Implied Motion Activation in Cortical Area MT Can Be Explained by Visual Low-level Features

Jeannette A. M. Lorteije1,2, Nick E. Barraclough3,4, Tjeerd Jellema3, Mathijs Raemaekers1, Jacob Duijnhouwer1, Dengke Xiao4, Mike W. Oram4, Martin J. M. Lankheet1,5, David I. Perrett4, and Richard J. A. van Wezel1,6

Abstract
To investigate form-related activity in motion-sensitive cortical areas, we recorded cell responses to animate implied motion in macaque middle temporal (MT) and medial superior temporal (MST) cortex and investigated these areas using fMRI in humans. In the single-cell studies, we compared responses with static images of human or monkey figures walking or running left or right with responses to the same human and monkey figures standing or sitting still. We also investigated whether the view of the animate figure (facing left or right) that elicited the highest response was correlated with the preferred direction for moving random dot patterns. First, figures were presented inside the cell’s receptive field. Subsequently, figures were presented at the fovea while a dynamic noise pattern was presented at the cell’s receptive field location. The results show that MT neurons did not discriminate between figures on the basis of the implied motion content. Instead, response preferences for implied motion correlated with preferences for low-level visual features such as orientation and size. No correlation was found between the preferred view of figures implying motion and the preferred direction for moving random dot patterns. Similar findings were obtained in a smaller population of MST cortical neurons. Testing human MT+ responses with fMRI further corroborated the notion that low-level stimulus features might explain implied motion activation in human MT+. Together, these results suggest that prior human imaging studies demonstrating animate implied motion processing in area MT+ can be best explained by sensitivity for low-level features rather than sensitivity for the motion implied by animate figures.

INTRODUCTION
A photograph of an object or a person in motion can evoke a perception of motion on the basis of the content of the static image in typically developed human observers (Jellema et al., 2009). Human fMRI studies have shown that static photographs depicting an object or a person without this implied motion evoke a higher response than photographs of the same objects or persons with this implied motion (Kourtzi & Kanwisher, 2000; Senior et al., 2000). These studies suggest integration of object and motion information at a relatively early level of visual processing, which makes implied motion an interesting case for studying feedback projections of high-level multimodality cortical areas to low-level single modality areas (Kaas & Collins, 2004).

We tested whether neurons in macaque MT and medial superior temporal (MST) cortex are sensitive to implied motion as expressed by animate agents in static images. These recordings allowed us to compare temporal response properties with implied and real motion from the same cell. There is evidence from human EEG and magnetoencephalography studies that implied motion responses in dorsal motion processing areas are delayed compared with responses to real motion (Fawcett, Hillebrand, & Singh, 2007; Lorteije et al., 2006, 2007), potentially resulting from high-level feedback. Cells in the anterior regions of the STS (STSa) are specialized for the perception of bodily actions and postures in macaque monkeys (Barraclough, Xiao, Oram, & Perrett, 2006; Jellema & Perrett, 2003a, 2003b; Jellema, Baker, Wicker, & Perrett, 2000). These cells typically have response latencies of about 80 to 110 msec (Barraclough et al., 2006). Neurons in areas MT have response latencies for real motion stimuli of about 50 to 80 msec (Perge, Borghuis, Bours, Lankheet, & van Wezel, 2005b; Raiguel, Xiao, Marcar, & Orban, 1999; Schmolesky et al., 1998; Lagae, Maes, Raiguel, Xiao, & Orban, 1994). If single neurons in MT show implied motion-related activity as a consequence of feedback from form-sensitive cells in STSa, then MT cells’ implied motion responses may appear delayed relative to pure motion responses. In addition, because STSa neurons typically have much larger receptive fields than MT neurons (Jellema & Perrett, 2006), such a
feedback projection from STSa neurons could also induce modulation of MT cell motion responses regardless of the position of an animate agent with respect to the MT neurons’ receptive field. Finally, real motion can adapt visually evoked responses to implied motion in humans in a direction specific manner (Lorteije et al., 2007). Because MT neurons are directionally selective for real motion, the question arises whether they show similar direction preference for implied motion and whether modulation of responses to one motion type by the other occurs at the level of MT.

To address the above questions, we performed extra-cellular single and multiunit recordings in area MT of awake macaque monkeys while presenting the animal with pictures with and without implied motion as well as control stimuli for low-level visual features. In addition, we recorded from single units in the MST cortex a motion-sensitive area that receives direct input from MT but generally has larger receptive fields and is thought to process more complex motion patterns such as optic flow and heading perception (Orban, 2008). In a second experiment, MT responses to dynamic noise in the cells’ receptive field were recorded while pictures were presented at the fixation spot. This presentation allowed the animals a clear view of the images, simultaneously removing the differences in low-level features inside the receptive field between conditions. Dynamic noise typically consists of motion energy at a broad range of speeds and in all directions and will evoke activity above spontaneous level in almost all MT cells (Van Wezel & Britten, 2002; Britten & Newsome, 1998; Britten, Shadlen, Newsome, & Movshon, 1993). Thus, modulation of real motion responses by implied motion, via a putative feedback mechanism, can be tested. Finally, a human fMRI experiment was conducted to correlate the single-unit data with previous fMRI studies.

METHODS

MT/MST Recordings: Surgical and Recording Procedures

The recordings in MT and MST (Experiments 1 and 2) were conducted at the Department of Functional Neurobiology, Utrecht University, The Netherlands. Housing, surgical procedures, recording, handling, and all other procedures used in Experiments 1 and 2 were approved by the Animal Use Committee (DEC) of Utrecht University, and procedures followed national and international guidelines.

