions were characterized by two features. First, the late increases were present only for coarse-grained event boundaries. Second, in the interval before fine-grained event boundaries there was generally an increase, whereas before coarse-grained boundaries there was generally a decrease. This gave rise to statistically significant timepoint-by-grain interactions in bilateral MT+ and right pSTS [smallest $F(15, 270) = 2.70$, $P < 0.001$]; this interaction approached statistical significance in left pSTS [$F(15, 270) = 1.69$, $P = 0.05$]. None of the effects involving interpretation approached statistical significance in any of the regions. To follow up the timepoint-by-grain interactions, two-way ANOVAs were performed for each grain separately for each region, with timepoint and group as the independent variables. In all four regions, statistically significant main effects of time were observed during coarse segmentation [smallest $F(15, 270) = 2.46$, $P = 0.002$]. Statistically significant main effects of time were found during fine segmentation in left MT+ [$F(15, 270) = 2.33$, $P = 0.004$] and right pSTS [$F(15, 270) = 2.86$, $P < 0.001$]. This effect was marginally significant in right MT+ [$F(15, 270) = 1.62$, $P = 0.07$], and did not approach statistical significance in left pSTS

---

[1] The question whether to use normative event boundaries or individual event boundaries for this type of analysis is an important methodological issue. On the one hand, individuals differ significantly in where they place event boundaries (Speer et al., 2003), so information may be lost in cumulating across individuals. This argues for using individual-level event boundaries. On the other hand, participants' individual event boundaries may be less reliable than group-level boundaries. For example, on a given viewing an observer may perceive an event boundary but simply forget to press a button to mark it. This militates for using group-level event boundaries. Previous neuroimaging research of segmentation of everyday activities has used participants' own event boundaries, rather than cumulating across a group (Hanson et al., 2001; Speer et al., 2003; Zacks et al., 2001a). However, patterns of segmentation for these simple animations appear to be less stable than for movies of everyday activities, rendering such an approach less desirable. Preliminary analyses comparing individual to group-level event boundaries indicated that in this case using group-level event boundaries led to more stable estimates.
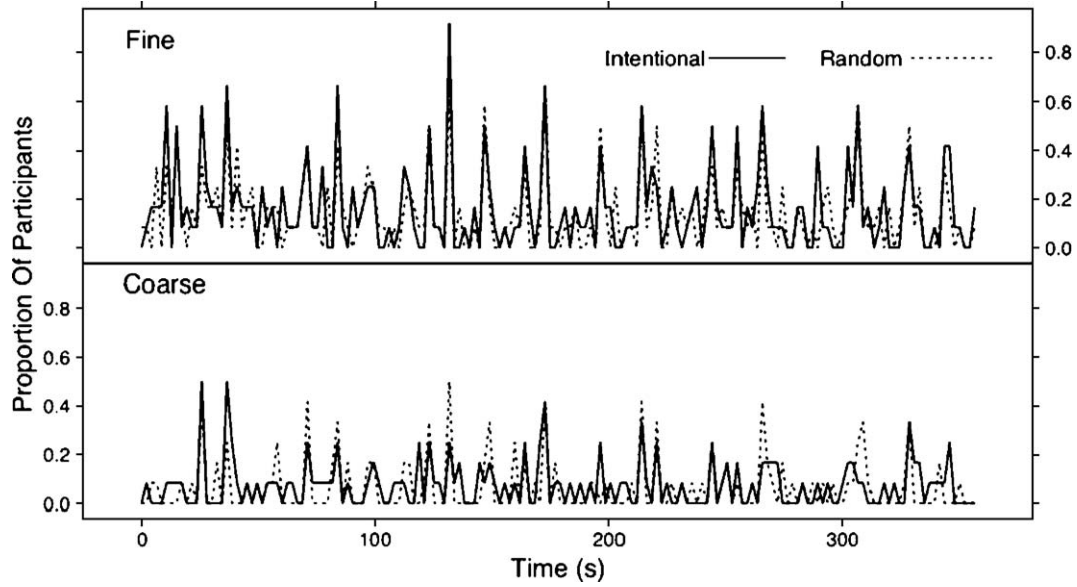
Fig. 5 – Participants agreed on the locations of event boundaries. Each line indicates the proportion of participants who identified an event boundary during each 2.16 s scanning frame of one of the movies (Hanson and Hanson, 1996). The solid lines represent the intentional group; the dashed lines represent the random group. Data from fine-grained segmentation are plotted in the top panel, and data from coarse-grained segmentation are plotted in the bottom panel.

[$F(15, 270) = 0.43$, $P = 0.52$]. In sum, all four regions showed robust increases in BOLD signal after the perception of coarse-grained event boundaries. These regions also showed increases in anticipation of fine-grained event boundaries, but these were not always statistically significant.

