Interaction of Retinal Image and Eye Velocity in Motion Perception

Herbert C. Goltz,[1] Joseph F.X. DeSouza,[1]
Ravi S. Menon,[2] Douglas B. Tweed,[4]
and Tutis Vilis[3],*
[1]CIHR Group on Action and Perception
[2]Department of Medical Biophysics and
   Robarts Research Institute
[3]Department of Physiology and Pharmacology
University of Western Ontario
London N6A 5C1
Canada
[4]Department of Physiology
Department of Medicine
University of Toronto
Toronto M5S 1A8
Canada

Summary

When we move our eyes, why does the world look stable even as its image flows across our retinas, and why do afterimages, which are stationary on the retinas, appear to move? Current theories say this is because we perceive motion by summation: if an object slips across the retina at $r°$/s while the eye turns at $e°$/s, the object's perceived velocity in space should be $r + e$. We show that activity in MT+, the visual-motion complex in human cortex, does reflect a mix of r and e rather than r alone. But we show also that, for optimal perception, r and e should not summate; rather, the signals coding e interact multiplicatively with the spatial gradient of illumination.

Introduction

Visual motion perception has been studied intensively, but current theories do not fully explain how we can perceive motion in our surroundings through eyes that are themselves almost constantly moving. Clearly, the visual system must use information about eye motion (Descartes, 1664) sensed by stretch detectors in the eye muscles (Sherrington, 1918), inferred from motor commands that move the eyes (von Helmholtz, 1867), derived directly from afferent retinal information (Gibson, 1968), or (most probably) all of these, used in combination (Wertheim, 1994).

The main question addressed here is not how the estimate of eye motion is derived but how it is used. It seems to be universally assumed that the process works by summation: if the image of an object moves across the retina at $r°$/s while the eye turns at $e°$/s, then the brain should perceive the object moving through space at $r + e°$/s (von Holst and Mittelstaedt, 1950; for reviews, see Shebilske, 1977; Carpenter, 1988). This summation theory is intuitive, but here we show that it is not in general correct, and that motion perception involves a different sensory-motor interaction.

*Correspondence: tutis.vilis@fmd.uwo.ca

To understand this interaction, we must consider the ambiguity of the raw data on which motion perception relies. Imagine a pattern of horizontal stripes filling your visual field. If the pattern moves horizontally, you will not see the motion. Nothing in your retinal image will change because the movement is orthogonal to the spatial gradient, or "slope," of retinal illumination: the movement is horizontal but the luminance varies only vertically. So the image does not reveal whether anything is moving or how fast. Of course, it seldom happens that the whole visual field is so ambiguous, but over smaller patches of retina, such as the receptive fields of motion-sensitive cells, the problem is recognized to be serious. These cells in effect view the world through small peepholes, so they are subject to the "aperture problem" (Marr, 1982): if a moving line is seen through a small aperture, then its movement is ambiguous, because the component of motion parallel to the line is invisible.

The brain tends to assume that this parallel component is zero (Wallach, 1976). In other words, the brain selects the smallest object velocity that is compatible with the retinal image. Mathematically, this minimal velocity $\mathbf{v}$ is

$$\mathbf{v} = (dI/dx)^+(dI/dt), \quad (1)$$

where dI/dx is the spatial gradient of retinal illumination, or in other words the rate of change of illumination, I, as a function of retinal location, $\mathbf{x}$; $(dI/dx)^+$ is the Moore-Penrose pseudoinverse of that gradient; and dI/dt is the temporal rate of change of illumination inside the aperture (see Ullman, 1983; Hildreth and Koch, 1987).