Two male rhesus macaques aged 6 and 7 years were implanted surgically with a head-holding device, a scleral search coil for measuring eye movements using the double induction technique (Malpeli, 1998; Reulen & Bakker, 1982), and a stainless-steel recording cylinder placed over a craniotomy above the left occipital lobe. For one animal, a second cylinder was placed dorsally over a craniotomy above the parietal/occipital region. All surgical procedures were performed under N2O/O2 anesthesia supplemented with isoflurane. After recovery, the monkeys were trained to fixate a rectangular red spot (0.4° × 0.4°). During experiments, each monkey sat in a primate chair 57 cm from a 19-in. monitor (refresh rate: 120 Hz in Experiment 1, 100 Hz in Experiment 2). When correctly fixating (±1°, sample rate 500 Hz), the monkey was rewarded with water or juice every second during the reverse correlation recordings and after every stimulus presentation during the implied motion recordings.

Extracellular single and multiunit recordings were carried out using standard methods as described previously (Perge, Borghuis, Bours, Lankheet, & van Wezel, 2005a). In short, a parylene-insulated tungsten microelectrode (0.5–2 MΩ impedance) was inserted manually through a guide tube and then manipulated by a micropositioning controller. Cortical areas MT and MST were identified by the recording position and depth, the transition between gray matter, white matter, and sulci along the electrode track, and by functional properties. For MT, these were the prevalence of direction-selective units, the similarity in direction tuning for nearby single-unit recordings, the receptive field size according to eccentricity, and the change of direction tuning along the electrode penetration. For MST, the functional properties were large receptive fields overlapping the fixation point and extending to the ipsilateral side and selectivity for complex motion patterns. In one monkey, the anatomical positions were confirmed with structural MRI scans of the brain containing a reference electrode inserted at the location of the recordings. Action potentials from single and multiunits were isolated with a window discriminator (BAK Electronics Inc., Mt. Airy, MD) for recordings in Monkey 1 and with online spike sorting software (ASD, Alpha Omega, Nazareth, Israel) for Monkey 2. Spike times were registered at a 0.5-msec resolution for online analysis and data storage using a Macintosh G4 computer with a National Instruments PCI 1200 data acquisition board.

MT/MST Recordings: Search Procedure

As a search stimulus, we used either the moving random dot patterns (RDPs) or the experimental images described below. Position and size of the receptive fields were mapped by projecting a light bar on a dark monitor with a hand-held projector while the monkey was fixating. Direction tuning for real motion was established using a motion reverse correlation paradigm (Perge et al., 2005a, 2005b; Borghuis et al., 2003). A translating RDP (subtending 8.5° × 11.5° in Experiment 1, same size as receptive field in Experiment 2) was presented centered on the cell receptive field. The pattern switched every frame between eight possible randomly chosen directions. The delay between successive steps was 10 or 8.3 msec, with a step size of 0.12° corresponding to velocities of 12°/sec and 14.4°/sec, respectively. The preferred direction determined with this reverse correlation paradigm is strongly correlated with the preferred direction measured conventionally with hand mapping or by presenting RDPs moving in different directions for long durations (>1 sec; Borghuis et al., 2003).
**Experiment 1 Procedure**

In Experiment 1, stimuli were color human or monkey figures implying motion facing left or right (Figure 1A and D) and standing or sitting facing forward (Figure 1C and F). In addition, images of bars that did not convey implied motion (Figure 1I and J), and scrambled images (Figure 1G and H) were presented as controls for low-level effects: orientation, position, and size. All figures subtended 8.5° × 11.5°. Pictures were presented in the receptive field of the cells, at an average eccentricity of 9° (SD = 2°). In human psychophysical studies, subjects are well able to discern shapes of similar size at these eccentricities (Näsinen & O’Leary, 1998). In the human pictures series, three different human agents were used (the three different persons are depicted in Figure 1A–C). The monkey pictures only showed one animal per condition (Figure 1D–F). Bars included only one type of tilted bar in two directions (Figure 1I) and one type of vertical bar (Figure 1J). All MT and MST units were tested with the same set of human figures, and 80% of the recordings also included the monkey figures and bars. The whole set of stimuli was presented in blocks, and within each block, the order was randomized. Blocks were repeated 10 to 20 times.

In Experiment 1, stimuli were presented at the center of the receptive field against a black background (Figure 2A) for 500 msec with an interstimulus interval of 250 msec (occasionally interstimulus interval up to 1000 msec). To increase the baseline firing rate of MT or MST cells during the presentation of the stimulus, thereby increasing the likelihood of finding modulating effects, we presented 18 cells in Monkey 1 with human figures flickering at 10 or 20 Hz for 500 msec. Responses of these cells to flickering and continuous stimulation were similar, and therefore these data were included in all analyses (except peristimulus time histograms) and are indicated separately in figures.

**Figure 1.** Humans and monkeys implying motion and sitting/standing were used in all experiments. The figures implying motion were identical in all experiments (A, D, profiled toward left or right), but the standing/sitting images could be viewed from the front, left, or right. In Experiment 1, monkey figures and human figures faced forward (C, F). In Experiment 2, both monkeys and humans sitting/standing were viewed from the left or right (B, E). To control for low-level stimulus features, we added control stimuli in Experiment 1. Bar images could be vertical to match standing images (J) or tilted left or right to match implied motion images (I). Furthermore, scrambled versions of the figures were used as control stimuli in Experiment 1 (G, H).

**Experiment 2 Procedure**

In Experiment 2, we presented foveally images of gray scale human and monkey figures implying motion (Figure 1A and D) or standing/sitting (Figure 1B and E), irrespective of the position of each cell’s receptive field. Images subtended 8.5° × 11.5° in most recordings; however, to prevent overlap of the images with the receptive field for eight units with near foveal receptive fields, images were reduced to 4.25° × 5.75°. Concurrently, a dynamic RDP (500 white dots subtending 0.2° × 0.2°, with a limited lifetime of 10 msec, black background) was presented optimally over the receptive field (Figure 2B). Both images and RDPs were presented simultaneously for 500 msec with 250-msec intervals.
MT/MST Recordings: Data Analysis

Data from the reverse correlation recordings were analyzed as described previously (Borghuis et al., 2003). In short, motion direction tuning functions were computed by reverse correlating the response to the rapid sequence of displacements of the RDP. The preferred direction was defined as the direction with maximum correlation and the non-preferred (null) direction as the direction opposite to the preferred direction. Furthermore, for each cell, we calculated the direction index (DI):

\[ DI = 1 - \left( \frac{\text{relative probability in null direction}}{\text{relative probability in preferred direction}} \right) \]

Response latency for the moving RDPs was established at the center of a 5-msec response window showing the highest direction-selectivity index. Cells were divided into four groups according to their direction tuning: (1) no direction selectivity, (2) tuned to upward or downward directions, (3) tuned to leftward direction (±45°), and (4) tuned to rightward directions (±45°). The responses of cells tuned for leftward or rightward motion were correlated with the responses to static images of figures implying motion facing left and right and bars that were tilted toward left and right.