2.3.2. Whole brain analysis

To explore the relationship between event boundaries and evoked brain activity throughout the brain, the timecourses for individual brain voxels were analyzed using ANOVAs of the same form as those for region-based analyses. The resulting F

statistic maps were converted to z statistics and corrected for multiple comparisons based on Monte Carlo simulations (McAvoy et al., 2001). We used a cluster size of 5 and a Z threshold of 4.5 to control the mapwise error rate at $P = 0.05$. Two regions with statistically significant main effects of timepoint were found, both in the right extrastriate occipital cortex (see Table 1). The locations of these regions were near the typical locations of MT+ and pSTS, but were more inferior and posterior. Evoked responses in these regions generally resembled those shown in Fig. 6. The behavior of the two regions was further characterized by averaging the
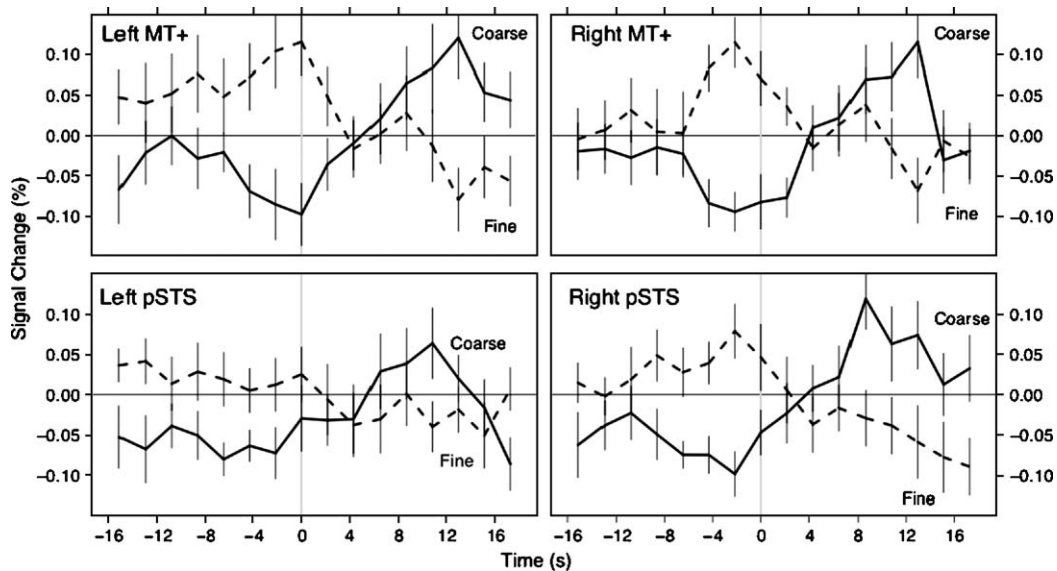

Fig. 6 – Evoked responses at event boundaries in MT+ and pSTS. The location of the event boundary is indicated by the vertical gray line in each pane.

Table 1 – Exploratory analysis of brain responses to event boundaries			
	X	Y	Z
Regions with a significant grain × time interaction			
R. middle occipital gyrus (BA 18/19)	44	−66	−3
R. middle occipital gyrus (BA 18)	34	−84	0
Regions with a significant timepoint-by-grain interaction			
L. cerebellum	−20	−78	−18
R. inf. frontal cortex (BA 11/47)	44	36	−12
R. lingual gyrus (BA 18)	10	−78	0
R. middle occipital gyrus (BA 19/37)	40	−75	3
R. middle occipital gyrus (BA 18)	26	−87	12
L. middle occipital gyrus (BA 19)	−26	−81	15
R. perisylvian cortex (BA 40/42/43)	62	−21	18
L. perisylvian cortex (BA 40/42/43)	−68	−24	18
R. temporoparietal junction (BA 39/40)	56	−51	24
R. occipitoparietal junction (BA 7/19)	20	−78	36
R. precuneus (BA 7)	4	−36	45
Regions with a significant timepoint-by-grain-by-interpretation interaction			
R. fusiform gyrus (BA 17/18)	22	−90	−15
L. caudate nucleus	−14	3	−6
L. inf. frontal cortex (BA 10)	−34	54	0
R. dorsolateral prefrontal cortex (BA 9)	−40	27	33

timecourses over voxels within each region and submitting the resulting average timecourses to ANOVAs of the same form as the original voxel-based ANOVAs. Unsurprisingly, both regions showed significant main effects of timepoint [smallest $F(15, 330) = 4.52$, $P < 0.001$]. Both regions also showed significant timepoint-by-grain interactions [smallest $F(15, 330) = 1.86$, $P = 0.03$], which were similar to those found for MT+ and pSTS. The more medial/posterior of the two regions showed a significant timepoint-by-grain-by-interpretation interaction [$F(15, 330) = 1.80$, $P = 0.03$]; however, for the other region this interaction was not significant [$F(15, 330) = 1.35$, $P = 0.17$]. For the medial/posterior region, this interaction was due to the absence of the late response in the intentional/coarse condition. To characterize this pattern, t tests were performed comparing BOLD signal in the four conditions for the peak of the coarse responses, 10.8 s after the event boundary. For the random group, the magnitude of response did not differ significantly across the grains [$t(11) = 1.73$, $P = 0.11$], whereas for the intentional group the magnitude of response was significantly greater for coarse-grained event boundaries [$t(11) = 2.62$, $P = 0.02$]. None of the other main effects or interactions approached statistical significance in either region [largest $F = 1.26$, $P = 0.27$].