This ambiguity of retinal image motion is well known, but the ambiguity extends further to the perceptual interpretation of motor signals coding eye movement. We can see this most clearly if we consider a case where there is no motion at all in the retinal image: the case of an afterimage (Brown, 1966). If you see an afterimage of horizontal stripes and you move your eyes horizontally, then your sensory and motor signals are consistent with both a space-fixed field of stripes and a horizontally moving field. Which do you perceive? According to the summation theory, *the afterimage should appear to move with the eye*. We hypothesized instead that the brain infers the smallest velocity, relative to space, that is compatible with its sensory and motor data. Mathematically (see Appendix for the derivation), the perceived velocity $\mathbf{v}$ in space would then be

$$\mathbf{v} = (dI/dx)^+(dI/dt + (dI/dx) \cdot e), \quad (2)$$

where e is eye velocity and $\cdot$ is the vector dot product. In this view, perceived velocity $\mathbf{v}$ is not a *sum* of e with anything; rather, e interacts multiplicatively with dI/dx, and it is the product $(dI/dx) \cdot e$ that then enters a sum. In the case of an afterimage, I is of course retinal activity rather than real illumination, and its rate of change, dI/dt, is close to zero, so the equation simplifies to

$$\mathbf{v} = (dI/dx)^+((dI/dx) \cdot e). \quad (3)$$

If e and the afterimage lines are both horizontal, then

dI/dx and e are orthogonal, so (dI/dx) • e = 0, and therefore **v** = 0 as well; i.e., Equation 3 predicts that *the image should appear stationary in space*. On the other hand, if the line orientation is perpendicular to the eye motion, **v** = **e**, and the lines should be perceived to move with the eye. For intermediate orientations, the perceived motion should be always perpendicular to the afterimage lines, regardless of the direction of motion of the eye. Here we test these predictions of Equation 3, and we use fMRI to look for the predicted visual-motor interaction in primary visual cortex, V1, and in the visual motion complex, MT+, in the posterior temporal lobe.

Results

To identify cortical areas activated by visual motion, we presented a moving texture of randomly oriented lines, then a stationary display of the same texture. Subjects viewed these two displays in alternation while fixating on a stationary centered red dot. We localized area MT+ at the junction of the inferior temporal and lateral occipital sulci (Watson et al., 1993; Tootell et al., 1995; Dumoulin et al., 2000; Dukelow et al., 2001; Culham et al., 2001). We then examined how activity in this visual-motion area relates to perceived motion in the absence of movement in the retinal image with two experiments.

Experiment 1: Afterimage versus Control

The first experiment determined whether MT+ responds to an afterimage that is perceived to be moving in space, even though the image is of course stationary on the retina. Seven subjects fixated on a stationary red dot for 12 s against a large white stationary grid on a black background (Figure 1A). To make this stimulus comparable to that used in the control (see below), the grid was blanked for 0.2 s at 1 s intervals. Then the grid vanished and subjects tracked the fixation dot as it oscillated horizontally against the black background for 24 s at 1°/s, maximum displacement 1° (Figure 1A). All subjects reported perceiving a moving afterimage of the grid, and pressed a button when it vanished. The control task was identical except that a much weaker afterimage was formed because the grid was blanked and randomly shifted right or left at 1 s intervals.

During the 12 s period of afterimage formation, the flashing grid activated MT+, provoking a slightly stronger response when it was repeatedly displaced than when it flashed in a fixed location, perhaps due to the jumping grid exciting more retinal cells (Figure 1A shows the average response of the seven subjects).

During the subsequent, tracking period, activity in MT+ correlated strongly with the subject's perception of a moving afterimage. After viewing the jumping grid, all subjects reported a moving, but weak and brief, afterimage lasting on average 3.1 ± 0.3 s SE (Figure 1B), and MT+ showed a weak, brief increase in activity (Figure 1B). After the stationary grid, afterimages lasted longer—on average 9.3 ± 0.8 s SE (Figure 1B)—and MT+ showed a more sustained response (Figure 1B).

