INTRODUCTION

Animals’ ability to correctly decode visual motion has likely been subject to significant evolutionary pressure. Tasks such as image segmentation, deducing structure from motion, judging time to collision, and recovering three-dimensional (3D) structure as well as goal-directed behavior are critically dependent on the appropriate analysis of visual motion signals (see Nakayama 1985 for an overview). Visual motion can be related to ego motion of the subject or, alternatively, can be produced by the movement of an object in the external world. Here, we will address aspects of the neuronal processing underlying the perception of a moving object.

It is well established that visual-motion processing in primates is intimately related to the neuronal activity observed in the middle temporal (MT) and middle superior temporal (MST) areas (for review, see Albright 1993). It was shown that activity recorded from areas MT and MST is closely related to the perception of visual motion. Neurons in area MT respond to stochastic motion signals, and the strength of the responses increase with the number of coherently moving dots (Britten et al. 1993). Using such stochastic motion signals, a clear relationship between the monkey’s behavioral choices and visual responses of area MT was found within single trials (Britten et al. 1996). For area MST, similar neuronal and psychophysical sensitivities were documented (Cebalbrini and Newsom 1994). Microstimulation in area MST affects the motion perception of monkeys (Celebrini and Newsom 1995) as well as the judgments of heading directions (Britten and van Wenzel 1998). Neurons recorded from the posterior polysensory part of the superior temporal sulcus (STPp) respond to moving stimuli in accordance with the monkeys’ perception of motion (Thiele and Hoffmann 1996). More recently, it was shown that microstimulation in area MT affects both the monkey’s choices and the speed of decision-making, i.e., monkeys decided more quickly in favor of the stimulation site’s preferred direction (Ditterich et al. 2003). Lesions in area MT and MST result in an impairment of motion perception (Newsome and Paré 1988; Rudolph and Pasternak 1999). Finally, using a glass pattern (Glass 1969), it was shown that response properties in primate area MT and MST are similar to the properties of perception in humans and monkeys (Krekelberg et al. 2003).

Neuronal activity in areas MT and MST was also shown to be related to the execution of smooth-pursuit eye movements (SPEM) (Ilg and Thier 2003; Kawano et al. 1994; Newsome et al. 1988; Thier and Erickson 1992). Small lesions of these areas yield an ipsiversive deficit in SPEM (Dürstler and Wurtz 1988; Yamasaki et al. 1991). Intracortical microstimulation is able to modify ongoing SPEM (Born et al. 2000; Groth et al. 1997; Komatsu and Wurtz 1989).

In previous studies, the motion signal has been created by luminance-defined moving objects. However, a moving object need not be separated from the background by luminance or color. It can be defined by coherent moving dots [Fourier motion (fm)] (Cavanagh and Mather 1989), by changes in luminance [drift-balanced motion (dbm)] (Chubb and Sperling 1988), or by opposed motion [theta motion (tm)] (Zanker 1993). Note that the dbm stimulus is segregated from the background by flicker, whereas the tm stimulus has the same amount of flicker as the background. These two motion stimuli are referred to as second-order motion as opposed to first-order motion such as Fourier motion (Cavanagh and Mather 1989). Previously, it was assumed that neurons in area MT might code for the direction of a moving stimulus invariant of the stimulus’ visual properties (Albright 1992; O’Keefe and Movshon 1998).

Most real-world moving objects provide not only visual motion cues but also auditory motion information. In the laboratory, the perception of a moving sound source can be produced in different ways: by binaural beat, by changing interaural time or intensity differences, by artificial auditory surround using individual head-related transfer functions, or by an actively moved loudspeaker. The area activated in the human brain depends on the kind of stimulus used (see Warren...
et al. 2002 for a overview). Here, we opted for apparent auditory motion generated by the sequential activation of a single loudspeaker within a linear array of loudspeakers. It might be quite reasonable to assume that neurons in area MT and MST respond to a moving sound source, especially if the monkey has to report the direction of this movement. This idea is supported by cortical connectivity: there are direct anatomical connections from auditory cortical areas to area MST (Boussaoud et al. 1990). In addition, the primary visual cortex receives projections from the core and parabelt areas of the auditory cortex (Falchier et al. 2002).