The exploratory whole brain analyses also revealed a number of regions in posterior temporal and occipital cortex that had statistically significant timepoint-by-grain interactions. These are listed in Table 1. Evoked responses in these regions also generally resembled those shown in Fig. 6. These regions did not show an overall main effect of time due to the lower responses for fine-grained event boundaries. There were also two regions in left frontal cortex, listed in Table 2, that showed significant interpretation-by-grain-by-time interaction. The patterns of these responses were not consistent across regions or easily interpretable.

2.4. MT+ and pSTS localization

As can be seen in Table 2 and Fig. 7, the locations of the MT+ and pSTS regions identified here corresponded reasonably well with those reported previously (Dumoulin et al., 2000; Grossman et al., 2000). The volumes of the regions identified varied substantially across participants, likely reflecting individual differences in signal-to-noise ratio in addition to anatomic variability. As indicated in Fig. 7, participants' MT+ and pSTS regions were adjacent but largely nonoverlapping, replicating previous results (Grossman et al., 2000). In the left hemisphere, the mean overlap was 101 mm³ (SD 182.6); in the right hemisphere the mean overlap was 45 mm³ (SD 119.2). For the region-based analyses reported above, only nonoverlapping voxels were included.

3. Discussion

The results of this study support the role of MT+ in the processing of movement, and extend that role from simple trial-based designs to extended events. Activity in MT+ bilaterally increased as the speed of the objects in the movies increased, replicating previous trial-based results for this range of velocities (Chawla et al., 1998). (At higher velocities, responses in MT decrease, resulting in a U-shaped function of response as a function of speed.) This converges with studies using live-action movies (Bartels and Zeki, 2004a,b; Hasson et al., 2004) to indicate that complex ongoing events can be used to study visual processing using fMRI.

Both MT+ and pSTS were selectively activated at those moments in time that participants later reported to be coarse-grained perceptual event boundaries. This replicates previous results with naturalistic events (Speer et al., 2003; Zacks et al., 2001a), and specifically confirms the involvement of pSTS. These regions also showed transient changes in activity associated with fine-grained event boundaries, but this was less robust. An exploratory analysis revealed similar patterns for nearby regions in right extrastriate cortex.

The differences between MT+ and pSTS shed light on the processing of movement in ongoing events. Whereas activity in MT+ was strongly and significantly correlated with object

Table 2 – Descriptive statistics for MT+ and pSTS functional localizers					
Region	N Identified (of 24)	Mean volume (mm³)	Mean location		
			X	Y	Z
Left MT+	20	800 (940)	−40.7 (11.2)	−71.0 (6.7)	6.6 (7.1)
Right MT+	20	699 (720)	42.0 (10.3)	−65.7 (5.3)	7.7 (4.9)
Left pSTS	20	1365 (1766)	−41.4 (7.9)	−72.0 (9.4)	7.7 (10.4)
Right pSTS	20	2209 (1924)	46.5 (4.1)	−63.8 (5.7)	6.0 (6.2)

Note. Numbers in parentheses are standard deviations across participants.
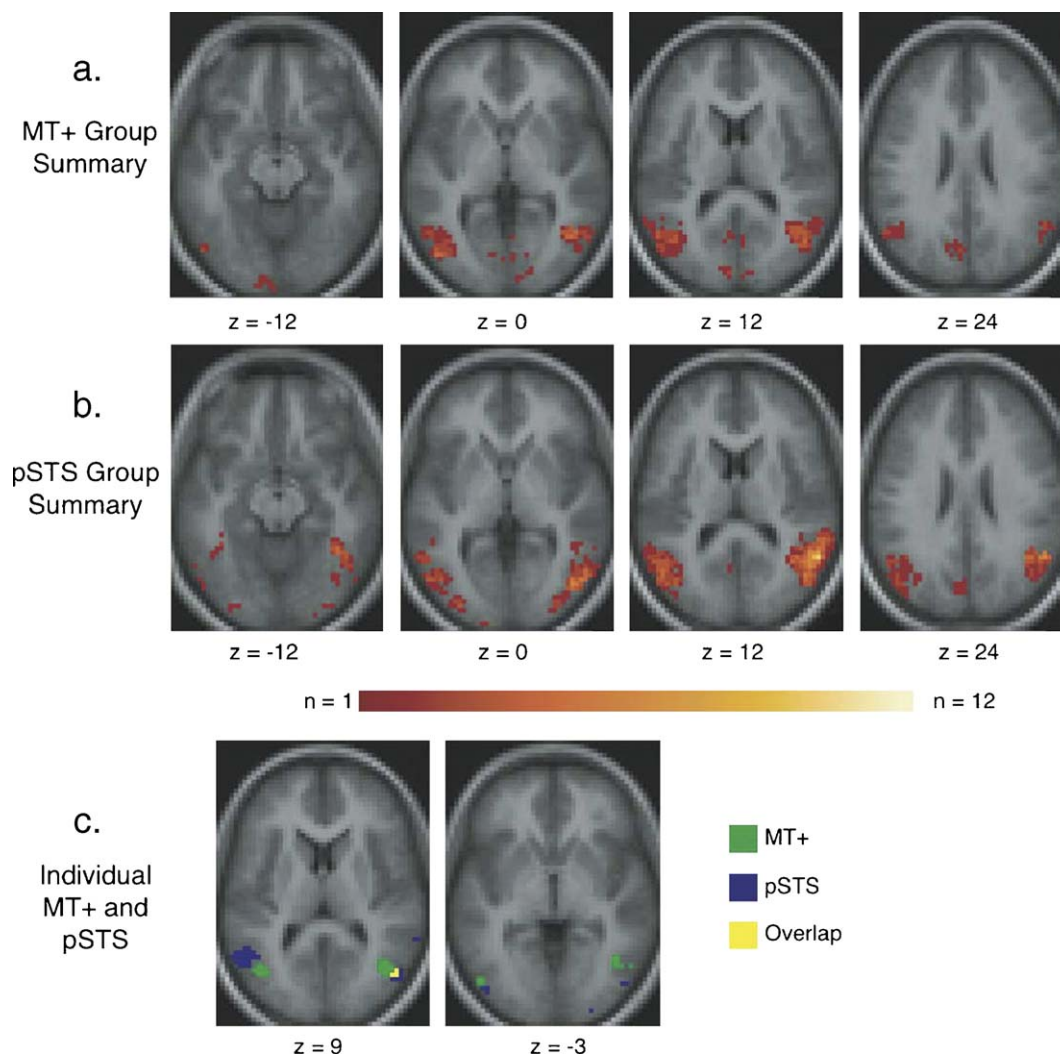
Fig. 7 – Identification of MT+ and pSTS in individual participants. (a) Number of participants for whom each brain voxel was included in MT+. (b) Number of participants for whom each brain voxel was included in pSTS. (c) Examples from two participants illustrating typical patterns of small overlap (left pane) or adjacency (right pane) between MT+ and pSTS. For all images, functional data are overlaid on the average anatomic template used for normalization.