Responses to static images in Experiments 1 and 2 were analyzed identically. For each cell, the response latency was calculated from the average response to all image conditions. Spontaneous activity and its standard deviation were calculated within a 100-msec window before stimulus onset. Response latencies to each image were calculated by determining the mean firing rate within a sliding 25-msec window (in 1-msec steps). The image response latency was established as the middle of the first window with a mean response of three standard deviations above spontaneous activity. For 4 MT units and 7 MST units in Experiment 1 and 1 MT unit in Experiment 2, with very small responses, latencies were set at the respective average MT or MST latency. Mean responses to each image were calculated within a 500-msec period after the respective response latency.

Responses of each cell to each stimulus were analyzed by applying a receiver operating characteristic (ROC) analysis (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Green & Swets, 1966). This analysis does not make any assumptions about the response distributions, and the ROC value expresses the amount of overlap of two response distributions for two conditions (where an ROC value = 0.5 means complete overlap and ROC value = 0 or 1.0 means no overlap of the response distributions). In our experiments, by definition ROC values larger than 0.5 signified stronger responses to implied motion stimuli. Furthermore, the ROC analysis allowed us to assess whether an ROC value is significantly different from 0.5 for each individual unit by performing a permutation test (n = 1000; Britten et al., 1996; Elfron & Tibshirani, 1993). Wilcoxon signed rank tests were used to test the differences in responses for different conditions across the population of units.

Human fMRI: Subjects, Scanning Protocol, and Task Design

Eight healthy subjects (five men and three women, average age = 24.8 years, SD = 6.1 years) who were recruited from the staff and students of Utrecht University participated in the experiment. All subjects gave informed consent for participation (approved by the Human Ethics Committee of the University Medical Center Utrecht).

All images were obtained with a Philips Achieva 3-T MRI scanner (Philips Medical Systems, Best, the Netherlands) with a Quasar Dual gradient set. Functional scans were acquired in sagittal orientation using a navigated 3D-PRESTO pulse sequence (Ramsey et al., 1998; van Gelderen et al., 1995) with the following parameters: repetition time =
21.75 msec (time between two subsequent RF pulses), effective echo time = 32.4 msec, field of view (anterior–posterior, inferior–superior, and right–left) = 224 × 256 × 128 mm, flip angle = 10°, matrix = 56 × 64 × 32 slices, voxel size = 4 mm isotropic, eight channel head coil, SENSE factors = 2.0 (left–right) and 1.8 (anterior–posterior). The total acquisition time per volume was 500.3 msec. A T1-weighted structural image was acquired after functional scanning. A PC, a rear projection screen, and a video projector were used for task presentation. All stimuli were projected on a gray background. All events were time locked to the fMRI scans.

Motion-sensitive areas were mapped by presenting RDPs intermittently for 1000 msec, with a 300-msec interval. RDPs alternated between moving left or right (10°/sec) and remaining stationary. Each block lasted 26 sec, and there were 18 blocks in total. The three main image types were human pictures, bars, and monkey pictures. Images of humans implying motion (running) and without implied motion (standing) were generated for three different actors when facing both left and right (Figure 3, top panels). All actors were seen on a gray background with a height of 350 pixels, subtending 13.2° visual angle. Images of bars consisted of three different vertical bars, subtending 13.2° that resembled a standing human actor, and three different combinations of two tilted bars that resembled either a human running left or right (Figure 3, middle panels). The relative luminance of all images (humans and bars) was matched. Images of monkeys implying motion (n = 40) and not implying motion (n = 40) consisted of digitized color photographs (250 × 350 pixels, 9.4° × 13.2° visual angle) of monkeys in their natural surroundings (Figure 3, bottom panels).

The experimental design that we used was similar to the study of Kourtzi and Kanwisher (2000), except that we used stimuli similar to those used during the electrophysiological recordings and bar stimuli controlling for low-level features, as described above. The fMRI experiment consisted of blocks of seven image types, including monkey and human figures implying and not implying motion and tilted and vertical oval bars. One subject was not presented with the tilted and vertical bars. A new picture was presented every 1000 msec (two scans) for 300 msec (total 20 images/block). In addition, there was a rest condition in which only the red fixation dot was displayed on the screen for 20 sec. There were three scanning sessions containing 21 blocks (3 blocks per condition). During the experiment, subjects were required to fixate a central red dot throughout all stimulus presentations. After the experiment, participants reported that they recognized the presence or the absence of implied motion in the human and monkey pictures, although they were not informed regarding the exact aim of the study nor received explanation about implied motion beforehand.

**Human fMRI: Analysis**

The fMRI time series data were preprocessed using SPM2 (http://www.fil.ion.ucl.ac.uk/spm/spm2.html). Preprocess-
contained six factors for stimulus-related changes in BOLD signal during the six conditions in which pictures were shown (the rest condition was used as a reference). All events in the design matrix were convolved with a predefined hemodynamic response function (Friston, Frith, Turner, & Frackowiak, 1995).

MT+ was defined in each subject by the contrast in activation between the presentation of static and moving random patterns in the ascending limb of the inferior temporal sulcus \((p < .05,\) Bonferroni corrected). Subsequently, we calculated the average regression coefficients for the six conditions of the implied motion experiment within area MT+. Differences between the implied and motionless stimuli were tested using a paired samples \(t\) test.

