*Cognitive, Affective, & Behavioral Neuroscience*
*2003, 3 (4), 335-345*

# Activation of human motion processing areas during event perception

NICOLE K. SPEER, KHENA M. SWALLOW, and JEFFREY M. ZACKS
*Washington University, St. Louis, Missouri*

Observers are able to segment continuous everyday activity into meaningful parts. This ability may be related to processing low-level visual cues, such as changes in motion. To address this issue, the present study combined measurement of evoked responses to event boundaries with functional identification of the extrastriate motion complex (MT+) and the frontal eye field (FEF), two regions related to motion perception and eye movements. The results provided strong evidence that MT+ is activated by event boundaries: Individuals' MT+ regions showed strong responses to event boundaries, and MT+ was collocated with a lateral posterior region that responded at event boundaries. The evidence regarding the FEF was less conclusive: The FEF showed reliable but relatively reduced responses to event boundaries, but the FEF was medial and superior to a frontal area that responded at event boundaries. These results suggest that motion cues, and possibly eye movements, may play key roles in event structure perception.

When observers view an everyday activity, such as making a bed, they are able to break the activity down into a sequence of smaller activities. Although the experience of making a bed comprises a continuous stream of activity, observers perceive the experience as a sequence of discrete meaningful units of action. This ability to perceive the individual parts of events is termed *event structure perception* (Zacks & Tversky, 2001), and it is the focus of the present study.

The development of event structure perception begins at a very young age. Infants are able to individuate actions within a sequence (Wynn, 1996) and relate those actions to overarching goals (Woodward & Sommerville, 2000), and at 3 years of age children have knowledge of the temporal ordering of actions within a causal sequence (Nelson & Gruendel, 1986). By adulthood, there is evidence that the ability to perceive event structure occurs without explicit instructions to do so (Zacks, Braver, et al., 2001) and that event structure perception affects interpretation of and memory for everyday activities (Newtson & Engquist, 1976).

Although these findings indicate that people are able to segment continuous ongoing activity into discrete meaningful units, the mechanisms that drive this process are unclear. One possibility is that the processing of motion features drives event structure (Newtson, Engquist, & Bois, 1977; Zacks, Braver, et al., 2001). For example, the changes in position that occur while a person at a restaurant is eating the main course may be smaller and more repetitive than the changes that occur when the person finishes the main course and begins to order dessert. The relatively larger change in position that occurs between finishing the main course and ordering dessert may signal the presence of a boundary point between the two meaningful units of activity. To explore the role of motion in event structure perception, Newtson et al. (1977) analyzed the positions of actors in movies and looked at how changes in the actors' positions related to the perception of event boundaries. Position changes between two event boundaries were significantly greater than position changes between two points not perceived as event boundaries. Thus, degree of movement may serve as a cue for the higher level processes involved in the conscious perception of event boundaries.

If visual changes in the movements of actors or objects involved in an event are related to the perception of event boundaries, brain regions involved in the perception of motion should also be involved in the perception of event boundaries. On the basis of this hypothesis, the MT complex (MT+), which is located anterior and ventral to the lateral occipital sulcus, is an obvious candidate for involvement in the perception of event boundaries. This area is commonly active during tasks involving visual motion (Tootell et al., 1995), as well as during tasks involving implied visual motion (Kourtzi & Kanwisher, 2000), and stimulation of the homologous area in nonhuman primates influences the perceived direction of visual motion (Salzman, Britten, & Newsome, 1990).

Another possible candidate for involvement in event structure perception is the frontal eye field (FEF). The FEF is located near the superior frontal and precentral sulci (Paus, 1996; Petit & Haxby, 1999; Rosano et al., 2002). It

Portions of these results were presented at the 2003 annual meeting of the Cognitive Neuroscience Society in New York. This research was supported by a grant to J.M.Z. from the James S. McDonnell Foundation, as well as a National Science Foundation Graduate Research Fellowship to N.K.S. The authors thank Margaret Sheridan for her essential role in the collection of the data reported here. Correspondence concerning this article should be addressed to J. M. Zacks, Department of Psychology, Washington University, Box 1125, St. Louis, MO 63130 (e-mail: jzacks@artsci.wustl.edu).

 

plays a role in guiding saccadic eye movements, which orient the eyes to a visual stimulus, and smooth pursuit eye movements, which maintain the visual stimulus in the fovea whenever the stimulus or the individual is moved (Petit & Haxby, 1999; Rosano et al., 2002). In the restaurant example mentioned previously, if the character gets up to pay the bill and the character is the focus of attention, the FEF might contribute to visually tracking the character from the table to the cashier. Therefore, the FEF may be involved in the shifting of visual attention when the observer is faced with the motion changes associated with event boundaries.

In a previous neuroimaging study, Zacks, Braver, et al. (2001) asked participants to watch movies of everyday activities and later asked them to note the points at which they believed one activity ended and another began. This event segmentation procedure is known to give a reliable measure of event structure perception (Newtson & Engquist, 1976; Zacks, Tversky, & Iyer, 2001). Each participant's behavioral data were applied to their neuroimaging data, to determine whether BOLD activity changed at points the participant later identified as event boundaries. This procedure identified a network of posterior regions and a single right frontal region, which showed increased neural activity well before the explicitly identified boundary points. These increases were observed in naive viewing conditions, before the participants were aware of the boundary identification task. In addition, these regions were modulated by the hierarchical structure of the events, with larger activation for larger event boundaries. Interestingly, the peak of the posterior activity was located near the MT+, and the right frontal region was located near the FEF, but due to the anatomical variability of these regions across participants, it was not clear whether the areas observed in the previous study were in fact the MT+ and the FEF.

The roles of the MT+ and FEF in the perception of events have implications for theories of event understanding. If activity in the MT+ and FEF were correlated with perceptual event boundaries, this would suggest that detecting motion changes is important for the perception of event structure. This could arise by two mechanisms. The first is a bottom-up mechanism: Distinctive movement features could activate the MT+ and lead to eye movements or shifts of attention, activating the FEF. These effects could then feed into processes that trigger the perception of an event boundary. The second is a top-down mechanism: The perception of an event boundary could lead to transient upregulation in the MT+ and FEF via recurrent connections.

However, if activity in the MT+ and FEF were not correlated with the perception of event boundaries, this would suggest that other external cues or internal representations have primacy for the perceiving of event structure. One set of candidates is knowledge structures that represent the structure of recurring activity, such as schemas or scripts (Rumelhart, 1977; Schank & Abelson, 1977). Another possibility is representations of goals or plans (Baldwin & Baird, 1999; Barker & Wright, 1954).

In the present study, we sought to determine whether the MT+ and FEF, functionally defined for each individual, would show evoked responses at the region level in response to event boundaries. Given the design of the present study, it was also possible to characterize the reliability of the evoked responses to event boundaries and to determine whether the posterior and frontal regions previously identified as responding to event boundaries corresponded to the MT+ and FEF.

The goals of the present study were twofold. The first was to determine whether motion changes serve as important cues for defining perceptual events by looking at whether the MT+ and FEF, two regions involved in motion perception, would show evoked responses to event boundaries. The secondary goal of the study was to determine how reliable event perception is over time, in terms of behavior and brain activation. Although event perception is thought to be the result of processes that are stable over time, such as motion detection or the top-down influences of schemata, little is known about its stability over time (Newtson, 1976).