speed, this relationship was weaker in nearby pSTS. However, both MT+ and pSTS showed robust transient increases immediately after event boundaries. Perceptually, observers tend to identify event boundaries at points at which objects are moving quickly (Zacks, 2004). Thus, it is possible that the relationship between MT+ activity and event boundaries is causally driven by object speed. However, this explanation is less likely for pSTS, given the weaker relationship between pSTS activity and object speed. Alternatively, activity in pSTS at event boundaries may be driven by top-down modulation in response to the detection of an event boundary. Further research is needed to test this hypothesis.

The differences between MT+ and pSTS also provide some support for the view that pSTS is specialized for biological motion processing (Beauchamp et al., 2002; Grossman et al., 2000). However, it is important to consider three caveats. First, activity in pSTS was positively correlated with object speed, though not significantly so. This null result should not be interpreted as establishing that there is no relationship

between object speed and pSTS activity. Second, differences between MT+ and pSTS in the relationship between object speed and activity reached significance only for one object, and only in the left hemisphere, though similar patterns were observed for both objects and both hemispheres (see Section 2.1.1). Finally, this experiment did not include stimuli depicting biological motion, and therefore the results constitute only a single dissociation. Converging evidence testing the dissociability of MT+ and pSTS during the perception of ongoing movement could be provided by studies comparing the relationship between pSTS activity and movement features during displays depicting biological motion and displays depicting nonbiological motion. If these relationships are stronger for biological motion than nonbiological motion in pSTS, this would complement the single dissociation observed here, suggesting distinct roles for the two regions in visual motion processing.

In bilateral MT+ and in right pSTS, evoked responses to fine and coarse boundaries differed. Evoked activity after an event

boundary was generally greater following coarse than fine boundaries. This pattern replicates previous findings for naturalistic movies (Speer et al., 2003; Zacks et al., 2001a). Activity prior to the event boundary showed the opposite pattern—an effect not previously observed. The exploratory whole-brain analysis of responses at event boundaries indicated that a number of regions in posterior cortex showed this pattern. One possibility is that pre-boundary increases, which were larger for fine-grained units, reflect stimulus-driven processing that feeds into the detection of event boundaries. If fine units are more determined by stimulus-driven processing (Zacks, 2004), such responses would be expected to be larger for fine units. Post-boundary increases, on the other hand, may reflect modulation of ongoing processing in response to a significant conceptual change; these would be expected to be larger for coarse-grained boundaries.

Exploratory whole-brain analyses revealed several regions whose activity was correlated with movement features. These should be interpreted with some caution regarding generalization, because it is possible that some of the observed effects reflect particulars of the small number of stimuli used. That said, two observations seem warranted. First, throughout extrastriate cortex, we observed regions whose activity was positively correlated with object speed. This is consistent with the observation that many visually responsive neurons fire most vigorously when the stimulus in their receptive field is moving quickly or is rapidly changing in contrast (Sekuler et al., 2002). Second, changes in object distance led to two distinct effects. One set of regions in posterior frontal and parietal cortex responded more when the two objects were far apart. This could reflect switching of attention and/or gaze location between the objects (Corbetta, 1998). However, a region in the right lateral occipital sulcus was more highly active when the objects were close together. This region was near the border of areas V4v and LO, the latter of which is thought to be involved in the initial stages of object processing (Tootell et al., 1996). Neuroimaging evidence suggests that when multiple objects are present, suppressive mechanisms reduce the response in V4 (and TEO), possibly implementing a mechanism of spatial selective attention (Kastner et al., 1998). One possibility consistent with that result and with the present findings is that when the two objects are close to each other, there is relatively more opportunity to analyze the *configuration* formed by the two objects as an object in its own right. Finally, one region in the right precuneus showed a group difference in how it responded to the speed of the animated square: It increased with increasing speed for the participants who treated the activity as intentional, but decreased with increasing speed for participants who treated the activity as random. This group difference was specific to one of the two objects; therefore, it is particularly likely that this reflects some stimulus-specific features of the animations used here.