**RESULTS**

**Direction Selectivity**

In total, 68 MT and 26 MST single and multiunits were recorded from two monkeys in Experiment 1 and 69 MT cells from one monkey in Experiment 2. Mean MT unit receptive field eccentricity was 9.6° \((SD = 4.6°)\) with an average diameter of 8.2° \((SD = 2.9°)\). MST cells had much larger receptive fields that were often hard to measure precisely, often extending to ipsilateral locations. Direction selectivity for real motion was obtained with the motion reverse correlation paradigm (Borghuis et al., 2003) using eight motion directions, and a DI was calculated. The mean DI of all recorded MT units was 0.4 \((SD = 0.3)\), with an average peak response latency of 59.1 msec \((SD = 18.2 msec)\); the mean MST DI was 0.7 \((SD = 0.2)\), with an average peak response latency of 71.9 msec \((SD = 13.4 msec)\). The DI for MT cells seems low compared with those obtained in previous studies (Snowden, Treue, & Andersen, 1992; Mikami, Newsome, & Wurtz, 1986). These studies varied the speed of their moving stimuli. A huge portion of MT neurons is speed selective, and DI values can be lower at nonoptimal speed (Mikami et al., 1986). Because of time restraints, we tested cells with only one speed (12°/sec or 14.4°/sec) to quickly obtain the preferred direction and to continue with the implied motion experiments. These nonoptimized presentation speed would cause a lower DI, but they are in the range reported earlier for MT (Perge et al., 2005a, 2005b; Borghuis et al., 2003).

**Experiment 1: MT/MST Cell Responses to Implied Motion within the Receptive Field**

In Experiment 1, we recorded 40 MT single units and 14 MT multiunits as well as 13 MST single units and 9 MST multiunits, with a nonflickering stimulus presentation. Other cells were tested with flickering stimuli (6 MT single units, 8 MT multiunits, and 4 MST single units). Stimuli were presented at the center of the receptive field where the visual responses and motion selectivity were strongest. The average eccentricity of the stimulus was 9° \((SD = 2°)\).
The results illustrated in Figure 4 show that there are no clear differences between the responses to figures implying motion and the responses to figures not implying motion in either MT (Figure 4A–C) or MST (Figure 4D) cells (see also Table 1). For units with a DI of at least 0.1 and a preferred direction along the horizontal axis (±45°) in the motion reverse correlation test, direction selectivity for implied motion was tested. No difference was found when comparing responses with figures implying motion (or tilted bars) when seen from views congruent or incongruent with the preferred motion direction (Figure 4E–H, see also Table 2).

The peristimulus time histograms in Figure 4 illustrate a transient response with an average peak response latency of 65.0 msec (SD = 16.6 msec) for MT units (n = 54) and 64.6 msec (SD = 18.9 msec) for MST units (n = 22) when presented with nonflickering stimuli. After this peak, average MT responses were sustained and higher than their spontaneous activity until the end of the stimulus presentation. MST responses dropped to or below spontaneous activity. The mean spontaneous firing rate of MT neurons was 7.9 spikes per second (SD = 7.1 spikes per second), and the mean sustained (200–400 msec after stimulus onset) firing rate was 15.1 spikes per second (SD = 16.8 spikes per second). For MST neurons, mean spontaneous activity was 11.0 spikes per second (SD = 6.5 spikes per second), whereas the mean sustained

<table>
<thead>
<tr>
<th>Area</th>
<th>Figures</th>
<th>No. Cells</th>
<th>Average Firing Rate Implied Motion/ Standing–Sitting</th>
<th>Average ROC Values</th>
<th>% Significant Cells for Implied vs. Standing–Sitting</th>
<th>Average ROC Value for Significant Cells</th>
<th>Average Firing Rate Implied/ Standing–Sitting for Significant Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>Humans</td>
<td>54</td>
<td>1.29 ± 1.61</td>
<td>0.53 ± 0.09</td>
<td>25.9</td>
<td>0.59 ± 0.12</td>
<td>1.28 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>Humans flicker</td>
<td>14</td>
<td>0.95 ± 0.24</td>
<td>0.52 ± 0.12</td>
<td>14.3</td>
<td>0.54 ± 0.3</td>
<td>0.99 ± 0.16</td>
</tr>
<tr>
<td></td>
<td>Monkeys</td>
<td>43</td>
<td>1.44 ± 2.61</td>
<td>0.48 ± 0.15</td>
<td>25.3</td>
<td>0.50 ± 0.28</td>
<td>1.13 ± 0.65</td>
</tr>
<tr>
<td></td>
<td>Bars</td>
<td>43</td>
<td>1.23 ± 0.57</td>
<td>0.6 ± 0.14</td>
<td>23.3</td>
<td>0.7 ± 0.13</td>
<td>1.68 ± 0.74</td>
</tr>
<tr>
<td>MST</td>
<td>Humans</td>
<td>22</td>
<td>1.00 ± 0.2</td>
<td>0.51 ± 0.01</td>
<td>9.1</td>
<td>0.5 ± 0.19</td>
<td>0.96 ± 0.41</td>
</tr>
<tr>
<td></td>
<td>Humans flicker</td>
<td>4</td>
<td>1.18 ± 0.11</td>
<td>0.56 ± 0.04</td>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Monkeys</td>
<td>12</td>
<td>1.11 ± 0.32</td>
<td>0.48 ± 0.1</td>
<td>25.0</td>
<td>0.46 ± 0.22</td>
<td>1.25 ± 0.62</td>
</tr>
<tr>
<td></td>
<td>Bars</td>
<td>12</td>
<td>1.06 ± 0.37</td>
<td>0.5 ± 0.1</td>
<td>8.3</td>
<td>0.67 (n = 1)</td>
<td>2.03 (n = 1)</td>
</tr>
</tbody>
</table>

Variance is indicated with standard deviation.