In the present study, the participants performed two sessions of an event segmentation task, as well as two tasks designed to localize the MT+ and FEF. The data from these localizer tasks were used to define functional areas for each participant (Huk & Heeger, 2000; Kourtzi, Bülthoff, Erb, & Grodd, 2002; Kourtzi & Kanwisher, 2001; Swallow, Braver, Snyder, Speer, & Zacks, 2003; Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). In this way, it was possible to address the role of the MT+ and FEF in event perception by computing time courses for each individual's MT+ and FEF regions on the basis of the behavioral and functional data from the first viewing of the movies. Because behavioral data were collected during the first viewing of the movies and again 1 year later, it was possible to assess the reliability of event structure perception over a period of more than a year.

## METHOD

### Participants

Eleven participants (ages, 20–51 years; 5 women) who had taken part in a previous event segmentation fMRI session (Session 1; Zacks, Braver, et al., 2001) volunteered to participate in a second testing session (Session 2). The delay between the two testing sessions ranged from 398 to 463 days ($M = 439$, $SD = 19.70$). The participants received $25 for each hour of their participation, and informed consent was obtained in accordance with guidelines set by the Human Studies Committee at the Washington University Medical School.

### Imaging

The imaging protocol was essentially identical to that reported previously (Session 1; Zacks, Braver, et al., 2001), with the exception that a new, slightly faster BOLD pulse sequence was used for data collection in Session 2. Scanning was performed on a 1.5 T Siemens Vision MRI scanner (Erlangen, Germany). Structural images were acquired using a sagittal MP-RAGE T1-weighted sequence with 2-mm (isotropic) resolution. Functional images were acquired using a T2*-weighted asymmetric spin-echo echo-planar sequence. Sixteen slices, 8 mm in thickness, with an in-plane reso-

lution of 3.75 × 3.75 mm, were acquired every 2.5 sec (Session 1; Zacks, Braver, et al., 2001) or every 2.16 sec (Session 2). Prior to analysis, the functional data were preprocessed and warped to a standard stereotactic space (Talairach & 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 of 6.0 mm).

Visual stimuli were presented using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993) running on an Apple Power Macintosh G4. An LCD projector was used to project stimuli onto a screen positioned at the head of the bore. The participants viewed the stimuli on the screen through a mirror attached to the head coil. A fiber-optic, light-sensitive keypress interfaced with the PsyScope button box was used to record the participants' responses during the segmentation runs.

### Behavioral Tasks and Procedure—Session 1

During the first session, the participants passively viewed movies of everyday activities and segmented them into meaningful units, all while undergoing fMRI scanning.

**Passive-viewing task**. The participants watched four movies of everyday activities. The activity in the movies portrayed making a bed, doing the dishes, assembling a saxophone, and ironing a shirt. The movies were 316, 258, 184, and 298 sec in duration, respectively, and subtended 13.5° of the visual field with a resolution of 640 × 480 pixels. In the passive-viewing task, the participants simply watched each movie passively and were instructed to learn as much about the movie as possible.

**Fine segmentation task**. In the fine segmentation task, the participants watched the same movies as those in the passive-viewing task but were asked to press a button at the points at which they believed one meaningful and natural unit of activity ended and another began. This procedure has been shown to reliably measure the perceptual units of ongoing behavior (Newtson & Engquist, 1976). The participants were instructed to identify the *smallest* units of activity that seemed natural and meaningful.

**Coarse segmentation task**. The coarse segmentation task was identical to the fine segmentation task, except that the participants were asked to identify the *largest* units that were natural and meaningful to them.

**Procedure**. After providing informed consent, the participants were made comfortable in the scanner, and a series of structural images were taken. The participants then completed 12 BOLD runs while performing the three tasks. All the participants completed the passive-viewing task first, to obtain recordings of brain activity before they had been informed of the segmentation task. They then were given instructions for performing either the fine or the coarse segmentation task and segmented a brief practice movie. Then they watched all four movies again and segmented them. The participants next received instructions for the segmentation task they had not yet performed, practiced that task, and watched the four movies a third time, performing the other segmentation task. The order of the fine and the coarse segmentation runs was counterbalanced across the 16 participants in Session 1.

### Behavioral Tasks and Procedure—Session 2

The 11 participants who returned for a second testing session performed the passive-viewing task in the scanner for each of the four movies seen during Session 1, in addition to two runs of an FEF localizer task and two runs of an MT+ localizer task. Behavioral fine-grained segmentation data for the movies from Session 1 were also collected.

**Passive viewing and fine segmentations**. The passive viewing and fine segmentation tasks were identical to those in Session 1.

**MT+ localizer task**. The details for the localizer task designs can be found in Swallow et al. (2003). During the MT+ localizer task, the participants were asked to fixate on a central fixation cross while one of three stimulus conditions was presented (see Figure 1A). In the control condition, the fixation cross was the only stimulus on the screen. In the still condition, 100 stationary dots were randomly distributed on the screen. In the motion condition, the 100 dots moved toward the edges of the screen. These conditions were presented using a block design, with the control condition starting the block and the motion and still conditions alternating with fixation. The order of motion and still blocks was counterbalanced across runs. Each run of the MT+ localizer task was 129 frames (279 sec) long, with 45 frames of fixation, 40 frames of the still condition, and 40 frames of the motion condition (as well as an additional 4 frames of fixation at the start of the run to allow the T2* signal to stabilize).

**FEF localizer task**. In the FEF localizer task, the participants were asked to keep their eyes on a cross that either jumped to a new location every second (saccade condition) or remained stationary in the center of the screen (fixate condition; see Figure 1B). These conditions were presented using a block design, with the still and the fixation conditions alternating. Each run of the FEF localizer task was 148 frames (320 sec) long, with 80 frames of the fixation condition and 64 frames of the saccade condition (as well as an additional 4 frames of fixation at the start of the run).

**Procedure**. After providing informed consent, the participants were made comfortable in the scanner, and a series of structural images was acquired. The participants then completed passive viewing of the four movies, fine segmentation of the four movies, two runs of the MT+ localizer, and two runs of the FEF localizer. The participants received instructions for the fine segmentation task before that task was begun, as in Session 1. As in Session 1, the passive-viewing task was always run before the fine segmentation task, to minimize memory effects.

### Analysis

**Regions responding to event boundaries during Session 1**. The analysis procedure for identifying regions responding to event boundaries during Session 1 was identical to that described by Zacks, Braver, et al. (2001), with the exception that only the data from the 11 participants who returned for Session 2 were analyzed. For each participant, brain activity during the Session 1 movie-viewing tasks was assessed during the 35-sec window surrounding the locations of fine and coarse segment boundaries identified during Session 1. The BOLD response to fine and coarse boundaries was estimated using the general linear model to produce two 35-sec time courses per voxel for each viewing of the movies. The data were analyzed across participants with a two-factor analysis of variance (ANOVA), using time point (frame) and grain (fine and coarse) as the two factors. Brain regions whose activity deviated from baseline during the 35-sec window of the event boundary (i.e., showed a main effect of time point) were identified by converting the $F$ statistics to $Z$ statistics and identifying clusters of 5 contiguous voxels with $Z$ statistics greater than 4.50. This procedure maintains a mapwise Type I error rate of .05 (McAvoy, Ollinger, & Buckner, 2001).