Other research has studied the perception of animacy and intentionality in simple animations such as these by changing the characteristics of the stimuli (Blakemore et al., 2003; Castelli et al., 2000; Martin and Weisberg, 2003). Two studies compared animations in which objects moved pseudorandomly according to simple physical rules with animations in which objects moved as if they were characters engaged in goal-directed activity (Blakemore et al., 2003; Castelli et al., 2000). A third study used a similar animate motion condition, but contrasted it with a condition depicting organized mechanical causality, as one might find in a factory or on a billiards table, rather than random movement (Martin and Weisberg, 2003). In all three studies, activity in ventral temporal cortex, in or near the fusiform gyrus, was greater for the animations depicting intentional agents. Regions in posterior superior temporal cortex showed a similar pattern in two of the three studies (Castelli et al., 2000; Martin and Weisberg, 2003). In the current experiment, manipulating participants' instructions about whether the movies were intentional did not have substantive effects on the neural response to event boundaries, or on the relationship between object movement and neural activity. One possibility is that the cover story manipulation was too weak and that manipulating the actual stimuli shown would produce robust group effects on these variables. Consistent with this possibility, stimulus manipulations have been found to produce reliable effects on the relationship between movement features and event segmentation (Zacks, 2004). Another possibility is that the effects observed in previous studies reflect a global up-regulation of activity in these areas, but that global change does not affect transient changes of the type studied here.

Given that activity in MT+ was related to object speed, and also to the presence of a perceptual event boundary, it is natural to ask two questions about the causal relationship between activity in MT+ and event perception. First, is activity in MT+ causally responsible for the detection of event boundaries? Second, is activity in MT+ modulated by the detection of event boundaries? Answering both of these questions requires varying the relationship between movement features—particularly object speed—and event segmentation. This will be an important task for future research.

In sum, the results presented here establish two key findings. First, specialized brain regions track movement features in ongoing activity that could form a basis for event segmentation. To our knowledge this is the first evidence for velocity sensitivity in MT+ during ongoing complex movement. Second, brain regions involved in processing motion in general and biological motion in particular respond selectively at the boundaries between events—even in the absence of nonrigid biological motion. Together, these findings provide the first evidence for a causal pathway such that changes in movement features lead to the perception of event boundaries.

4. Experimental procedures

4.1. Participants

Twenty-four volunteers were recruited from the Washington University community. All were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971), reported that they were neurologically healthy, and reported no conditions contraindicating MRI scanning. They were paid $25 per hour for their participation.

4.2.   MRI acquisition and pre-processing

MRI data were acquired using a 1.5 T Siemens Vision scanner (Erlangen, Germany). For each participant we acquired a high-resolution (1 mm isotropic) whole brain structural image using a 3D magnetization prepared rapid acquisition gradient recalled echo (MP-RAGE) T1-weighted pulse sequence. Functional imaging was performed using an asymmetric spin-echo echo-planar pulse sequence optimized for BOLD contrast (Conturo et al., 1996; Ogawa et al., 1990). The sequence had an echo time of 37 ms, acquisition time per slice of 135.2 ms, and a total acquisition time (TR) of 2.16 s. Sixteen axial slices were acquired with a thickness of 8.0 mm and in-plane resolution of 3.75 mm. To facilitate stereotactic alignment, a set of T2-weighted structural images was acquired for each participant, in the same planes as the functional images, with an in-plane resolution of 0.938 mm.

Prior to analysis, the functional data were pre-processed and warped to a standard stereotactic space (Talairach and Tournoux, 1988). Timing offsets between slices were corrected using cubic spline interpolation, and slice intensity differences were removed using suitably chosen scale factors. The data were spatially smoothed with a Gaussian kernel (full width at half maximum 6.0 mm).

4.3.   Stimulus presentation and behavioral data acquisition

Stimuli were presented by an Apple Power Macintosh computer (Cupertino, CA) with PsyScope experimental presentation software (Cohen et al., 1993). During MRI scanning, an LCD projector was used to project the images onto a screen behind the scanner, where they were viewed through a mirror attached to the scanner head coil. This provided an image subtending approximately 16 degrees of horizontal vertical angle, and 12 vertical degrees. Responses were recorded with a custom-made fiber-optic button box. In the post-scanning session, stimuli were presented on a CRT monitor and responses were collected using the PsyScope button box.