The results illustrated in Figure 4 show that there are no clear differences between the responses to figures implying motion and the responses to figures not implying motion in either MT (Figure 4A–C) or MST (Figure 4D) cells (see also Table 1). For units with a DI of at least 0.1 and a preferred direction along the horizontal axis (±45°) in the motion reverse correlation test, direction selectivity for implied motion was tested. No difference was found when comparing responses with figures implying motion (or tilted bars) when seen from views congruent or incongruent with the preferred motion direction (Figure 4E–H, see also Table 2). The peristimulus time histograms in Figure 4 illustrate a transient response with an average peak response latency of 65.0 msec (SD = 16.6 msec) for MT units (n = 54) and 64.6 msec (SD = 18.9 msec) for MST units (n = 22) when presented with nonflickering stimuli. After this peak, average MT responses were sustained and higher than their spontaneous activity until the end of the stimulus presentation. MST responses dropped to or below spontaneous activity. The mean spontaneous firing rate of MT neurons was 7.9 spikes per second (SD = 7.1 spikes per second), and the mean sustained (200–400 msec after stimulus onset) firing rate was 15.1 spikes per second (SD = 16.8 spikes per second). For MST neurons, mean spontaneous activity was 11.0 spikes per second (SD = 6.5 spikes per second), whereas the mean sustained

<table>
<thead>
<tr>
<th>Area</th>
<th>Figures</th>
<th>No. Cells</th>
<th>Average Firing Rate Preferred/ Nonpreferred</th>
<th>Average ROC Values</th>
<th>% Significant Cells for Preferred vs. Nonpreferred</th>
<th>Average ROC Value for Significant Cells</th>
<th>Average Firing Rate Preferred/ Nonpreferred for Significant Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>Humans</td>
<td>42</td>
<td>1.08 ± 0.35</td>
<td>0.51 ± 0.09</td>
<td>14</td>
<td>0.57 ± 0.18</td>
<td>1.23 ± 0.34</td>
</tr>
<tr>
<td></td>
<td>Humans flicker</td>
<td>12</td>
<td>1.15 ± 0.46</td>
<td>0.49 ± 0.15</td>
<td>17</td>
<td>0.56 ± 0.44</td>
<td>1.06 ± 0.37</td>
</tr>
<tr>
<td></td>
<td>Monkeys</td>
<td>35</td>
<td>1.09 ± 0.33</td>
<td>0.56 ± 0.13</td>
<td>17</td>
<td>0.60 ± 0.24</td>
<td>1.14 ± 0.31</td>
</tr>
<tr>
<td></td>
<td>Bars</td>
<td>35</td>
<td>1.46 ± 2.89</td>
<td>0.51 ± 0.13</td>
<td>14</td>
<td>0.65 ± 0.23</td>
<td>1.30 ± 0.31</td>
</tr>
<tr>
<td>MST</td>
<td>Humans</td>
<td>17</td>
<td>0.98 ± 0.32</td>
<td>0.53 ± 0.11</td>
<td>17</td>
<td>0.73 ± 0.07</td>
<td>0.98 ± 0.85</td>
</tr>
<tr>
<td></td>
<td>Humans flicker</td>
<td>1</td>
<td>1.03 (n = 1)</td>
<td>0.54 (n = 1)</td>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Monkeys</td>
<td>11</td>
<td>1.11 ± 0.29</td>
<td>0.51 ± 0.12</td>
<td>9</td>
<td>0.77 (m = 1)</td>
<td>1.63 (n = 1)</td>
</tr>
<tr>
<td></td>
<td>Bars</td>
<td>11</td>
<td>1.09 ± 0.42</td>
<td>0.47 ± 0.15</td>
<td>29</td>
<td>0.56 ± 0.36</td>
<td>1.33 ± 0.9</td>
</tr>
</tbody>
</table>

Variance is indicated with standard deviation.
response was 10.0 spikes per second ($SD = 7.8$ spikes per second). Spontaneous and sustained firing rates were compared using nonparametric tests and differed significantly for MT neurons (Wilcoxon signed rank test, $p < .001$) but not for MST neurons (Wilcoxon signed rank test, $p > .05$).

**Experiment 1: Cell-by-Cell Analysis**

Responses to every condition were averaged for all cells and plotted against each other (Figure 5, see Supplemental Figure 1 for histograms of response ratios). Response preferences of the MT and MST neuronal population were
tested with a nonparametric paired test (Wilcoxon signed rank test, \( p < .05 \) is considered significant). No significant difference was found between responses to figures implying motion and sitting/standing figures, neither for the human figures (continuously presented or flickering; Figure 5A) nor for the monkey figures (Figure 5C). No difference was found between responses to tilted and vertical bars (Figure 5E). In addition, no significant differences were found in responses to implied motion figures facing the preferred motion direction versus facing the nonpreferred motion direction for both human (Figure 5B) and monkey (Figure 5D) figures or tilted bars (Figure 5F). Comparing response magnitudes of MST neurons in the nonflickering condition (indicated in Figure 5 with square symbols) between figures implying motion and sitting/standing, tilted and vertical bars as well as preferred versus nonpreferred direction did not reveal any significant differences. Average responses to scrambled images versus all unscrambled images for 49 MT and 26 MST units also did not differ significantly (data not shown).

Although we did not find evidence for implied motion processing in areas MT and MST in the population data, ROC analysis of units separately is necessary to exclude the possibility that a subset of neurons is modulated by motion implied by the figures (Table 1). Neurons with a significant ROC value are marked with filled symbols in Figures 5 and 6. Under our definitions, an ROC value \( > 0.5 \) means a higher response to figures implying motion versus standing figures. Because the average ROC values for the human figures were above 0.5 for MT cells (but equal to 0.5 in MST cells), one might conclude that this small subset of MT neurons was significantly more responsive in the expected direction. However, as will be looked at later, we think this is due to low-level effects, as the response to bar figures shows a similar direction yet larger amplitude. Furthermore, the preference for implied motion stimuli was not present for the monkey images in MT cells and was even opposite in MST cells.

We compared eccentricities of the receptive field for MT cells with significant ROC values versus MT cells with insignificant ROC values in independent \( t \) tests for the human, monkey, and bar conditions separately. Eccentricity was not significantly different between the two groups (\( p \) values > .05).