**Functionally defined FEF and MT+**. The analysis procedure for the MT+ and FEF localizer tasks followed that described by Swallow et al. (2003). To create individual MT+ and FEF regions of interest (ROIs), voxels within MT+ and FEF were identified for each individual on the basis of the localizer tasks (Session 2) and anatomical masks. Anatomical masks for right and left MT+ and FEF were created on the basis of the standard atlas to which all images were registered. The mask for the MT+ included those brain voxels posterior to the Sylvian fissure and the central sulcus that were not part of the cerebellum or deep brain structures. The mask for the FEF included those brain voxels anterior to the central sulcus and dorsal to a line connecting the anterior and the posterior commissures. The functional data from the localizer tasks were entered into the general linear model (Friston et al., 1995) in order to estimate the BOLD signal. Left and right FEF were identified by contrasts comparing the

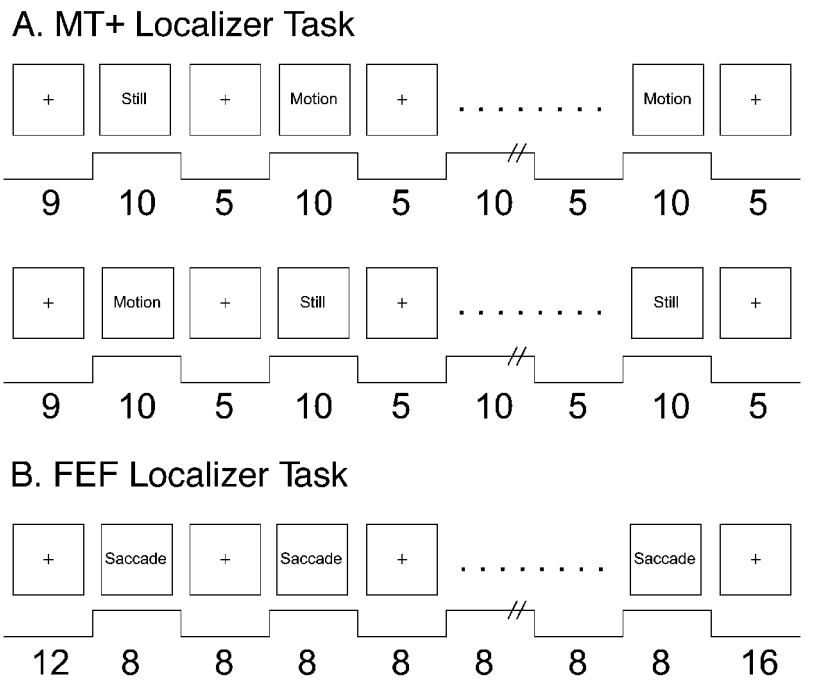


**Figure 1. Task design for the MT complex (MT+) and frontal eye field (FEF) localizer tasks. In the MT+ task, the participants saw blocks of still and motion dots alternating with blocks of fixation. The order of the two runs of the MT+ task was counterbalanced across participants. In both runs of the FEF task, the participants saw blocks of fixation (presented in red) alternating with blocks of the saccade task. The fixation stimuli for both tasks were presented in red. The number below each block refers to the number of frames in that block.**

saccade and the fixation conditions across both runs of the FEF localizer task for each voxel. The contrast values were divided by the estimated standard deviation to generate *t* statistics, which were then used to generate a *z* statistic map. This map was thresholded at 4.5 ($p < .001$), retaining only those voxels that had a *z* statistic greater than 4.5, and was masked within the anatomic mask. The left and right individual FEF ROIs were, therefore, defined as those showing significantly greater activation in the saccade condition, relative to fixation, that fell within the liberal anatomical boundaries for the FEF.

Left MT+ and right MT+ were identified by contrasts comparing the motion and the still conditions across the two runs of the MT+ localizer task. These contrasts were converted to *z* statistics and thresholded at $z = 4.5$. The still–fixation contrasts were also computed for each voxel, in order to exclude voxels that responded to nonmoving stimuli. These contrasts were converted to *z* statistics and thresholded so that voxels with a *z* statistic greater than 2.0 (corresponding to $p = .023$) were excluded from further analyses. These maps were then masked with the anatomic mask for MT+. Thus, only voxels that did not show a large response to nonmoving stimuli, did show a response to moving stimuli, and fell within the anatomical boundaries for the MT+ were included in the individual left and right MT+ ROIs.

To create the group FEF and MT+ ROIs, the individual contrasts were submitted to group-level random effects *t* tests. These *t* statistics were converted to multiple comparison corrected *z* statistic images (Monte Carlo simulation values: minimum cluster size = 12 voxels, minimum *z* statistic = 3.25). The group FEF ROI was defined as any voxel within the FEF anatomical mask with a *z* statistic greater than 3.25 ($p < .0006$) for the saccade–fixation contrast. Similarly, the group MT+ ROI was defined as any voxel within the MT+ anatomical mask with a *z* statistic greater than 3.25 ($p < .0006$) for the motion–still contrast and a *z* statistic less than 2.0 ($p > .0228$) for the still–fixation contrast.

**FEF and MT+ responses to event boundaries**. Evoked BOLD responses for all the movie-viewing tasks in Session 1 (passive viewing, coarse segmentation, and fine segmentation) were calculated for each participant on the basis of the individual segment boundaries identified during the Session 1 coarse and fine segmentation tasks, following the analysis procedure described by Zacks, Braver, et al. (2001). The BOLD response was estimated for each MT+ and FEF voxel in the passive-viewing and segmentation conditions, using a 35-sec (14-frame) window surrounding the coarse and fine segment boundaries. This produced two 35-sec time courses for each MT+ and FEF voxel for each participant in the passive-viewing condition and each of the segmentation conditions in Session 1. Time courses were averaged across voxels within each individual's MT+ and FEF regions (right MT+, left MT+, right FEF, and left FEF) to obtain two time courses for the passive-viewing and segmentation conditions within each region. The average time courses within each region were subjected to a $2 \times 14$ ANOVA, with temporal grain (fine and coarse) and time point as the two within-subjects variables.

**Reliability of evoked responses across sessions**. A final analysis was conducted to evaluate whether the regions identified as showing reliable changes at event boundaries during passive viewing in Session 1 also showed reliable changes at event boundaries during passive viewing in Session 2. The functional data from the Session 2 passive-viewing condition were used to assess the reliability of the evoked responses to event boundaries across sessions. For each

voxel in the passive-viewing condition, the BOLD response was estimated for each of the 8 regions identified during Session 1 (Zacks, Braver, et al., 2001), as well as for the individually defined FEF and MT+ regions. These responses were estimated using a 34.56-sec (16-frame) window surrounding the coarse and fine segment boundaries identified during Session 1 to produce two 34.56-sec time courses for each voxel in the 12 regions (8 regions identified during Session 1, left and right FEF and MT+) and for each participant. Time courses were averaged across voxels within the 12 regions for each individual to obtain two time courses for the passive-viewing task within each region. The average time courses within each region were subjected to 2 × 16 ANOVAs with temporal grain (fine and coarse) and time point as the two within-subjects variables.