4.4.   Animation viewing and segmentation tasks

Participants viewed three animations, each 360 s long, depicting the movements of a small orange circle and a small green square on a white square background. The stimulus was 480 pixels square, and subtended a visual angle of approximately 12°. The circle and square were each 20 pixels (0.5°) in diameter. The animations were similar to those used in a previous study (Zacks, 2004). They were generated by initially placing the two objects 10% above (square) or below (circle) the center of the background, and updating their position 10 times per second according to the following equations:

$$x_i = x_{i-1} + D\partial_{x,I-1} + Ar_{x,I}(1-D) \tag{1a}$$
$$y_i = y_{i-1} + D\partial_{y,I-1} + Ar_{y,I}(1-D) \tag{1b}$$

The variables x and y denote the x and y position of the object, and the subscript i denotes the current time step. The value $\partial_{x,i-1}$ is the amount the object moved in the x direction on the previous time step, i.e., $\partial_{x,i-1} = x_{i-1} - x_{i-2}$. Similarly, $\partial_{y,i-1} = y_{i-1} - y_{i-2}$. The variables $r_{x,i}$ and $r_{y,i}$ are random numbers drawn from the uniform distribution [–0.5, 0.5]. The parameters D and A control the relative contributions of previous velocity and noise to the object's movement. For this experiment, A was set to 1.0, and D was set to 0.9425, which produced movements that were relatively rapid and included sudden accelerations. A depiction of the animation is shown in Fig. 1, and the full stimuli can be downloaded from http://dcl.wustl.edu/stimuli.html.

During the *passive viewing* task (performed in the scanner), participants were asked simply to watch an animation and try to remember as much as possible. During the *active segmentation* task (performed later, outside the scanner), they were instructed to press a button on a button box whenever, in their judgment, one natural and meaningful unit ended and another began (Newtson, 1973). On a given viewing, a participant was asked to segment the activity into either the largest units that participant found meaningful (*coarse* segmentation) or the smallest units she or he found meaningful (*fine* segmentation).

Before viewing any of the animations, each participant was told one of two cover stories. Half of the participants were told that the animations were randomly generated (the *random* group). These participants were given a description of the animation generation process, including equations 1a and 1b. The other half of the participants were told that the animations were generated by the intentional actions of two people playing a video game (the *intentional* group). Participants in the intentional group were shown a game in which they and an experimenter controlled the motions of the circle and square by pressing keys on the computer keyboard.

4.5.   MT+ localizer

To localize motion-sensitive regions of visual cortex, we used a block-design paradigm based on previous reports (Beauchamp et al., 2002; Swallow et al., 2003). During separate blocks, three types of stimuli were presented: The *still* stimuli consisted of white dots placed randomly on a black background, within a circular annulus. The *high contrast motion* stimuli consisted of white dots moving in a randomly chosen direction on a black background, within a circular annulus. The *low contrast motion* stimuli were identical to the high contrast motion stimuli, except that the light dots and dark background were adjusted toward the mean luminance possible on the display, reducing the contrast to 0.06. For all conditions, 100 dots with diameter 9 pixels (0.225°) were presented, inside an annulus 440 pixels in diameter (11°). In the motion conditions, pixels moved at a velocity of 140 pixels/s (3.5°/s) in one of 11 randomly chosen directions ranging from 0° to 345° in 15° increments.

Each still picture or moving dot field was presented for 1 s, followed by a 1.16 s neutral gray inter-stimulus interval. Participants completed two 372 s fMRI scans of the MT+ localizer task. (For one participant, a scanner malfunction led to the omission of one scan.) Each scan consisted of 4 initial frames to allow transient magnetic effects to dissipate, followed by 168 stimuli. The 3 stimulus conditions were grouped into blocks of 8 trials, which repeated in a fixed order. For one scan, the order was high contrast motion, still, low contrast motion; for the other scan, it was low contrast motion, still, high contrast motion. The order of administration of these two scans was counterbalanced across participants. To maximize attention to the stimuli, participants performed a one-back memory task. For each stimulus, they were asked to press a button with their right hand if it was identical to the previous stimulus, and to press a button with their left hand if it was not. Whether the current stimulus did match the previous one was randomly determined, except for the first trial of each block, which was always nonmatching. No feedback was given.

MT+ is defined as a region in lateral posterior cortex that responds selectively to movement, independent of stimulus contrast (Huk et al., 2002). It can be distinguished from movement-sensitive regions in frontal and parietal cortex, and from posterior regions that respond to moving stimuli but also to static stimuli with high contrast (e.g., V1). To identify regions satisfying these criteria, voxelwise t tests were performed for each participant, comparing the two motion conditions to the still condition, and comparing the high contrast motion condition to the low contrast motion condition. MT+ was operationalized as voxels that satisfied the following three criteria: (1) greater response in the motion conditions than the still condition ($z > 7.0$); (2) response in the high-contrast motion condition not greater than in the low

contrast motion condition ($z \leq 2.7$); and (3) within a specified anatomic region. The anatomic restriction was performed to exclude motion-selective regions in superior parietal and frontal cortex. A large region was chosen, and was used for both the MT+ localizer and the pSTS localizer described in the following section, so that the localization was truly functionally defined. The anatomic region was defined on a standard atlas, and included portions of occipital cortex, posterior temporal cortex and the inferior parietal lobule, excluding the lingual and fusiform gyri. The region's anterior border was created by dropping a line from the posterior end of the Sylvian fissure through a descending branch of the superior temporal sulcus and down to the most ventral point of the inferior temporal gyrus dorsal. The region's dorsal border was approximately parallel to the Sylvian fissure and extended posteriorly from the most dorsal point on the region's anterior border.