The aim of the present study was twofold: first, we tried to prove that rhesus monkeys are able to correctly detect the direction of a stimulus defined either by second-order motion or by a moving sound source. Second, we asked whether the activity of individual neurons recorded from areas MT and MST code the movement of the stimuli in a directionally selective manner independent of the specific parameters of the moving object.

METHODS

While our highly trained rhesus monkeys performed a direction discrimination task as explained in the following text, we recorded single-unit activity from areas MT and MST.

Animal preparation

Three adult male rhesus monkeys (F, B, and G) were used in this study. Under sterile conditions and intubation anesthesia using isoflurane, the monkeys received a dental cement implant including head post and recording chambers as well as a subconjunctival search coil to monitor precisely the eye movements (Judge et al. 1980; Robinson 1963). All animal procedures were carried out in accordance with National Institutes of Health guidelines and German law and were approved by the local ethics committee. The position of the recording chambers allowed for electrode tracks leading to areas MT and MST. The chamber axis was aligned with a parasagittal plane, tilted 30° upward from the horizontal plane. This axis was aimed at the stereotactic location of area MST (lateral 18, posterior 3.5, and dorsal 16 mm). Single-unit activity was recorded with self-made glass-insulated tungsten electrodes the high stability and stiffness of which allowed for transdural electrode tracks. The signal from the electrode was preamplified, low-pass filtered, and fed to a multiwire detector (Alpha Omega, Model MSD). The temporal resolution of single-unit activity was 4 kHz, and horizontal as well as vertical eye positions were sampled at 1 kHz per channel. All data acquisition was performed by an i586 PC (LINUX). Eye-position signals were calibrated by having the monkeys fixated targets at known positions.

Passive visual stimulation

To determine the visual response properties of an individual neuron, such as the location and size of the receptive field, preferred direction, preferred velocity, tuning width, and strength of directional selectivity, we displayed a moving random-dot pattern within a stationary aperture while the monkeys simply fixated a stationary target. The preferred direction of this passive visual response was determined by testing the responses to eight different directions (with 45° spacing). The responses to visual motion in these directions were fitted by a Gaussian function. The mean value of this function indicates the preferred direction of the recorded neuron and the SD gives its tuning width. The directional index (DI) was computed according Eq. 1 (see following text). The location, size, and eccentricity of the visual receptive field were determined by the investigator using a mouse-driven bar adjusted to the parameters of the individual neuron.

Direction discrimination task (DDT)

At the onset of each trial, the monkeys fixated a small red stationary target (diameter: 20 arc min, 0.5 cd/m²) placed in the center of the tangent screen (size: 86° × 66°, viewing distance: 85.5 cm). All visual stimuli were generated by a second PC (i586, LINUX) and displayed by a video projector (Electrohome ECP 4100, 1200*1024 Hz at 60 Hz). Gaze fixation was monitored throughout the trial. The size of the fixation control window was 2° in each dimension. If the deviation of gaze exceeded the control window, the trial was aborted instantaneously, the data were discarded, and the monkey was not rewarded for the trial. After the initial fixation period of 500 ms, a randomly selected motion stimulus as explained in the following text was presented for 2,000 ms. The stimulus moved either in the preferred or nonpreferred direction of the recorded neuron (i.e., left- or rightward). The preference was determined beforehand during the passive visual stimulation. The stimulus trajectory crossed the center of the recorded neuron’s receptive field halfway through its presentation. The velocity was adjusted to the previously determined preferred velocity of the recorded neuron. After the presentation of the stimulus, the monkeys had to maintain fixation (delay of 500 ms) until two green targets (diameter of 20 arc min) were presented 20° to the left and right of the central fixation target. To report the direction of the perceived movement, the monkeys had to perform a saccade toward the target in the direction of the movement. The monkeys had to execute this saccade within an interval of 3,000 ms to complete the task correctly and to receive the drop of apple juice or water for reward.