**Figure 6.** A comparison of ROC values for responses to figures of humans and monkeys and tilted bars in Experiment 1. (A) As selectivity for human figures implying motion versus figures of standing humans increased, so did selectivity for tilted bars versus vertical bars. Cells showing significantly different responses for either the human figures or the bar figures are indicated with filled symbols. (B) No clear trend is apparent for ROC values for response preferences to monkey figures and bars. (C) Implied motion preference for human figures and monkey figures was not consistent; five cells with a significant preference for human implied motion even had a significant preference for figures of sitting monkeys. The mean ROC values with standard deviations of all MT cells are shown as black bars.
In our total MT population, 26% (14/54) of the units had significantly different responses for implied motion versus standing human images in the nonflickering condition. For this subgroup of cells, the ratio of the average response to humans implying motion versus standing human images was, on average, 1.3, corresponding to a difference of 2.2 spikes per second. Response ratios and average response differences for figures with monkeys implying motion versus standing or skewed versus vertical bars for these significant cells were 1.13 with 0.19 spikes per second and 1.68 with 5.55 spikes per second, respectively. This indicated that although a quarter of the MT cells responded significantly more to either figures implying motion or figures standing or sitting, the magnitude of this difference was very small. In MST, the percentages of significant cells were lower than in MT, except for the monkey condition; again, the absolute differences in average firing rate were very small.

We tested whether cells with a selectivity for random dot motion in either left or right direction had a preferences for
implied motion (humans and monkeys) and tilt direction (bars) in the same direction (Table 2 and Figure 5B, D, and F). A small number of cells exhibited a significant preference for implied motion but also tilt in the same direction as their preferred random dot motion direction, although again with only a small increase in spike rate. In addition, 11% (10/94) of all cells recorded in Experiment 1 were selective for upward or downward motion but were also significantly selective for implied motion human figures facing left or right (4 MT cells), in flickering human figures (1 MT cell and 1 MST cell), in monkey figures (2 MT cells), or for tilt direction of bars figures (2 MT cells). These cells had a significant preference for implied motion in a specific direction, which could not be explained by the preferred direction for real motion.

We further examined whether MT neurons that were selective for tilted bars versus vertical bars preferred implied motion versus standing/sitting images by comparing the ROC values (Figure 6). Units with a preference for implied motion often preferred tilted bars, as is indicated by the relatively high number of cells (23) in the upper right quadrant, compared with the lower right quadrant (8) in Figure 6A. No trend was visible for monkeys with implied motion versus bar figures (Figure 6B). For human versus monkey figures implying motion, units with a significant preference for human figures implying motion tended to have a significant preference for monkey figures without implied motion, although units with a significant preference for either monkey or human figures were much more scattered (see Figure 6C). Indeed, regression analysis for these data revealed a significant linear correlation between ROC values for human figures and bar stimuli \((R = .64, ANOVA p \text{ value } = .000)\), but not for human and monkeys or bars and monkeys. Together, these results indicate that preferences for human figures implying motion versus standing might be caused by the same low-level stimulus features that were responsible for the tilted bar preference but that may be lacking in the monkey figures.

**Table 3.** MT Cell Numbers in Experiment 2 and the Modulation of Their Response to Figures Implying and Not Implying Motion or Bar Tilt Are Shown by the Average Firing Rates and the Average ROC Values for the Whole Population and for Cells with a Significant Selectivity for Particular Conditions as Established in the ROC Analysis

<table>
<thead>
<tr>
<th>Figures</th>
<th>Figures</th>
<th>No. Cells</th>
<th>Ratio of Average Firing Rate</th>
<th>Average ROC Values</th>
<th>% Significant Cells for Preferred vs. Nonpreferred</th>
<th>Average ROC Value for Significant Cells</th>
<th>Ratio of Average Firing Rate</th>
<th>Implied Motion/Standing–Sitting or Preferred/Nonpreferred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied motion vs. standing/sitting</td>
<td>Humans</td>
<td>67</td>
<td>0.99 ± 0.07</td>
<td>0.49 ± 0.05</td>
<td>10</td>
<td>0.46 ± 0.12</td>
<td>0.93 ± 0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monkeys</td>
<td>44</td>
<td>1.01 ± 0.12</td>
<td>0.49 ± 0.05</td>
<td>9</td>
<td>0.40 ± 0.16</td>
<td>0.97 ± 0.13</td>
<td></td>
</tr>
<tr>
<td>Standing/sitting (face direction)</td>
<td>Humans</td>
<td>38</td>
<td>1.02 ± 0.11</td>
<td>0.51 ± 0.08</td>
<td>5</td>
<td>0.70 ± 0.03</td>
<td>1.05 ± 0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monkeys</td>
<td>29</td>
<td>1.01 ± 0.23</td>
<td>0.51 ± 0.12</td>
<td>10</td>
<td>0.58 ± 0.30</td>
<td>1.54 ± 0.46</td>
<td></td>
</tr>
<tr>
<td>Implied motion</td>
<td>Humans</td>
<td>38</td>
<td>1.00 ± 0.08</td>
<td>0.48 ± 0.06</td>
<td>5</td>
<td>0.36 ± 0.01</td>
<td>0.90 ± 0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monkeys</td>
<td>29</td>
<td>1.03 ± 0.24</td>
<td>0.51 ± 0.12</td>
<td>10</td>
<td>0.59 ± 0.27</td>
<td>1.45 ± 0.54</td>
<td></td>
</tr>
</tbody>
</table>

Variance is indicated with standard deviation.

**Experiment 2: Can MT Responses to Dynamic Noise Patterns Be Modulated by Implied Motion?**

We recorded the responses from 69 MT units (43 single units and 26 multiunits) to dynamic noise in the receptive field when static figures were presented at the fovea. We plotted the responses to figures implying motion and figures standing/sitting (Figure 7). We found no significant differences between average responses (Wilcoxon \(p \text{ values } > .05\)) to dynamic noise when foveal figures implied motion or foveal figures were standing/sitting (Figure 7A and B) or figures implied motion while facing the preferred motion direction versus facing the nonpreferred motion direction (Figure 7E and F).