Behavioral responses collected during scanning indicate that the one-back memory task was demanding but within participants' abilities. All participants performed significantly better than chance (as assessed by binomial tests, $P < 0.05$). Mean accuracy across participants in the still condition was 90.5% (SD 6.7%), in the high contrast motion condition was 91.2% (SD 8.5%), and in the low contrast motion condition was 90.8% (SD 9.0%) (failures to respond, which occurred infrequently, were counted as errors).

4.6. pSTS localizer

To localize regions of visual cortex selectively activated by biological motion, we again used a block-design paradigm based on previous reports (Beauchamp et al., 2002, 2003; Grossman and Blake, 2001; Grossman et al., 2000). This paradigm contrasted point-light *human motion* displays, in which 12 points on a person were tracked as they moved, and *scrambled human motion* displays in which the initial positions of the points were randomized, but the points' trajectories were preserved. All stimuli were rendered using square black dots 3 pixels in diameter (0.074°) on a neutral gray background. The stimuli were 252 pixels wide by 440 pixels tall (6.3° by 11°). Each animation lasted 1 s, followed by a 1.16 s neutral gray inter-stimulus interval. Participants performed the same one-back memory task as for the MT+ localizer.

Participants completed two 354-s scans of the pSTS localizer task. Each scan began with 4 frames to allow transient magnetic effects to dissipate, followed by 160 trials of the one-back memory task. The two stimulus conditions were grouped into blocks of eight trials each. The order of block alternation (human motion then scrambled human motion, or vice versa) was varied across the two scans, and the order of the two scans was counter-balanced across participants.

pSTS is defined to be a region in lateral posterior cortex that responds more to biological motion than to nonbiological motion with equivalent velocity, spatial distribution, etc. To identify pSTS, voxelwise t tests were performed separately for each participant to identify voxels whose response in the human motion condition was greater than in the scrambled human motion condition ($z > 5.5$). The same anatomic criteria as were used for the MT+ localizer were also used to restrict the search space for pSTS. Voxels that met the functional and anatomic criteria were included in a given participant's pSTS region.

Mean accuracy in the one-back task across participants was 92.5% (SD 6.1%) in the human motion condition, and 91.6% (SD 6.7%) in the scrambled motion condition. (Failures to respond, which occurred infrequently, were counted as errors.)

4.7. Procedure

Before scanning, participants were trained outside the scanner room on the tasks for the MT+ and pSTS localizers. They were also told that they would be watching animations in the scanner, told of the cover story corresponding to their experimental condition (random or intentional), and watched a 1-min example animation.

The scanning sessions began with a series of anatomical scans, followed by the three animation scans, the two MT+ localizer scans, and the two pSTS localizer scans. The order of presentation of the movies was counterbalanced across participants. Importantly, participants were naïve with regard to the segmentation task until after the scanning session; therefore, they could not have been covertly performing the segmentation task during scanning.

After the scanning session, participants were trained on the segmentation task outside the scanner, and segmented each of the three animations twice: once to mark fine-grained units and once to mark coarse-grained units. To reduce individual differences in the temporal grain of coarse and fine segmentation, participants were trained via a feedback procedure to segment at rates similar to those observed in previous studies of segmentation of naturalistic events (Zacks et al., 2001b): 4–5.5 fine-unit breakpoints per minute, and 1–2.5 coarse-unit breakpoints per minute. Each participant was given the opportunity to segment the 120 s practice movie and their breakpoint count was compared to these ranges. If the number of breakpoints produced was outside the appropriate range, the participant was given feedback and asked to segment the practice movie again. The order of presentation of the three movies was the same as in the scanner. Whether coarse or fine segmentation was performed first was counterbalanced across participants. Following the segmentation tasks, participants were fully debriefed and excused.

REFERENCES

Baldwin, D.A., Baird, J.A., 2001. Discerning intentions in dynamic human action. Trends Cogn. Sci. 5, 171–178.

Bartels, A., Zeki, S., 2004a. The chronoarchitecture of the human brain—Natural viewing conditions reveal a time-based anatomy of the brain. NeuroImage 22, 419–433.

Bartels, A., Zeki, S., 2004b. Functional brain mapping during free viewing of natural scenes. Hum. Brain Mapp. 21, 75–85.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34, 149–159.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2003. fMRI responses to video and point-light displays of moving humans and manipulable objects. J. Cogn. Neurosci. 15, 991–1001.

Blakemore, S.-J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Decety, J., 2003. The detection of contingency and animacy in the human brain. Cereb. Cortex 13, 837–844.

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

Castelli, F., Happe, F., Frith, U., Frith, C., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. NeuroImage 12, 314–325.

Chawla, D., Phillips, J., Buechel, C., Edwards, R., Friston, K.J., 1998. Speed-dependent motion-sensitive responses in V5: an fMRI study. NeuroImage 7, 86–96.

Cohen, C.E., Ebbesen, E.B., 1979. Observational goals and schema activation: a theoretical framework for behavior perception. J. Exp. Soc. Psychol. 15, 305–329.

Cohen, J.D., MacWhinney, B., Flatt, M., Provost, J., 1993. PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. Behav. Res. Methods Instrum. Comput. 25, 257–271.