To compare the strengths of the fMRI responses, we computed the average activation for each subject (Figure 1C) across the same 10 s period (shaded area in Figure 1A). Activation was stronger after the stationary grid than after the jumping grid, both across all seven
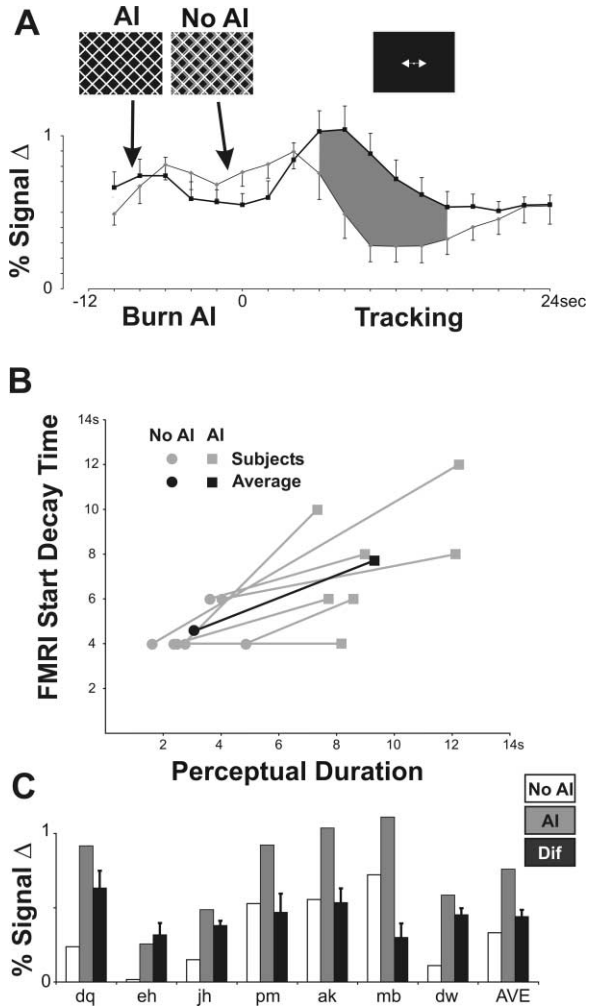

Figure 1. MT+ Responses to a Moving Afterimage

(A) Average activation across seven subjects during afterimage creation (Burn AI) and during the subsequent tracking period (Tracking). During tracking, activation was stronger and more prolonged (thick line) after the subjects viewed the stationary grid, which created a prolonged afterimage (AI), than it was after the periodically displaced grid (thin line), which created a weak, brief afterimage (No AI). The *y* axis shows the % change in MT+ activity (= (activation − baseline) × 100/baseline, where baseline was the average activation during those epochs, at the beginning and end of each sequence, in which a stationary fixation dot was displayed against a black background). Error bars mark one standard error. The shaded area marks the time period used to compute the average responses of individual subjects in (C).

(B) Durations of afterimages and fMRI responses are correlated. The *x* axis shows the average perceptual duration of the afterimage for each subject (gray lines and symbols) and their average (black line and symbol). The *y* axis shows the time when the fMRI signal began to decay. Squares show responses after viewing the stationary grid (AI), while circles show those after the displaced grid (No AI).

(C) During the tracking period, all seven subjects showed stronger activation after the stationary grid (white bars)—when the afterimage was strong—than after the displaced grid (gray bars); black bars indicate the difference. Error bars mark standard errors of individual subjects based on 12 measurements each.