Motion stimuli

Three different motion stimuli were generated by the stimulation PC and presented during the performance of the DDT. Each stimulus consisted of a rectangular object the size of which was adjusted to the size of the receptive field of the recorded neurons. In the case of the Fourier motion stimulus, the rectangle was an unchanging pattern of coherently moving dots as indicated in space-time diagram of Fig. 1 by the diagonal lines. In the case of the drift-balanced stimulus, the rectangle can be imagined as a moving window through which a second distant plane of stationary dots was seen. Here the moving stimulus is indicated in the space-time diagram as short vertical lines, i.e., stationary dots. Finally, in the case of theta motion, the rectangle consisted of coherently moving dots. However, as opposed to the Fourier motion stimulus, the dots are moving in the opposite direction

FIG. 1. Sketch of space-time diagrams of the motion stimuli used in this study: Fourier motion (fm), drift balanced motion (dbm), and theta motion (tm). Each row in the diagrams represents the actual distribution of black-and-white pixels in a single row on the simulation monitor. The vertical spacing represents the refresh rate of the graphic card (60 Hz).
with respect to the entire rectangle. In the space-time diagram of Fig. 1, the dots are moving toward the left while the rectangle is moving rightward. The stimuli were presented on top of a dynamic random-dot pattern (2% coverage, maximal 500 dots, mean luminance 1.9 cd/m²). Figure 1 shows sketches of the space-time diagrams of these three different motion stimuli.

In addition to the visual stimuli displayed by the video projector, apparent visual, auditory, or bimodal motion stimuli were also presented over the course of the DDT. We had 48 light-emitting diodes (LEDs) and loudspeakers mounted on the tangent screen in a linear arrangement with a spacing of 1° at a distance of 85.5 cm in front of the animal. The LEDs and/or loudspeakers were subsequently switched on for 25 ms and off for 25 ms by a digital IO card in a i286 PC (DOS) that was synchronized with the data-acquisition PC. The resulting apparent velocity of the stimuli was 20°/s in these experiments. The brightness of the LEDs was 5 cd/m². The auditory stimuli were band-pass noise pulses (–10dB bandwidth: 1.7–3.4 kHz) presented at 60 dB SPL. Frequency response curves of the 48 loudspeakers used were measured with a Bruel & Kjaer phone system and analyzed with Cool Edit (Syntrillium). The frequency responses deviated by <5 dB and overall sound pressure level deviated by <2 dB between speakers. The trajectory of these stimuli was centered on the visual receptive field of the recorded neuron. Human observers perceived a smoothly moving stimulus in every case.

Training of the animals

Before the monkeys underwent the aforementioned fixation surgery, we trained them in the primate chair on a standard fixation paradigm using a single-touch bar (Wurtz 1969). After ~4 wk, the monkeys were subjected to a simple version of the DDT. To start a single trial, the monkeys had to hold the central touch bar. After a randomized delay, a bright bar moved either to the left or right. On color change of the fixation target, the monkeys had to move their hand to either the left or right touch bar, respectively. Over the course of these training lessons, the monkeys learned to report the direction of bar movement independent of the bar position at the onset of each trial. The monkeys could not use the bar position at the offset of each trial to complete the task correctly. Once the monkeys’ performance stabilized (>80% correct trials, they underwent surgery. Subsequently, the DDT paradigm was changed to its final design. Instead of using three touch bars, the monkeys reported their perception by making saccades. In addition, the stimuli used during the recording sessions were displayed. Overall, the training on the DDT lasted ~3 mo. Note that during the recording sessions, the position of the receptive field of the recorded neuron dictated the trajectory of the stimulus.

Data analysis

The neuronal responses were obtained as the number of action potentials within a time interval during which the stimulus crossed the receptive field. The dimensions of the visual receptive field had previously been determined as explained earlier. Using this information, the time interval for data analysis was adjusted for each individual recorded neuron. Based on these responses, we calculated the strength of directional tuning as

\[ \text{direction index DI} = 1 - \frac{\text{response non-preferred}}{\text{response preferred}} \]

In the case of the tm stimulus, the neuronal response apparently inverts its preferred direction. To document this effect, we calculated the DI as explained in the preceding text, i.e., we followed the apparent inversion of the preferred direction, but set the DI to be negative.

To quantify possible inhibitory effects, we also calculated the net responses for which we subtracted the activity recorded during the initial fixation period of each trial from the above mentioned response. To address the statistical significance of the responses to the different stimuli or different directions of stimulus movement, we performed t-test on the responses or net responses, respectively, of single trials. For the purpose of displaying the neuronal activity, we calculated and plotted spike density functions (sigma = 25 ms).