Conturo, T.E., McKinstry, R.C., Akbudak, E., Snyder, A.Z., Yang, T.Z., Raichle, M.E., 1996. Sensitivity optimization and experimental design in functional magnetic resonance imaging. Abstr.-Soc. Neurosci. 22, 7.

Corbetta, M., 1998. Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? Proc. Natl. Acad. Sci. U. S. A. 95, 831–838.

Dumoulin, S.O., Bittar, R.G., Kabani, N.J., Baker Jr., C.L., Le Goualher, G., Bruce Pike, G., Evans, A.C., 2000. A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb. Cortex 10, 454–463.

Graziano, W.G., Moore, J.S., Collins, J.E., 1988. Social cognition as segmentation of the stream of behavior. Dev. Psychol. 24, 568–573.

Grossman, E.D., Blake, R., 2001. Brain activity evoked by inverted and imagined biological motion. Vision Res. 41, 1475–1482.

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

Hanson, C., Hanson, S.J., 1996. Development of schemata during event parsing: Neisser's perceptual cycle as a recurrent connectionist network. J. Cogn. Neurosci. 8, 119–134.

Hanson, C., Hirst, W., 1989. On the representation of events: a study of orientation, recall, and recognition. J. Exp. Psychol. Gen. 118, 136–147.

Hanson, C., Hirst, W., 1991. Recognizing differences in recognition tasks: a reply to Lassiter and Slaw. J. Exp. Psychol. Gen. 120, 211–212.

Hanson, S.J., Negishi, M., Hanson, C., 2001. Connectionist neuroimaging. In: Wermter, S., Austin, J., Willshaw, D. (Eds.), Emergent Neural Computational Architectures Based on Neuroscience: Towards Neuroscience-inspired Computing. Springer, Berlin, pp. 560–577.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640.

Huk, A.C., Dougherty, R.F., Heeger, D.J., 2002. Retinotopy and functional subdivision of human areas MT and MST. J. Neurosci. 22, 205–795.

Kastner, S., De Weered, P., Desimone, R., Ungerleider, L.G., 1998. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282, 108–111.

Lassiter, G.D., Slaw, R.D., 1991. The unitization and memory of events. J. Exp. Psychol. Gen. 120, 80–82.

Liu, T.T., Frank, L.R., 2004. Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part I: theory. NeuroImage 21, 387–400.

Martin, A., Weisberg, J., 2003. Neural foundations for understanding social and mechanical concepts. Cogn. Neuropsychol. 20, 575–587.

McAvoy, M., Ollinger, J.M., Buckner, R.L., 2001. Cluster size thresholds for assessment of significant activation in fMRI. NeuroImage 13, S198.

Mooney, C.Z., Duval, R.D., 1993. Bootstrapping: a nonparametric approach to statistical inference. Sage Publications, Newbury Park, CA, vi, 73 p.

Newtson, D., 1973. Attribution and the unit of perception of ongoing behavior. J. Pers. Soc. Psychol. 28, 28–38.

Newtson, D., Engquist, G., 1976. The perceptual organization of ongoing behavior. J. Exp. Soc. Psychol. 12, 436–450.

Newtson, D., Engquist, G., Bois, J., 1977. The objective basis of behavior units. J. Pers. Soc. Psychol. 35, 847–862.

Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. U. S. A. 87, 9868–9872.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Saxe, R., Xiao, D.K., Kovacs, G., Perrett, I., Kanwisher, N., 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. Neuropsychologia 42, 1435–1446.

Sekuler, R., Watamaniuk, S.N.J., Blake, R., 2002. Perception of visual motion. In: Pashler, H., Yantis, S. (Eds.), Steven's Handbook of Experimental Psychology, vol. 1. John Wiley and Sons, New York, pp. 121–176.

Speer, N.K., Swallow, K.M., Zacks, J.M., 2003. On the role of human areas MT+ and FEF in event perception. Cogn. Affect. Behav. Neurosci. 3, 335–345.

Swallow, K.M., Braver, T.S., Snyder, A.Z., Speer, N.K., Zacks, J.M., 2003. Reliability of functional localization using fMRI. NeuroImage 20, 1561–1577.

Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain: 3-dimensional Proportional System, An Approach to Cerebral Imaging. G. Thieme, Stuttgart.

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, 3215–3230.

Tootell, R.B., Dale, A.M., Sereno, M.I., Malach, R., 1996. New images from human visual cortex. Trends Neurosci. 19, 481–489.

Wilder, D.A., 1978a. Effect or predictability on units of perception and attribution. Pers. Soc. Psychol. Bull. 4, 281–284.

Wilder, D.A., 1978b. Predictability of behaviors, goals, and unit of perception. Pers. Soc. Psychol. Bull. 4, 604–607.

Zacks, J.M., 2004. Using movement and intentions to understand simple events. Cogn. Sci. 28, 979–1008.

Zacks, J.M., Tversky, B., 2001. Event structure in perception and conception. Psychol. Bull. 127, 3–21.

Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., Buckner, R.L., Raichle, M.E., 2001a. Human brain activity time-locked to perceptual event boundaries. Nat. Neurosci. 4, 651–655.

Zacks, J.M., Tversky, B., Iyer, G., 2001b. Perceiving, remembering, and communicating structure in events. J. Exp. Psychol. Gen. 130, 29–58.