## RESULTS

### Mean Locations of MT+ and FEF

Each participant's center of mass for the four ROIs was averaged to determine the mean location and variability for each region. For the left and right FEF and MT+ regions, the mean coordinates in the Talairach and Tournoux (1988) atlas, as well as the variability of these coordinates across participants, are shown in Table 1. The mean coordinates for the left and right FEF regions observed in the present study are slightly more medial than those that have been observed in previous studies (Paus, 1996), as are the mean coordinates for the left and right MT+ regions (Kourtzi et al., 2002). However, the majority of the mean coordinates from the present study fall within one standard deviation of the mean coordinates reported in previous studies.

### Comparing Independently Identified MT+, FEF, and Regions Responding to Event Boundaries

Figure 2A shows the centers of mass for the functionally defined individual MT+ and FEF regions used to generate time courses for the region-wise analyses, as well as the regions showing evoked responses to event boundaries during Session 1. Figure 2B shows the functionally defined group MT+ and FEF regions, as well as the regions showing evoked responses to event boundaries during Session 1. The individual and group MT+ regions fall largely within the extrastriate region identified during Session 1 (Zacks, Braver, et al., 2001): Twelve of the centers of mass for the left and right MT+ regions fell on left and right extrastriate voxels that responded to event boundaries (six for the left extrastriate region and six for the right extrastriate region), and the group MT+ regions show a good deal of overlap with the regions identified during Session 1. However, the individual and group FEF regions do not appear to be the same right precentral region identified previously: Only one center of mass for the left and right FEF regions fell on an activated voxel in the right precentral region that responded to event boundaries, and none of the voxels in the group FEF regions overlapped with the regions from Session 1.

To quantify the relationship between the individual and group MT+ and FEF regions identified in Session 2 and the regions that responded to event boundaries in Session 1, we counted the number of overlapping voxels between each MT+ and FEF region and the group regions identified in the previous study. On average, 20.25% ($SD$ = 19.78%) of the voxels in each individual left MT+ region and 19.72% ($SD$ = 18.47%) of the voxels in each individual right MT+ region overlapped with voxels that responded to event boundaries in Session 1. In contrast, only 1.51% ($SD$ = 2.84%) of the voxels in each individual right FEF region overlapped with voxels that responded to event boundaries in Session 1. (Because there was no left frontal region that responded to event boundaries in the previous study, there was no comparison for the individual left FEF regions.) The statistical significance of the overlap amounts was tested by computing the expected amount of overlap for each participant for each region on the basis of a binomial distribution assuming random MT+ and FEF voxel locations and comparing it with the observed amount of overlap. For the left and right MT+, the observed overlap was significantly more than would be expected by chance [left MT+ expected overlap = 2.92%, $t(10) = 3.69$, $p = .004$; right MT+ expected overlap = 2.58%, $t(10) = 4.04$, $p = .002$]. The difference between the number of observed and expected overlapping voxels in the right FEF was not reliable [expected overlap = 1.21%, $t(10) = 0.69$, $p = .51$].

At the group level, 55.94% of the voxels in the group right MT+ region and 52.94% of the voxels in the group left MT+ region overlapped with voxels activated at event boundaries in Session 1. There were no voxels in the group right FEF regions that overlapped with voxels activated at event boundaries in Session 1. (As for the individual overlap measures, there was no comparison for the group left FEF region.) The statistical significance of each of the MT+ overlap amounts was tested by a binomial test, assuming random MT+ voxel locations. For the left MT+ the expected percentage of overlapping voxels was 2.92%, and for the right MT+ this value was 2.58%. For both the left and the right MT+, the observed number of voxels was significantly greater than would be expected by chance ($p < .001$).

**Table 1**
**Mean Location and Variability for Each of the Four Regions of Interest**

| | Left | | | | | | Right | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *x* | | *y* | | *z* | | *x* | | *y* | | *z* | |
| Region | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| FEF | −22.9 | 2.5 | −2.4 | 4.4 | +50.0 | 3.4 | +23.3 | 4.4 | +3.6 | 8.4 | +48.6 | 6.1 |
| MT+ | −34.0 | 10.5 | −73.1 | 7.8 | +7.1 | 6.7 | +34.6 | 9.4 | −69.6 | 6.1 | +6.0 | 4.4 |

Note—FEF, frontal eye field; MT+, MT complex.