ROC values were calculated for all units for figures implying motion versus standing/sitting figures (top rows in Table 3, filled symbols in Figure 7). Roughly 10% of MT cells showed significant ROC values, which could be favorable for either figure type, but with an average preference for the standing/sitting figures (i.e., mean ROC value below 0.5). To quantify what these ROC values mean in terms of firing rate, we calculated the ratio of average response rates to figures with versus without motion. For human figures, this ratio was 0.95, which corresponded to a difference in spike rate of 0.17 spikes per second in favor of the sitting/standing figures. For monkey pictures, the ratio was 0.97, which corresponded to a difference of 0.03 spikes per second in favor of the figures implying motion.
Cells showing preferential responses to horizontal motion were also tested for the difference in response to implied motion figures facing in the preferred and nonpreferred motion direction (bottom rows in Table 3, Figure 7E and F). Few cells had a significant preference for implied motion, and the preferred direction was not consistent over cells. The same result was obtained for comparison of standing/sitting figures facing in the preferred or nonpreferred direction for real motion (middle rows in Table 3, Figure 7C and D).

Human fMRI Experiment

Robust MT+ activation was detected in seven subjects during the motion-mapping task. For human as well as abstract pictures, there was more activation in MT+ when viewing implied motion pictures than when viewing pictures without implied motion (see Figure 8): human pictures, \( t(6) = 4.65, p < .005 \); abstract pictures, \( t(5) = 4.34, p < .01 \). The comparison of activation in MT+ between pictures of monkeys implying and not implying motion revealed no differences, \( t(6) = 1.218, p > .05 \).

A voxel-based comparison between images of humans and abstract figures implying motion and stationary figures revealed increased activation during observation of implied motion stimuli in several areas, including Brodmann areas 17 and 18 in early visual cortex, MT+ in the MT/occipital region, and parts of the inferior and superior parietal lobe. Differences in activation between images of monkeys implying motion and stationary monkeys were nearly entirely restricted to areas in the parietal lobe (Figure 9; Talairach x, y, and z coordinates: 42, −79, 22 for the left parietal lobe and −30, −79, 91 for the right parietal lobe). The difference in signals to images of humans implying motion and stationary humans in this area was not significant, \( t(7) = 0.81, p > .05 \) (Figure 10); however, the difference in signals to tilted and vertical bars was significant, \( t(6) = 2.82, p < .05 \).

DISCUSSION

In this study, we measured implied motion processing in cortical areas MT and MST of macaque monkeys and in human MT+. Modulatory input about the form of a moving (in)animate object could, in principle, be integrated in neurons processing real motion. This type of integration of different modalities could be formed by association (Schlack & Albright, 2007), as brief glimpses of moving subjects are often accompanied by actual motion.

There are several reasons why modulatory effects in areas MT and MST can be expected. First, previous human fMRI

Figure 8. Average signal in MT+ during the six conditions relative to passive fixation. Bars indicate standard errors, and asterisks show significant differences \( (p < .05) \).

Figure 9. Results of the contrast between implied and motionless stimuli for monkey, human, and abstract pictures. Significant voxels were thresholded at \( p < .05 \) (corrected) and superimposed on the averaged anatomical scan. MNI z-coordinates are displayed in white on the top left of each slice.
experiments have shown significantly different BOLD activations in the MT+ when comparing responses to images of humans, animals, or objects moving versus images of humans, animals, or objects standing still (Kourtzi & Kanwisher, 2000; Senior et al., 2000). Second, single-cell recordings in monkeys have shown neurons in the STS respond selectively to different degrees of implied motion (Barraclough et al., 2006; Jellema & Perrett, 2006) and could provide input to MT. Third, responses evoked by implied motion in a human EEG study could be modulated by pre-adaptation with real motion, which indicates the presence of neurons that are (directionally) sensitive to both types of motion (Lorteije et al., 2007). Neurons in MT and MST are adaptable to real motion and typically directionally sensitive and thus may be responsible for this adaptation effect. Although TMS over MT/MST fails to impair judgments of implied motion within 150 msec of stimulus presentation (Alford, van Donkelaar, Dassonville, & Marrocco, 2007), in an earlier TMS study, MT/MST has shown to be functionally necessary for representational momentum processing (Senior, Ward, & David, 2002). Fourth, modulatory input to MT and MST has been shown in many macaque single-cell studies, for example, modulation by working memory (Schlack & Albright, 2007; Zaksas & Pasternak, 2005; Bisley, Zaksas, Droll, & Pasternak, 2004) and attention (Maunsell & Treue, 2006; Cook & Maunsell, 2002, 2004; Martínez-Trujillo & Treue, 2002; Recanzone & Wurtz, 2000; Treue & Martínez-Trujillo, 1999; Treue & Maunsell, 1996, 1999), and thus any implied motion signal might also modulate MT neuronal activity. In addition, motion implied by Glass patterns has been shown to evoke activity in both human and monkey MT and MST (Krekelberg, Vatakis, & Kourtzi, 2005; Krekelberg, Dannenberg, Hoffmann, Bremmer, & Ross, 2003).

In our study, however, we did not find evidence for implied motion processing in macaque areas MT and MST. At the population level, cells did not respond differently to human or monkey figures with and without implied motion. A quarter of the cells in the MT population (less in MST, but with a smaller sample size) showed a small but significantly different response to images of humans running and standing. Surprisingly, the cells that were significantly selective for humans implying motion also tended to have larger differences in responses to the vertical and tilted bar figures, suggesting that the effects that were found may be attributable to low-level orientation and size differences in the images. In addition, preferences for human and monkey implied motion were not consistent: Cells selective for humans implying motion appeared to prefer images of stationary monkeys. These results contrast human fMRI results (Kourtzi & Kanwisher, 2000) that show activation within area MT+ by implied motion expressed by both human and animal agents, and this suggests that the response to implied motion is not a response to the implied motion content of the images but could reflect responses to low-level visual features. Our fMRI results support this hypothesis.