Pursuit study

To strengthen the argument that our monkeys indeed report the perceived direction of a moving target, we performed an additional pursuit eye-movement study similar to previously published studies of human eye movements (Butzer et al. 1997; Lindner and Ilg 2000). The pursuit of the animal was centered on the visual receptive field of the recorded neuron. The stimuli or different directions of stimulus movement, we performed t-test on the responses or net responses, respectively, of single trials. For the purpose of displaying the neuronal activity, we calculated and plotted spike density functions (sigma = 25 ms).

FIG. 2. Recording sites in monkey F. Four parasagittal sections are shown based on drawings of Nissl-stained sections. Bold line, the surface of the cortex; thin line, the transition between gray and white matter. The reconstruction of MT area along the posterior bank of STS is based on Gallyas-stained sections. Note that the recordings of middle superior temporal (MST) neurons were performed along the floor of the STS.
stimuli were identical to those used in the direction discrimination task except that we did not display the stationary fixation target. Instead, the eye-position control window moved with the stimulus. After a training period of several weeks, monkey B was able to pursue the moving target. We quantified the initial acceleration of presaccadic pursuit initiation as described previously (Lindner and Ilg 2000). In addition, we determined steady-state pursuit velocity within a fixed time interval extending from 500 to 1,200 ms after the onset of target movement. The saccades were removed automatically based on an acceleration criterion (see Lindner and Ilg 2000).

**Histology**

After perfusion of one (F) of the animals, the brain was cut parasagittally in 40-μm sections. The sections were stained for cell bodies (Nissl) and for myelination (Gallyas). Area MT in the posterior bank of superior temporal sulcus (STS) was identified by the dense myelination visible in the Gallyas staining. The reconstruction of recording sites was made possible by injections of fluorescent tracers at the end of the recording period into sites that were previously identified as areas MT or MST, respectively, based on their single-unit response properties. The histology of the two remaining monkeys (B and G) is not available at present.

**RESULTS**

While our monkeys performed the DDT, we recorded the activity of 38 neurons located in area MT and 68 neurons located in area MST. We first present the recording sites and the general visual response properties of our neuron sample, then we present the psychophysical results of the DDT, and finally we describe the single-unit activities obtained.

**Properties of single-unit responses during passive visual stimulation**

The microelectrode approached areas MT and MST within a parasagittal plane, entering the brain in the lunate sulcus, at the transition between areas V1 and V2. The neurons recorded from area MST were concentrated along the floor of the superior temporal sulcus (STS) where the lateral part of area MST (MST-l) is located. The anatomical reconstruction of recording sites shown in Fig. 2 was performed in one (F) of the three monkeys. In the remaining two animals (B and G), the classification of area MT and MST was performed based on the specific single-unit response properties observed during the experiments as well as the 3D positions of the electrode tip. The anatomical reconstruction of monkey F clearly overlaps with the classification of areas MT and MST based on single-unit response properties and the 3D position of the electrode tip.

Because of methodological limitations, we were restricted to a horizontal stimulus trajectory during execution of the DDT. Therefore our data sample shows a clear bimodal bias toward horizontal (see Fig. 3A). The histogram of the tuning widths reveals a clear maximum at 90° (see Fig. 3B). Most neurons in our sample are highly directionally selective irrespective of whether they were recorded from area MT or MST (see Fig. 3C). The only significant difference between the responses obtained from area MT and MST neurons was the receptive fields’ size (see Fig. 3D). In area MST, the receptive fields were slightly but significantly (P = 0.005, t-test) larger than those of neurons recorded from area MT.

**Monkey psychophysics**

The results of the DDT are shown in Fig. 4. The performances for all stimuli used were significantly above chance level (significantly different from 50%, t-test, P < 0.003). The statistical analysis (2-way ANOVA) revealed a significant influence of the factor stimulus (P < 0.001) on the performance. However, there was no significant difference in the performance of our monkeys (factor monkey: P = 0.138). Because the interaction of both factors was also not significant.

![Figure 3](https://example.com/image3.png)

**FIG. 3.** General visual response properties of all recorded neurons from area MT and MST in this study. The clear bias in preferred direction toward horizontal orientations (shown in A) was due to a methodological limitation that constrained stimulus movement to horizontal directions. The distribution of tuning widths showed a clear maximum around 90° (B). All neurons were directionally selective (C). The white bars in B and C, the distribution of directionality of neurons recorded from area MT, exclusively. Note the similarity between the distribution of all recorded neurons and those recorded from area MT. Only the size of receptive fields in area MST (shown in D) are slightly but significantly (P = 0.005, t-test) larger than the receptive field size observed in area MT.
(\(P = 0.732\)), we can conclude that the differences between the stimuli were very similar in all monkeys.