subjects ($p < 0.00001$, paired t test) and within each subject ($p < 0.033$, paired t test). These differences likely do not reflect differences in attention (Treue and Maunsell, 1999; Huk et al., 2001) or eye motion (New-
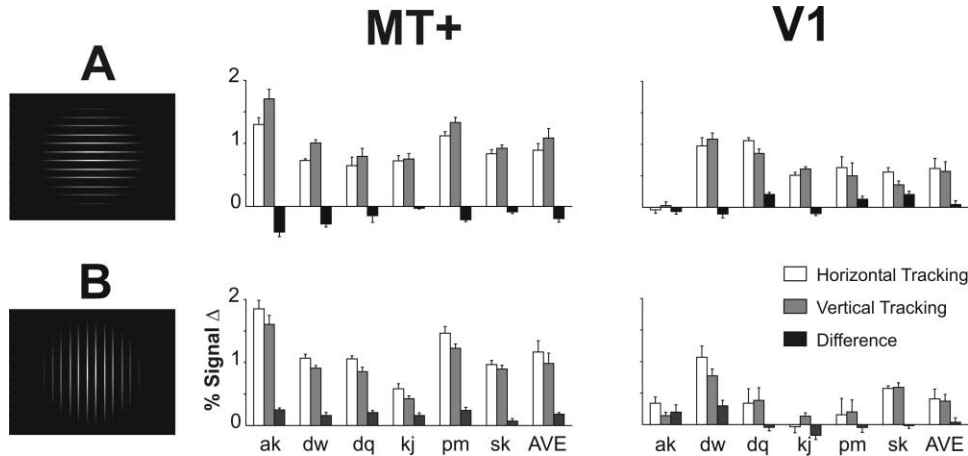
Figure 2. MT+ Responses to Afterimages of Horizontal and Vertical Lines Depend on an Interaction between Eye Motion and Line Orientation

Left graph: activations from area MT+ in six subjects during horizontal eye motion (white bars), during vertical eye motion (gray bars), and the difference (black bars), while subjects saw afterimages of horizontal lines (A) and vertical lines (B). Error bars mark standard errors of individual subjects based on 12 measurements each. The difference was greater during vertical tracking when the afterimage was of horizontal lines ($p < 0.02$, paired t test) and greater during horizontal tracking when the afterimage was of vertical lines ($p < 0.02$, paired t test). Right graphs: the same but from area V1. The difference was not significant for either horizontal or vertical lines ($p > 0.5$, paired t test).

some et al., 1988) because we maintained attention with the same behavioral task in both conditions, and the amplitude of the tracking eye movements did not differ significantly ($p > 0.05$, t test). Rather, the differences relate to the presence or absence of the afterimage. But do they specifically reflect the perceived *motion* of the image?

Experiment 2: Horizontal and Vertical Lines

To determine whether activation reflects perceived motion, we conducted a second experiment that was suggested by the proposed visual-motor interaction in Equation 3. Six subjects formed afterimages of either horizontal (Figure 2A) or vertical (Figure 2B) lines, then tracked an oscillating dot horizontally or vertically in darkness ($\pm 2°/s$, maximum displacement $2°$). Recall that, according to the standard, summation theory of motion perception, the afterimage lines should in all cases appear to move in the same direction as the eyes, while according to Equation 3, the perceived motion should depend on the spatial pattern, specifically on the orientation of the lines. With our goggles we could not make the lines span the visual field; therefore, our subjects could see the ends of the lines, where the gradient of illumination was not unidirectional. This complicates the predictions of Equation 3, but we minimized the complication by having the lines fade gradually at their ends, so that the disruptive gradients were small (Figure 2, left). Subjects pressed a button when the afterimage disappeared.

Afterimages lasted about equally long in all cases: the horizontal lines persisted for 13.0 ± 1.3 s SE during horizontal tracking and 11.9 ± 1.8 s SE during vertical tracking ($p > 0.4$, paired t test); vertical lines persisted for 13.0 ± 1.2 s SE during horizontal tracking and 14.6 ± 1.2 s SE during vertical ($p > 0.09$, paired t test). All subjects, however, reported perceiving stronger motion when tracking was perpendicular to the lines, contrary to the summation theory but in keeping with Equation 3.