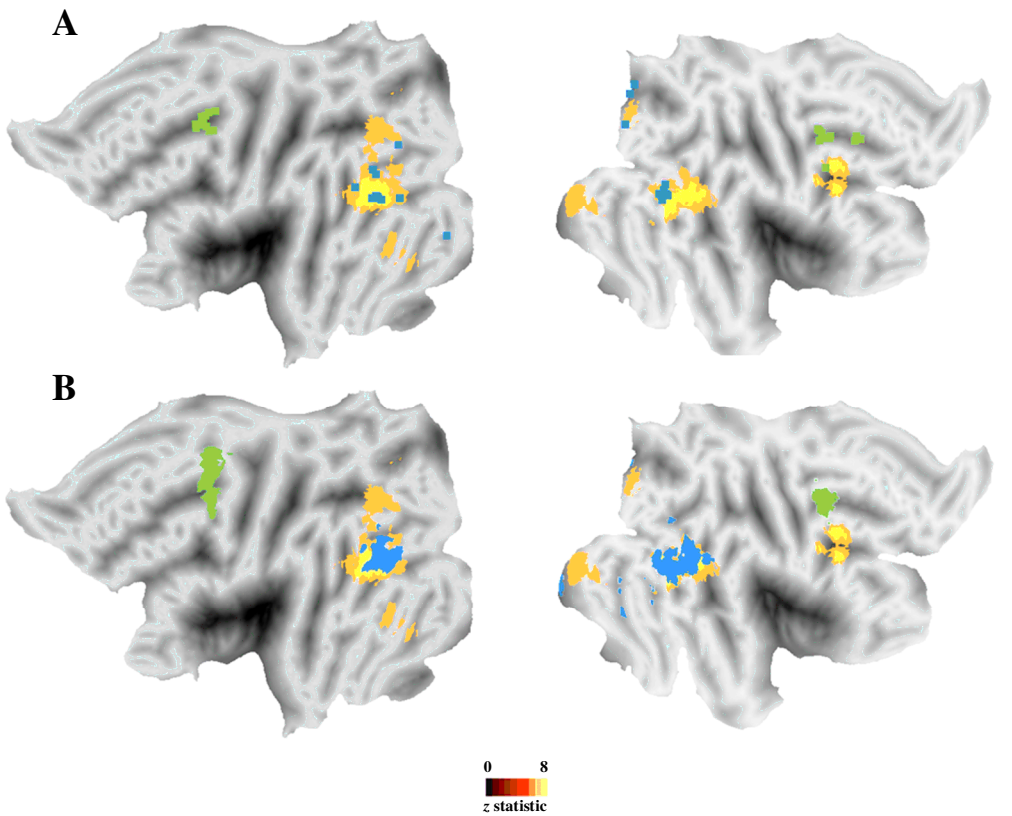


**Figure 2. (A) Centers of mass (squares) for individual MT complex (MT+, blue) and frontal eye field (FEF, green) regions of interest (ROIs) are shown with functional data from the Session 1 movie-viewing tasks. The centers of mass for the MT+ ROIs show a good deal of overlap with the left and right extrastriate regions found in Session 1, but the centers of mass for the right FEF ROIs are more medial than the previous right frontal region. (B) The group MT+ (blue) and FEF (green) ROIs are shown with the functional data from the Session 1 movie-viewing tasks. As with the centers of mass, the group MT+ region is collocated with the previous extrastriate region, but the group FEF region shows no overlap with the previous right frontal region (the right side of the figure corresponds to the right hemisphere). These figures were produced using CARET (Van Essen, 2002a, 2002b; Van Essen et al., 2001; Van Essen, Drury, Harwell, & Hanlon, 2002).**

### Region-Based Analysis of FEF and MT+

Figure 3 shows the evoked responses to the individually defined coarse and fine unit boundary points from the Session 1 passive-viewing and segmentation conditions for the functionally defined FEF and MT+ regions. All the regions showed reliable responses to boundary points, regardless of the task condition or segmentation grain, although this response was largest to coarse unit boundaries in the segmentation conditions.

**Passive-viewing runs**. All four regions showed a reliable evoked response at the individually defined boundary points during the passive-viewing condition in Session 1 [left MT+, $F(13,130) = 4.19$, $p < .001$; right MT+, $F = 7.54$, $p < .001$; left FEF, $F = 2.61$, $p < .005$; right FEF, $F = 2.56$, $p < .005$]. The right MT+ region showed a significant interaction of time point and grain [$F(13,130) = 2.08$, $p = .02$], due to a larger evoked response in the coarse-grained condition than in the fine-grained condition. None of the responses in the regions showed a main effect of grain [largest $F(1,10) = 1.35$, $p = .27$].

The evoked response in the MT+ appears to have been much larger than the evoked response in the FEF (see Figure 3). To determine the effect of region on the evoked response to event boundaries in the passive-viewing condition, we carried out a second ANOVA, using region and time point (frame) as the two independent variables. This analysis confirmed our observation that the MT+ showed a larger evoked response to event boundaries, revealing a significant effect of region [$F(1,10) = 6.45$, $p < .05$] and a significant interaction of region and time point [$F(13,130) = 4.47$, $p < .001$].

**Segmentation runs**. All four regions showed statistically significant evoked responses at the individually defined boundary points from Session 1 [left MT+, $F(13,130) = 17.33$, $p < .001$; right MT+, $F = 15.04$, $p < .001$; left FEF, $F = 16.85$, $p < .001$; right FEF, $F = 12.21$, $p < .001$]. The effect of time point interacted with the effect of grain [left MT+, $F(13,130) = 3.09$, $p < .001$; right MT+, $F = 2.70$, $p = .002$; left FEF, $F = 3.16$, $p < .001$; right FEF, $F = 2.10$, $p = .02$], with larger modulation of evoked responses at

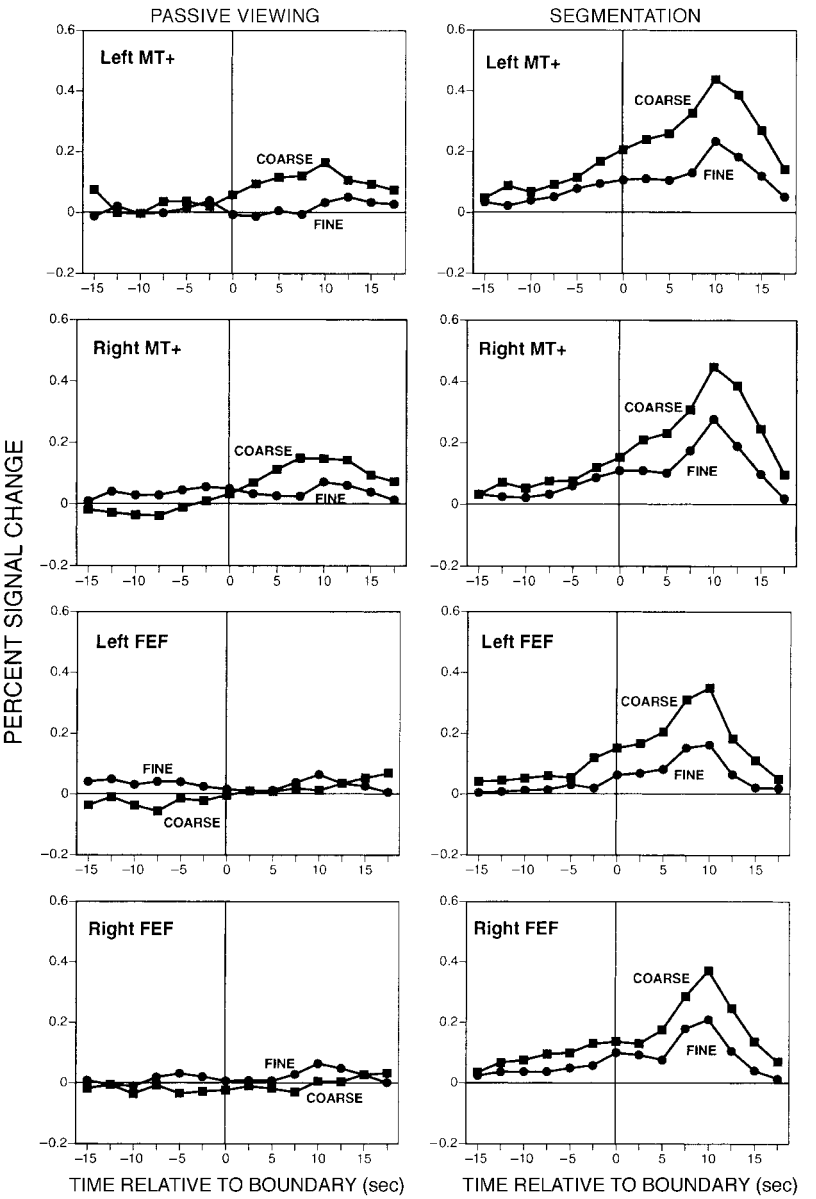


**Figure 3. Time courses for each region of interest indicate that the MT complex (MT+) and the frontal eye field (FEF) show evoked responses to event boundaries. The evoked responses were larger when the participants were actively segmenting the movies than when they were passively viewing the movies, and these responses were larger at coarse unit boundaries than at fine unit boundaries.**

individually defined boundary points during the coarse-grained segmentation condition than during the fine-grained segmentation condition. All four regions also showed significant main effects of temporal grain [left MT+, $F(1,10) = 10.72, p < .01$; right MT+, $F = 7.77, p = .02$; left FEF, $F = 19.01, p = .001$; right FEF, $F = 6.48, p = .03$].