**Smooth pursuit eye movement elicited by second-order motion**

It is possible to argue that our monkeys did not report the direction of the moving stimuli in the DDT but rather reported the position of the stimuli at the offset of the stimulus presentation. To check this hypothesis, we asked whether a monkey is able to track the moving second-order motion stimulus by smooth-pursuit eye movements and whether the way in which this monkey tracks this stimulus is similar to the way in which humans do it (Butzer et al. 1997; Lindner and Ilg 2000).

The monkey was able to pursue the theta motion and drift-balanced stimuli as well as the Fourier motion stimulus, as shown in Fig. 5. However, steady-state pursuit gain was less
in case of the paradoxical motion stimulus than for the Fourier motion stimulus. In addition, we analyzed the initiation of SPEM. The initial eye acceleration depended on the motion stimulus being tracked. Fourier motion elicited higher acceleration values \((-39^\circ/s^2)\) compared with drift-balanced motion \((-14^\circ/s^2)\). In the case of theta motion, the initial acceleration occurred in the opposite direction (i.e., in the direction of the moving dots), denoted by the positive sign of the acceleration value \((36^\circ/s^2)\).

Taken together, the results of this pursuit study support the notion that monkeys reported the direction of motion in our task not simply the position of the stimulus at the offset. In addition, the monkey's eye movements were quite similar to human eye movements elicited by these stimuli (Butzer et al. 1997; Lindner and Ilg 2000).

**Single-unit responses during performance of DDT**

As already mentioned, we focused our study on directionally selective neurons, which were tuned for horizontal directions. Figure 6 shows the response of a typical neuron recorded from

*FIG. 6. Response of a typical MST neuron (No. 110/1, monkey G) to Fourier, drift-balanced, and theta motion stimuli moving leftward (left) and rightward (right). Target and eye position are shown together with the neuronal activity as raster display and as a spike density function. Leftward motion is indicated by negative numbers. The dotted vertical lines indicate the time when the stimulus entered the receptive field and left it, respectively.*
area MST to the presentation of Fourier, drift-balanced, and theta motion stimuli in preferred and nonpreferred directions. The eye-position profiles documented that the monkey maintained fixation during stimulus presentation and subsequently reported the direction of the moving stimulus correctly.

The response of the presented neuron to Fourier motion was clearly directionally selective. The horizontal dimension of the receptive field were from 16° left (contralateral) to 4° right (ipsilateral) as indicated by the vertical dashed lines in Fig. 6. We determined the neuronal response during a time interval of 1,000 ms while the stimulus was moving across the receptive field. The response in the preferred direction (74 spikes/s) was significantly higher than in the nonpreferred direction (11 spikes/s) resulting in a DI of 0.85. The directionality of this neuron was slightly diminished, but its response to drift-balanced motion stimuli was significant. Here the response in the preferred direction was only 31 spikes/s, whereas the response in the nonpreferred direction was 5 spikes/s, resulting in a DI of 0.83. During the presentation of the theta motion stimulus, the directionality was apparently inverted. The response to the theta motion stimulus in the nonpreferred direction (58 spikes/s) was larger than the response in the preferred direction (13 spikes/s), yielding a DI of −0.77. The response in the preferred direction disappeared in the case of theta motion, whereas the response in nonpreferred direction was increased. This apparent inversion indicates that the neuron’s response was determined by the direction of the individual pixel constituting the target not by the direction of the moving target itself. Not only was the preferred direction apparently inverted, but the absolute value of directionality was also reduced in the case of theta motion as compared with Fourier motion.

With respect to the responses of all recorded neurons (see Fig. 7), we observed weaker directionality in the responses to drift-balanced motion stimuli than in responses to Fourier motion stimuli. However, 36% of neurons from area MT (14 of 38) and 53% of neurons from MST (36 of 68) responded significantly to drift-balanced stimuli (P < 0.01, t-test). We did not find significant differences between the response properties such as preferred direction, width of tuning, preferred velocity, strength of directional selectivity, receptive field size, or receptive field eccentricity of neurons responding significantly to drift-balanced motion and neurons lacking this selectivity (see Table 1). Therefore we are not able to deduce the ability of a neuron to code for drift-balanced motion from any other visual response property determined by the passive visual stimulation.