For a long time, we have known that MT neurons can respond selectively to static bars (e.g., Albright, 1984). Our results show that in addition to a transient response to stimulus onset, MT neurons have highly significant sustained responses above spontaneous activity for the static images. As MST neurons did not show this sustained response, small eye movements causing motion on the retina probably do not explain this phenomenon. In an fMRI study in monkeys, it has been shown that area MT (but not MST) has a differential response to static images compared with fixation only (Nelissen, Vanduffel, & Orban, 2006). It is thus surprising that we do not find large differences in responses to the different images. The small differences that we find can probably be attributed to low-level visual attributes like orientation and position.

In the second experiment, we found a small percentage (about 10%) of cells whose dynamic dot pattern response could be selectively modulated by static test images at the fovea. The change in firing rate, however, was so small that we question whether an implied motion signal is fed back onto area MT from position and size-invariant STSa neurons coding implied motion. These weak activations and low cell numbers can probably not explain the BOLD activity in MT+ found with human fMRI studies (Kourtzi & Kanwisher, 2000; Senior et al., 2000). Furthermore, the few cells showing differentially sensitive responses had ROC values below 0.5, indicating a preference for images without implied motion rather than images implying motion.

Our human fMRI experiment also strongly indicated that low-level differences in the images (partly) causes differences in responses to images with versus without implied motion because the bar stimuli gave an even stronger response difference than the human stimuli. More importantly, there was no difference in MT+ activity when observing the large set of images of monkeys implying motion and not implying motion. These images all differed considerably in low-level visual features that, on average, there is no consistent visual difference between pictures with and
without implied motion, except the presence of implied motion itself.

A direct comparison between human fMRI experiments and single-unit recordings in the monkey is difficult because it has been shown that BOLD signal correlates more with local field potentials (incoming signals) than direct neural activation (Logothetis & Pfeuffer, 2004). Although human homologues of macaque areas MT and MST have been proposed (Goossens, Dukelow, Menon, Vilis, & van den Berg, 2006; Huk, Dougherty, & Heeger, 2002; Dukelow et al., 2001; Peusksens, Sunaert, Dupont, Van Hecke, & Orban, 2001; Morrone et al., 2000), there are also several clear differences in the function and location of motion and object-sensitive cortical areas in humans and monkeys (Nelissen et al., 2006; Orban et al., 2003; Vanduffel et al., 2001). Evidence from fMRI in macaque monkeys suggests that combining object and motion information may mainly occur in areas fundus of the superior temporal region (FST) and a newly defined area lower superior temporal region (LST) (Nelissen et al., 2006). FST and LST would be strong candidate areas for processing implied motion. Because in the monkey these areas are located near area MT, in the human they could be part of MT and its satellites, where implied motion activation was demonstrated (Kourtzi & Kanwisher, 2000; Senior et al., 2000).

STSa contains neurons sensitive to animate implied motion (Barraclough et al., 2006; Jellema & Perrett, 2006). Many cells in this area will also respond to moving random dots (Barraclough et al., unpublished observations). The properties of STSa cells, however, are very diverse and complex; about 60% of those cells that respond to static images of human figures are sensitive to the degree of articulation shown by the human. Of STSa cells sensitive to images of human figures, half prefer implied motion, whereas the other half prefer standing/sitting images (Barraclough et al., 2006). On the basis of these results, one would not necessarily expect a differential response in the human homologue of STSa for implied motion versus standing/sitting images in an fMRI experiment. Furthermore, STSa cell responses can depend on the view of the human or monkey figure. Although cells selective for static images of articulated bodies were more likely to respond to movies of bodies walking forward (Barraclough et al., 2006), typically there is no correlation between responses to implied motion and moving RDPs (Barraclough et al., unpublished observations).

Several arguments could be given as to why we were not able to show evidence for implied motion activity in areas MT and MST. First, although our search stimulus consisted of both moving RDPs and the static images that we used in this study, we cannot completely exclude the possibility that our sample was biased because of search strategies. Also, we may not have sampled neurons from every subregion of MT and MST. However, we searched extensively for implied motion-responsive neurons because we varied recording position and depth to cover different locations within MT and MST, and we recorded in two different animals and for one animal in both hemispheres. Second, it could be argued that our images did not convey a strong signal for implied motion that monkeys were not able to recognize or interpret these images or that the monkeys were not paying attention to the images because their task was only to fixate the fixation dot. These suggestions are unlikely as similar images during the same fixation task can evoke implied motion-selective responses within the STSs of macaques (Barraclough et al., 2006). Furthermore, the same images have been shown to elicit differential responses in human-evoked potential recordings (Lorteije et al., 2006) arising from direction-selective motion-sensitive neurons in dorsal cortex (Lorteije et al., 2007), and the images are very similar to the type of images that were used in the human fMRI study of Kourtzi and Kanwisher (2000). Furthermore, there is a plethora of studies showing that Rhesus macaque monkeys can interpret the content of images depicted on a computer screen as observed from their behavior, and many studies have shown differential neural activation in object recognition areas for different images in monkey fMRI or single-unit recording studies (Kourtzi, Krekelberg, & van Wezel, 2008).

Our results show that MT cells may respond selectively to low-level stimulus features within images with and without implied motion. However, although STSa cells have been shown to respond strongly to the implied motion content within an image, MT cells (and we have shown evidence indicating that this may hold for MST cells too) are insensitive to motion implied by biological forms.

Acknowledgments

The authors thank Bert van den Berg and Wim van de Grind for their contributions to this project. The project was supported by grants awarded to R. v. W.: Innovational Research Incentives Scheme (VIDI) of the Netherlands Organization for Scientific Research (NWO), the Interuniversity Attraction Poles Programme (IUAP) of the Belgian Science Policy and a High Potential grant of the Utrecht University (UU). This work was further supported by grants from the European Union and the Wellcome Trust awarded to D. P. and M. O.

Reprint requests should be sent to Richard J. A. van Wezel, Division of Pharmacology, Utrecht University, Sorbonnelaan 16, 3584 CL Utrecht, The Netherlands, or via e-mail: r.j.a.vanwezel@uu.nl.

REFERENCES