As in the passive-viewing runs, it appeared that the evoked response to event boundaries in the MT+ was larger than the evoked response in the FEF. An ANOVA using region and time point (frame) as the two independent variables showed a significant interaction of region and time point [$F(3,130) = 3.99, p < .001$], although there was not an overall effect of region [$F(1,10) = 1.99, p = .19$]. This analysis confirmed the observation that the MT+ showed a larger evoked response to event boundaries than did the FEF.

## Reliability of Responses to Event Boundaries

**Behavioral reliability.** The reliability of the participants' perceptual segmentation was evaluated by comparing the locations of fine-unit segment boundaries identified by each participant during each of the two sessions. Each movie was divided into 1-sec intervals, and each interval

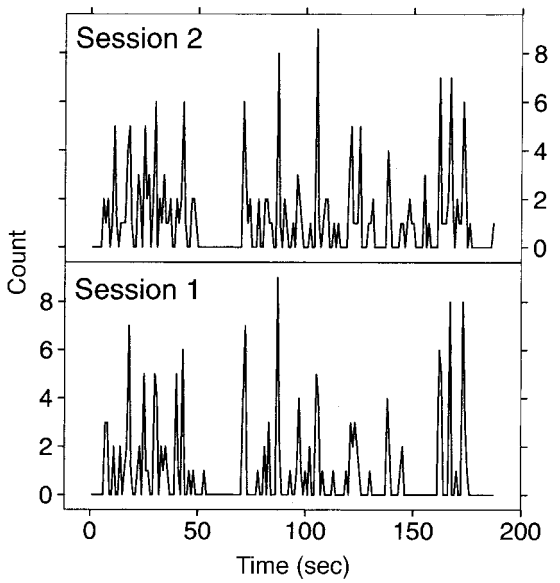


**Figure 4. Observers consistently identified the same moments as perceptual segment boundaries. Each panel plots the number of observers who identified a segment boundary between each 1-sec interval of one of the movies (assembling a saxophone). The top panel shows data from Session 2, and the bottom panel shows data from Session 1, which occurred 13–15 months earlier.**

was coded as a segment boundary for a particular observer if that observer pressed the segmentation key during the interval. We then counted the number of observers who identified each interval as a segment boundary during each of the two sessions. At the group level, the observers picked out similar intervals as segment boundaries during the two sessions: The correlations between the Session 1 and the Session 2 boundary counts for the four movies ranged from .71 to .77. Figure 4 gives an example of segmentation patterns during the two sessions.

However, good group agreement would be possible even if each individual did not reliably identify the same segment boundaries. To measure test–retest reliability at the individual level, we calculated the proportion of intervals identified as segment boundaries during either session that were identified as boundaries in both sessions. By this measure, perfect agreement leads to a score of 1, and perfect disagreement to a score of zero. The mean proportion was .38 ($SD$ = .16). Chance agreement would be equal to the simple probability of identifying an interval as a segment boundary, which had a mean of .10 ($SD$ = .03). Agreement across the two sessions was significantly greater than chance [$t(10) = 6.19, p < .001$].

The proportion-of-agreement measure also provides a means by which to characterize whether individual differences in perceptual segmentation were stable over time. If they were, one would expect agreement between the same individual across the two sessions to be higher than agreement between that individual and the other observers during each of the two sessions. This was indeed the case. In Session 1, an observer agreed with each other observer with a mean proportion of .28 ($SD$ = .06). In Session 2, the mean proportion of agreement was .28 ($SD$ = .05). In both cases, these proportions were significantly lower than the degree to which observers agreed with themselves across sessions [for Session 1, $t(10) = 2.57, p = .02$; for Session 2, $t(10) = 2.72, p = .02$]. In short, perceptual segmentation demonstrated both stable intersubject agreement and stable individual differences over a period of more than a year.

**Reliability of evoked neural responses.** To evaluate the reliability of evoked responses to event boundaries, we used the imaging data from the passive viewing of the event movies in Session 2. The behavioral coarse- and fine-grained segmentation data were taken from Session 1. (This was done to maximize comparability of the two analyses and to eliminate memory effects in the Session 2 segmentation data, and also because there were no coarse segmentation data collected in Session 2.) The BOLD response was estimated for each voxel in the Session 2 passive-viewing data. A 34.56-sec (16-frame) window surrounded the coarse and fine segment boundaries identified from the behavioral data. This produced two 34.56-sec time courses per voxel for each participant in the passive-viewing condition. The time courses of voxels within the previously defined regions, as well as within the individually defined FEF and MT+ regions, were averaged, to generate one coarse-grained and one fine-grained time course for each region for each participant. A 2 × 16 ANOVA (grain × time point) was conducted for each region. As can be seen in Figure 5, most of the regions that showed evoked responses to event boundaries in the first testing session (left and right posterior inferior temporal sulcus, left and right fusiform gyrus, right precuneus, right precentral sulcus, left cuneus, and right superior temporal sulcus; see Zacks, Braver, et al., 2001) also showed significant effects of event boundaries (i.e., showed significant main effects of time point) during the second passive viewing of the movies [smallest $F(15,150) = 2.36, p = .004$]. The one exception was the right prefrontal region (Brodmann's Area 40, 9, 39), which showed only a marginally significant effect of time point [$F(15,150) = 1.66, p = .07$]. The activation maps from Sessions 1 and 2 were qualitatively similar, although the magnitude and extent of activation in Session 2 were not as large as those in Session 1.

In addition, both left and right individually defined MT+ regions showed a significant response at event boundaries in the Session 2 passive-viewing task [$F(15,150) = 1.86$, $p = .03$, and $F(15,150) = 2.66$, $p = .001$, respectively]. Only the left individually defined FEF region showed a significant effect of timepoint [$F(15,150) = 1.95, p = .02$; right FEF, $F(15,150) = 0.72, p = .76$].

## DISCUSSION

The primary goal of the present study was to characterize the behavior of the MT+ and the FEF, as defined by tasks that localized motion processing and eye movements, during perceptual event boundaries. The data provided strong evidence that the MT+ is involved in event