Based on the responses shown in Fig. 6, one might argue that the responses to the theta motion stimulus were exclusively determined by retinal image motion (i.e., by the movement of individual dots). However, as Fig. 8A shows, the absolute value of theta-motion-related directionality was significantly less than the Fourier motion related directionality (t-test, P < 0.001). This indicates that although the direction of pixel motion determined the preferred direction of the neuronal response, the strength of directionality was different for theta and Fourier motion. Similar to the neuronal responses to Fourier and theta motion, the absolute values of eye acceleration during initiation of pursuit elicited by theta motion were significantly less than when pursuit was elicited by Fourier motion (t-test, P < 0.001) as shown in Fig. 8B. The consistency of neuronal activities and initial eye acceleration during SPEM indicates that areas MT and MST are involved in the generation of these eye movements.

We analyzed the responses to the auditory and bimodal moving stimulus only in those neurons the directionally selective response of which to the visual motion stimulus was significant (t-test preferred versus nonpreferred direction, P < 0.01). Figure 9 shows the response of a typical neuron recorded from area MT to visual, auditory, and bimodal moving stimuli during DDT. The receptive field of this neuron extended from 16° left (contralateral) to 12° right (ipsilateral), therefore the time interval used to determine the neuronal response was 1,400 ms. While the neuron clearly responded with an increase in discharge rate if the visual (net response of 22 spikes/s) or bimodal stimulus (net response of 20 spikes/s) moved across its receptive field, there was a small decrease in discharge rate if the auditory stimulus moved across the receptive field (net response of 2 spikes/s).

Overall, the responses of 33 neurons (6 of 28 neurons from area MT, 27 of 68 neurons from area MST) fulfilled the above-mentioned directionality criterion and were included in this analysis. Because we did not observe any significant differences between net responses recorded from area MT and MST, respectively, we pooled the responses of all 33 neurons.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Selective Neurons</th>
<th>Nonselective Neurons</th>
<th>P (t-Test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>58</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Deviation of preferred</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>direction from horizontal, °</td>
<td>34 ± 19</td>
<td>35 ± 25</td>
<td>0.90</td>
</tr>
<tr>
<td>Tuning width, °</td>
<td>79 ± 40</td>
<td>79 ± 41</td>
<td>0.95</td>
</tr>
<tr>
<td>Preferred velocity, %/s</td>
<td>23 ± 13</td>
<td>17 ± 12</td>
<td>0.07</td>
</tr>
<tr>
<td>DI</td>
<td>0.85 ± 0.13</td>
<td>0.85 ± 0.14</td>
<td>0.98</td>
</tr>
<tr>
<td>Size of rf, °</td>
<td>662 ± 830</td>
<td>421 ± 529</td>
<td>0.21</td>
</tr>
<tr>
<td>Eccentricity of rf, °</td>
<td>8.0 ± 5.8</td>
<td>6.6 ± 4.3</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Values are means ± SD. DI, directional index; rf, receptive field.
As Fig. 10 shows, the net responses to the visual and bimodal stimulus were not significantly different ($t$-test, $P = 0.83$). However, the inhibition induced by the moving auditory stimulus shown in Fig. 9 was not observed at the population level. The population response to the auditory stimulus was not significantly different from zero ($t$-test, $P = 0.70$).

**DISCUSSION**

The performance of our monkeys in the DDT strongly suggests that the animals were able to correctly report the direction of a moving stimulus, independent of whether the stimulus was defined by first- or second-order motion or by a moving sound source. The results of our pursuit study seem to indicate that the monkeys reported the motion direction not simply the position of the stimulus at least for the second-order motion stimulus. However, the activity in areas MT and MST coded only for those moving stimuli that were segregated from the background by luminance or flicker.

We are convinced that our data sample is quite representative for areas MT and MST because the general response properties of our recorded neurons to visual motion fit well with those previously described in the literature. The size of the receptive fields in areas MT and MST, respectively, were of the same dimension as described earlier (Komatsu and Wurtz 1988). Similarly, the mean tuning width of the neurons in our data sample was close to 90° as previously reported (Albright 1984; Britten and Newsome 1998; Snowden et al. 1992).

The absence of selective neuronal responses during perception of theta motion and moving sound sources might be surprising because it was previously shown that some neurons in area MST respond to inferred motion (Assad and Maunsell 1995). The neurons in their study responded to an inferred motion stimulus without any physical stimulus. However, it must be noted that the monkeys were trained to report the movement of a visual stimulus, so it is likely that they inferred the presence of the visual motion stimulus that they had been trained on. In contrast, there is no evidence to assume that our monkeys inferred visual motion in the presence of a moving sound source. Therefore the lack of response in our study does not contradict results reported during inferred motion perception (Assad and Maunsell 1995). However, the finding that a moving sound source did not lead to neuronal excitation does seem surprising because anatomical connections from the monkey's auditory cortical areas to primary visual cortex (Falchier et al. 2002) as well as to area MST (Boussaoud et al. 1990) have been previously described.

**Alternative interpretation of our results**

Our data show that the neuronal activity recorded from areas MT and MST did not always parallel the motion perception of the monkeys. From this, we conclude that these areas are not the final stages in motion processing. However, an alternative interpretation is possible based on the fact that in our study, the monkeys could theoretically perform the DDT correctly without processing the motion signals at all and simply using position information, i.e., reporting the position of the stimulus at the offset of every trial. The outcome of our experiments might have been different if we had used a task in which positional information could not have been used, for instance, if the monkeys had to respond to a change in stimulus velocity. However, there are three arguments against this alternative explanation. First the outcome of the pursuit study indicates that the monkey could use the velocity of the second-order motion stimuli to generate smooth-pursuit eye movements. Second, during the training history, the monkeys were able to report the direction of stimulus movements at randomized positions within the visual field. Finally, in case of visual motion, drift-balanced motion, and Fourier motion, the neuronal activity paralleled the motion perception of the monkeys, and it is unclear why the monkeys would change their strategy depending on the stimulus type. The alternative explanation cannot be completely ruled out, but it seems less likely that the lack of response to auditory stimulation as well as the apparent inversion of the preferred direction in the case of theta motion could be explained by the fact that specially in these conditions, the monkeys did only report the position of the stimulus, not its motion.

**Processing of first- and second-order motion**

When a flicker component segregated the moving stimulus from the background (dbm), some neurons in our sample coded for object motion, a result that has also been reported previously (Albright 1992; O'Keefe and Movshon 1998). As found in those experiments, the strength of directionality in response
to drift-balanced motion was clearly less than in response to first-order motion.

On the other hand, the apparent inversion of preferred direction in case of the theta motion indicates that the motion processing is not yet complete in areas MT and MST. It is important to note that this does not contradict findings that microstimulation in these areas affects motion perception (Britten and van Wezel 1998; Celebrini and Newsome 1995) as well as the execution of smooth pursuit eye movements (Born et al. 2000; Groh et al. 1997; Komatsu and Wurtz 1989). These effects convincingly demonstrate that these areas are involved in processing visual motion but do not support the notion that these areas are final stages in this processing. Because these areas are retinotopically organized, subsequent areas could be able to recover object motion based on the processing of the various arrangements of receptive fields in an appropriate temporal sequence. Along the same line of argument, a computational model consisting of two layers of elementary motion...
The similarity of eye acceleration during the initiation of SPEM and the neuronal responses to theta motion strongly supports the idea that areas MT and MST are definitively involved in the processing of first- and second-order motion. When a theta motion stimulus was used, the absolute value of eye acceleration as well as the directionality of neuronal responses to theta motion differed significantly from eye acceleration and directionality of responses when a Fourier motion stimulus was used even though the absolute amount of retinal motion was identical for both stimuli. This suggests that both eye acceleration during SPEM initiation and neuronal responses might not depend exclusively on the movement of individual pixels. It is possible that the processing of pixel motion (first-order) interacts with the processing of object motion (second-order), which would in turn predict a decreased response if the two components were in opposite directions. The similarity in eye acceleration and neuronal responses shows that the initiation of SPEM and neuronal processing in areas MT and MST are tightly connected.

Motion processing in higher cortical areas of rhesus monkeys

Taken together, the results clearly confirm previous findings that areas MT and MST are not the final stages in the cortical processing of a moving visual stimulus. Areas higher up in the hierarchy of visual motion processing such as the anterior and posterior part of the superior temporal polysensory area (STPa and STPp) (Felleman and van Essen 1991), area 7A (Siegel and Read 1997), the ventral intraparietal area (VIP) (Colby et al. 1993), and the cortical vestibular areas as the parietoinsular vestibular cortex (PIVC) (Grüsser et al. 1990) and area 2v (Büttner and Buettner 1978) may be involved.