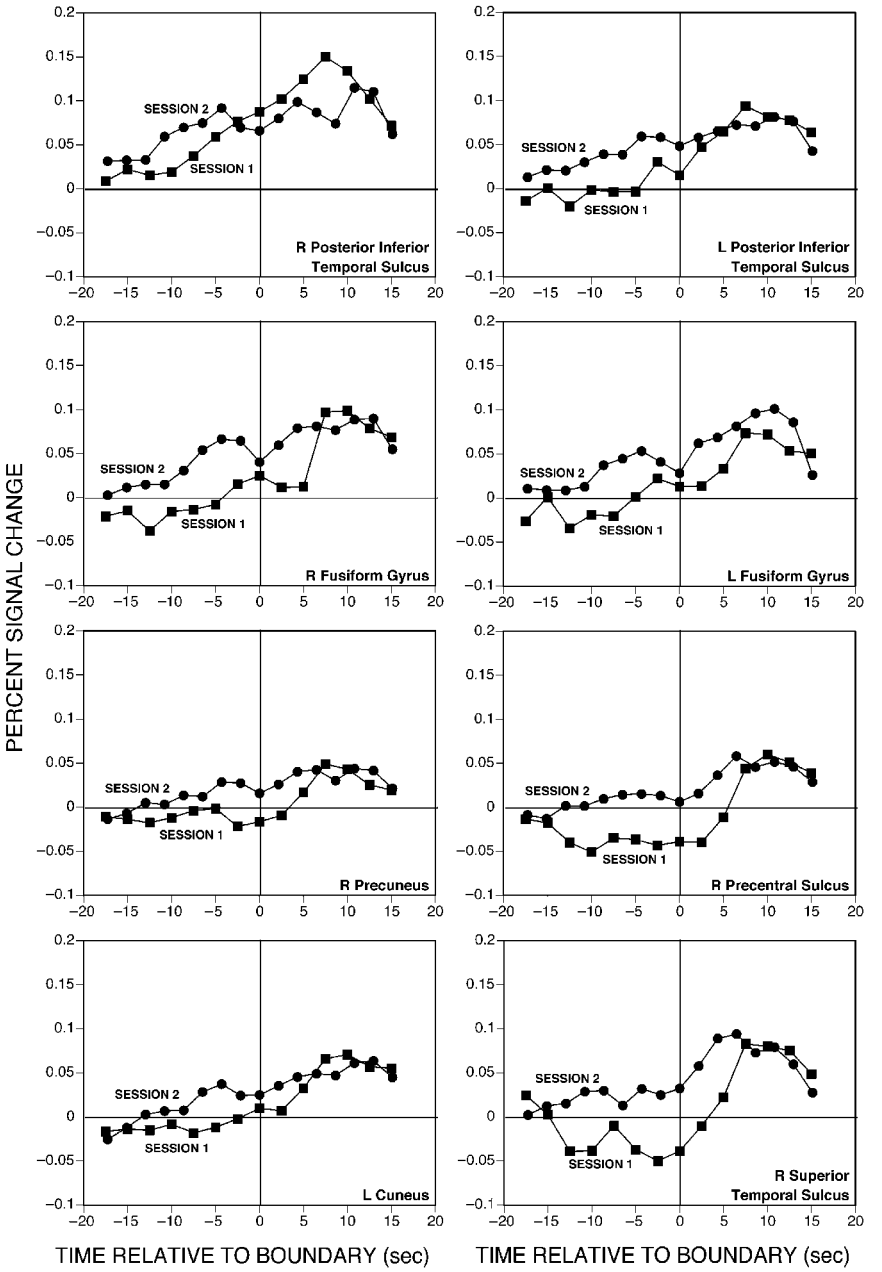


**Figure 5. The eight regions that showed evoked responses to fine and coarse unit boundaries in Session 1 (Zacks, Braver, et al., 2001) were tested using functional data from the passive-viewing condition in Session 2. Each of the eight regions showed reliable changes in activation in response to unit boundaries that were defined 13–15 months earlier during Session 1 (squares), and the overall patterns of this activation (increasing activity leading up to and following the unit boundary) were preserved during Session 2 (circles).**

perception. Functionally defined MT+ regions showed evoked responses at the points that the participants identified as event boundaries, and this activation was greater during the segmentation condition than during the passive-viewing condition (see Figure 3). This MT+ activity during the segmentation condition was greater for coarse-grained segmentation than for fine-grained segmentation, and this activity began before the event boundary. This pattern of results is qualitatively very similar to the pattern observed in posterior regions that have previously shown responses to event boundaries. In addition, the locations of the functionally defined individual and group MT+ regions were re-

liably aligned with the locations of posterior regions previously found to be active at event boundaries (see Figure 2).

In contrast, the data provided mixed evidence that the FEF is involved in event perception. First and foremost, the right frontal event boundary-related activation appears to be lateral to the FEF (see Figure 2). Functionally defined FEF regions did show statistically significant changes in activation in response to event boundaries. However, as compared with the changes in activation observed in the MT+, this activity was greatly reduced when the participants were passively viewing the movies (see Figure 3). The magnitude of this effect was much larger in the active segmentation conditions. In addition, although the pattern of activity observed in the FEF was qualitatively very similar to the pattern of activity in the right frontal region previously identified as responding to event boundaries, the functionally defined individual and group FEF regions failed to show any reliable overlap with the right frontal region. The results of the FEF analysis do not rule out involvement of the FEF in the neural processing of event structure, but they fail to provide strong evidence for such a role. The absence of activity in parietal areas associated with visual attention shifts is consistent with the interpretation that attention shifts are not critical to event structure perception.

The finding that the MT+ and, to a lesser degree the FEF, responded to event boundaries and that these responses were increased by explicit instructions to attend to event structure supports theories proposing that event structure perception is related to changes in the visual environment (Newtson et al., 1977). If event perception depends in part on detecting visual changes in the environment, event boundaries should emerge near the points at which there is a high degree of visual change in the activity of the characters in the movies. Activity in the MT+ has been related to the perception of visual motion (Kourtzi & Kanwisher, 2000; Salzman et al., 1990; Tootell et al., 1995), and activity in the FEF has been related to saccadic and smooth pursuit eye movements (Petit & Haxby, 1999; Rosano et al., 2002), as well as to covert direction of attention (Corbetta et al., 1998), indicating that these regions should also respond to points at which there is a high degree of visual change. These results provide indirect evidence that event perception is related to the detection of visual changes in the environment.

The present data leave open the question of whether this detection is driven by conceptual processes, such as cognitive representations, that anticipate the emergence of visual change or by lower level motion changes, which trigger the perception of an event boundary when a sufficient level of visual change is present. It may be the case that regions involved in representing knowledge related to the activities depicted in the movies signal the impending event boundary to the MT+, leading to the observed pattern of activation. However, it is also possible that greater motion changes at event boundaries directly triggers activation in lower level visual regions that monitor these changes, such as the MT+. Increased activity in the MT+ might then signal the presence of an event boundary to regions higher in the processing stream, such as regions in the prefrontal cortex that are important in planning and comprehending everyday activities (Crozier et al., 1999; Partiot, Grafman, Sadato, Flitman, & Wild, 1996; Sirigu et al., 1996).

There are many regions involved in motion processing in humans (see Culham, He, Dukelow, & Verstraten, 2001, for a review), and many involved in attention shifts (e.g., Corbetta & Shulman, 2002). However, here we have focused on the MT+ and the FEF. This was primarily because initial results had implicated these regions (Zacks et al., 2001). These two regions were also of interest because (1) the response properties of cells in these areas had been well characterized and (2) each region is relatively selective in the situations in which it is activated.

The design of the present study also made it possible to determine the reliability of neural and behavioral responses to event boundaries over a period of more than a year. There was a high degree of overlap in the points that the participants identified as event boundaries across the two testing sessions (see Figure 4), despite the fact that the delay between testing sessions was more than a year. It is possible that this overlap was due to the participants' memories for the locations of event boundaries. That is, the participants' responses during Session 2 may have been based on their memories for their responses in Session 1, rather than on where they perceived the event boundaries during Session 2. However, subjective reports from the participants indicated that many of them did not remember having participated in Session 1. This, in addition to the length of the interval between sessions, argues against a memory explanation for the high degree of overlap in the locations of boundaries across sessions.