With respect to multisensory processing of visual and auditory information, it has previously been shown that neurons recorded from the superior temporal sulcus (STS), especially from the anterior parts of the sulcus, responded to stimulation in both modalities (Benevento et al. 1977; Bruce et al. 1981). From the lateral intraparietal area (LIP), it was reported that neurons responded to an auditory stimulus only if this stimulus was used as a saccade target. This auditory response was only observable after training; in naïve monkeys, there was no response to auditory stimulation (Grunewald et al. 1999). Neurons recorded from the VIP displayed responses to somatosensory, vestibular, and visual stimulation (Bremmer et al. 2002; Duhamel et al. 1998). Finally, in the frontal cortex (area F4), multimodal signal processing was described in mirror neurons. These neurons discharge if the monkey either performs a specific action or watches another subject, either human or monkey, performing the same action (Rizzolatti et al. 1996). Some of these mirror neurons responded if the monkey heard the action (Kohler et al. 2002). Taken together, these results suggest that there is a widespread overlap of visual and auditory processing within the rhesus monkeys’ brain. This notion was further supported by a 2-deoxy-glucose study that showed widespread overlap between visual and auditory areas (Poremba et al. 2003): many areas within the superior temporal sulcus as well as in the frontal cortex showed responsiveness to auditory and visual stimulation.

Multimodal motion processing in the human brain

The rapid development of functional brain imaging capabilities has made it possible to compare the outcome of single-unit recordings in awake and behaving monkeys with activation studies of the human brain. The involvement of area V5 or human MT (hMT) in visual motion perception has been demonstrated (Beauchamp et al. 1997; Castelo-Branco et al. 2002). Even a moving tactile stimulus yielded an activation of hMT (Hagen et al. 2002). The activity in hMT as well as the perception of the motion aftereffect have been modified by an auditory attention task (Berman and Colby 2002). When human subjects had to perform a similar direction discrimination task with a moving visual stimulus, clear activation of hMT and surrounding areas, human homologues of areas MST and FST, was shown. However, when subjects had to report the movement of an auditory stimulus, a slight decrease in activation was observed in hMT (Lewis et al. 2000). This fMRI finding is very similar to the single-unit responses from area MST in monkeys reported here. In addition, it has been shown by PET and fMRI experiments that a bilateral posterior network including the planum temporale and parieto-temporal operculum is activated during the perception of sound-source motion (Warren et al. 2002). When subjects were exposed to second-order motion stimuli, areas such as V3 and hMT were activated (Smith et al. 1998). However, the signals detected by fMRI do not allow retinal image motion and object motion to be differentiated. This result is consistent with our finding that the neurons are driven by tm stimuli but do not code for the direction of the object. Furthermore, when subjects were exposed to visual, auditory, and tactile motion stimuli, activation was found in the intraparietal sulcus, which is most likely the human homologue of primate area VIP (Bremmer et al. 2001). Interestingly, these authors did not report polymodal activation of hMT, which has been reported by others (Lewis et al. 2000). Thus there is a high degree of overlap between fMRI studies in humans and the single-unit responses from rhesus monkeys presented here, especially with respect to the responses of hMT and surrounding areas.

In summary, there is clear evidence that areas MT and MST are involved in the processing of a moving stimulus that

**FIG. 10.** The net responses of the population (n = 33) to a visual, auditory, or bimodal moving stimulus. The net responses elicited by the visual and bimodal stimulus are highly significant from 0 (P < 0.001), whereas the responses due to the auditory stimulus were not (P = 0.7, t-test).
underlies motion perception, but there is also clear evidence that this processing is not completed by these areas.

**Acknowledgments**

Thanks to C. Cavada for histology, T. Cooke for language assistance, J. Ostwald for measuring the spectral properties of the loudspeakers used, and P. Thier for innumerable discussions on various aspects of this work.

**Grants**

This project was supported by Deutsche Forschungsgemeinschaft (Heisenberg and SFB 550, A3) as well as the Hermann and Lilly Schilling Foundation.

**References**