The points in the movies that the participants defined as event boundaries during the original viewing session showed evoked responses in brain activation in the second testing session, more than a year after the behavioral data were collected. Although the magnitude of activation in the second testing session was decreased relative to the magnitude of activation in the first testing session, the regions that previously responded to event boundaries showed significant responses to event boundaries over 1 year later. Because both the behavioral data and the neural activity were reliable over time, the reliability of the behavioral and neural responses may be related.

The results of the present study demonstrate that motion perception and, possibly, eye movements or shifts of attention are fundamentally related to the perception of event boundaries and that the perception of these boundaries is relatively stable over time. These results firmly establish the importance of the MT+ to event perception and leave open the possibility that the FEF contributes to event perception. Future studies will need to address whether these regions contribute to event perception because of modulation by top-down processes, bottom-up activation by the stimulus, or some combination of the two.

## REFERENCES


BALDWIN, D. A., & BAIRD, J. A. (1999). Action analysis: A gateway to intentional inference. In P. Rochat (Ed.), *Early social cognition* (pp. 215-240). Hillsdale, NJ: Erlbaum.

BARKER, R. G., & WRIGHT, H. F. (1954). *Midwest and its children: The psychological ecology of an American town.* Evanston, IL: Row, Peterson.

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. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.

CORBETTA, M., AKBUDAK, E., CONTURO, T. E., SNYDER, A. Z., OLLINGER, J. M., DRURY, H. A., LINENWEBER, M. R., PETERSEN, S. E., RAICHLE, M. E., VAN ESSEN, D. C., & SHULMAN, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, **21**, 761-773.

CORBETTA, M., & SHULMAN, G. L. (2002). Control of goal-directed and stimulus driven attention in the brain. *Nature Reviews Neuroscience*, **3**, 201-215.

CROZIER, S., SIRIGU, A., LEHÉRICY, S., VAN DE MOORTELE, P.-F., PILLON, B., GRAFMAN, J., AGID, Y., DUBOIS, B., & LEBIHAN, D. (1999). Distinct prefrontal activations in processing sequence at the sentence and script level: An fMRI study. *Neuropsychologia*, **37**, 1469-1476.

CULHAM, J., HE, S., DUKELOW, S., & VERSTRATEN, F. A. J. (2001). Visual motion and the human brain: What has neuroimaging told us? *Acta Psychologica*, **107**, 69-94.

FRISTON, K. J., HOLMES, A. P., WORSLEY, K. J., POLINE, J. P., FRITH, C. D., & FRACKOWIAK, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, **2**, 189-210.

HUK, A. C., & HEEGER, D. J. (2000). Task-related modulation of visual cortex. *Journal of Neurophysiology*, **83**, 3525-3536.

KOURTZI, Z., BÜLTHOFF, H. H., ERB, M., & GRODD, W. (2002). Object-selective responses in the human motion area MT/MST. *Nature Neuroscience*, **5**, 17-18.

KOURTZI, Z., & KANWISHER, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, **12**, 48-55.

KOURTZI, Z., & KANWISHER, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, **293**, 1506-1509.

MCAVOY, M. P., OLLINGER, J. M., & BUCKNER, R. L. (2001). Cluster size thresholds for assessment of significant activation in fMRI [Abstract]. *NeuroImage*, **13**, S198.

NELSON, K., & GRUENDEL, J. (1986). Children's scripts. In K. Nelson (Ed.), *Event knowledge: Structure and function in development* (pp. 21-46). Hillsdale, NJ: Erlbaum.

NEWTSON, D. (1976). Foundations of attribution: The perception of ongoing behavior. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 223-248). Hillsdale, NJ: Erlbaum.

NEWTSON, D., & ENGQUIST, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, **12**, 436-450.

NEWTSON, D., ENGQUIST, G., & BOIS, J. (1977). The objective basis of behavior units. *Journal of Personality & Social Psychology*, **35**, 847-862.

PARTIOT, A., GRAFMAN, J., SADATO, N., FLITMAN, S., & WILD, K. (1996). Brain activation during script event processing. *NeuroReport*, **7**, 761-766.

PAUS, T. (1996). Location and function of the human frontal eye-field: A selective review. *Neuropsychologia*, **34**, 475-483.

PETIT, L., & HAXBY, J. V. (1999). Functional anatomy of pursuit eye movements in humans as revealed by fMRI. *Journal of Neurophysiology*, **82**, 463-471.

ROSANO, C., KRISKY, C. M., WELLING, J. S., EDDY, W. F., LUNA, B., THULBORN, K. R., & SWEENEY, J. A. (2002). Pursuit and saccadic eye movement subregions in human frontal eye field: A high-resolution fMRI investigation. *Cerebral Cortex*, **12**, 107-115.

RUMELHART, D. E. (1977). Understanding and summarizing brief stories. In D. Laberge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 265-303). Hillsdale, NJ: Erlbaum.

SALZMAN, C. D., BRITTEN, K. H., & NEWSOME, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, **346**, 174-177.

SCHANK, R. C., & ABELSON, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.

SIRIGU, A., ZALLA, T., PILLON, B., GRAFMAN, J., AGID, Y., & DUBOIS, B. (1996). Encoding of sequence and boundaries of scripts following prefrontal lesions. *Cortex*, **32**, 297-310.

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.* Stuttgart: Thieme.

TONG, F., NAKAYAMA, K., MOSCOVITCH, M., WEINRUB, O., & KANWISHER, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology*, **17**, 257-279.

TOOTELL, R. B. H., 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. *Journal of Neuroscience*, **15**, 3215-3230.

VAN ESSEN, D. C. (2002a). *Surface management system* [Computer database of surface-based atlases for the macaque and human cerebral cortex]. Retrieved from http://pulvinar.wustl.edu:8081/sums/search.do?filename=ATLAS.

VAN ESSEN, D. C. (2002b). Windows on the brain: The emerging role of atlases and databases in neuroscience. *Current Opinion in Neurobiology*, **12**, 574-579.

VAN ESSEN, D. C., DICKSON, J., HARWELL, J., HANLON, D., ANDERSON, C. H., & DRURY, H. A. (2001). An integraged software system for surface-based analyses of cerebral cortex. *Journal of American Medical Informatics Association*, **41**, 1359-1378.

VAN ESSEN, D. C., DRURY, H. A., HARWELL, J., & HANLON, D. (2002). *CARET*: *Computerized anatomical reconstruction and editing toolkit* [Computer software and manual]. Retrieved from http://brainmap.wustl.edu/caret

WOODWARD, A. L., & SOMMERVILLE, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science*, **11**, 73-77.

WYNN, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science*, **7**, 164-169.

ZACKS, J. M., BRAVER, T. S., SHERIDAN, M. A., DONALDSON, D. I., SNYDER, A. Z., OLLINGER, J. M., BUCKNER, R. L., & RAICHLE, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, **4**, 651-655.

ZACKS, J. M., & TVERSKY, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, **127**, 3-21.

ZACKS, J. M., TVERSKY, B., & IYER, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, **130**, 29-58.


(Manuscript received July 28, 2003;
revision accepted for publication December 14, 2003.)