We measured fMRI responses in MT+. These increased during tracking, as compared to stationary fixation on a black background, but activation was consistently higher when tracking was perpendicular to the lines (Figure 2, middle column): it was greater during vertical tracking when the afterimage was of horizontal lines ($p < 0.02$, paired t test) and greater during horizontal tracking when the afterimage was of vertical lines ($p < 0.02$, paired t test). As in the first experiment, these differences cannot plausibly be attributed to differences in attention (subjects attended to the afterimages for comparable lengths of time) or in eye motion (eye movements were recorded in five of the six subjects; these showed no consistent difference between the amplitude of horizontal and vertical tracking in a paired t test, $p > 0.6$; individually, three subjects made larger tracking movements horizontally than vertically, and one subject did the reverse, but none of these differences correlated with fMRI activity). And in the present experiment, the difference also cannot be attributed to differences in afterimage intensity or duration. Instead, the activity in MT+ seems to correlate specifically with the perception of a *moving* afterimage, and the motion coded there is not the percept predicted by the summation theory, but something more like the minimal velocity in space predicted by Equation 3, because it depends on the orientation of the afterimage lines.

We also measured activity in primary visual cortex, V1. Here, in contrast to MT+, there were no significant differences in activation between tracking perpendicular to the afterimage lines and tracking along them, for either horizontally oriented lines (Figure 2A, right column) or vertical lines (Figure 2B, right column) ($p > 0.5$, paired t test).

Experiment 3: Behavioral Measures

To test the predictions of Equation 3 more fully, we performed two further experiments outside the magnet. Equation 3 predicts that perceived speed should be

minimal when people move their eyes parallel to the afterimage lines and maximal when they move orthogonally to the lines. Further, the direction of perceived motion should depend only on line orientation and not on the direction of eye motion. To exclude all retinal input except for the afterimage itself, we had subjects track in a completely dark room (1) a sound emitted from a small speaker attached to their slowly moving finger and (2) the afterimage of a small dot located 4° off the fovea. Both the orientation of the lines and the direction of tracking were randomized.

Figure 3A shows the amount of perceived motion as a function of the angular difference between line orientation and the direction of tracking in the auditory/finger tracking experiment. In this subject, the smallest perceived velocity occurs around zero, when tracking was close to parallel to the line orientation. The maximum occurred near ±90°, when tracking was perpendicular to the line orientation. For each subject, we plotted perceived speed versus the angle between the afterimage lines and the tracking direction, we fitted a sine function to these plotted points, and we interpolated, using the sine function, to find the angle at which the perceived speed was maximal. Averaged across subjects, that angle was 87° ± 9° SE in the auditory/finger tracking experiment (Figure 3B) and 92° ± 6° SE for the dot afterimage tracking experiment (Figure 3C). In both experiments, these angles were not significantly different from 90° (95% confidence interval).

In agreement with the second prediction of Equation 3, the data also show that the direction of perceived motion correlates strongly with line orientation (Figure 4, left column) and not with the direction of tracking (Figure 4, right column). In the auditory/finger tracking experiment, a linear fit to the data as a function of line orientation yielded a mean slope of 0.8 ± 0.1 SE and an intercept of 93° ± 5° SE (Figure 4B, left column) (mean r^2 = 0.8). In contrast, when the data were plotted as a function of tracking direction, the mean slope was 0.04 ± 0.2 SE (Figure 4B, right column) (mean r^2 = 0.2). The data from the dot afterimage tracking confirmed the second prediction even more strongly. A linear fit to the data as a function of line orientation resulted in a mean slope of 0.97 ± 0.07 SE and an intercept of 90° ± 3° SE (Figure 4C, left column) (mean r^2 = 0.9). When the data were plotted as a function of tracking direction, the mean slope was 0.00 ± 0.3 SE (Figure 4C, right column) (mean r^2 = 0.07). Thus, in both experiments the perceived direction of motion was perpendicular to the afterimage lines and not correlated to the direction of tracking.

Discussion

Our results show that, contrary to intuition and the standard view, motion perception does not work by summing estimates of eye and retinal image velocity. Rather, the visual system computes object velocity in space by combining eye velocity signals with temporal and spatial visual information by a process closely equivalent to the multiplicative local interaction described in Equation 2.

Our neuroimaging data suggest that the resulting multimodal percept of motion in space is reflected in the activity of the visual motion complex, MT+. This
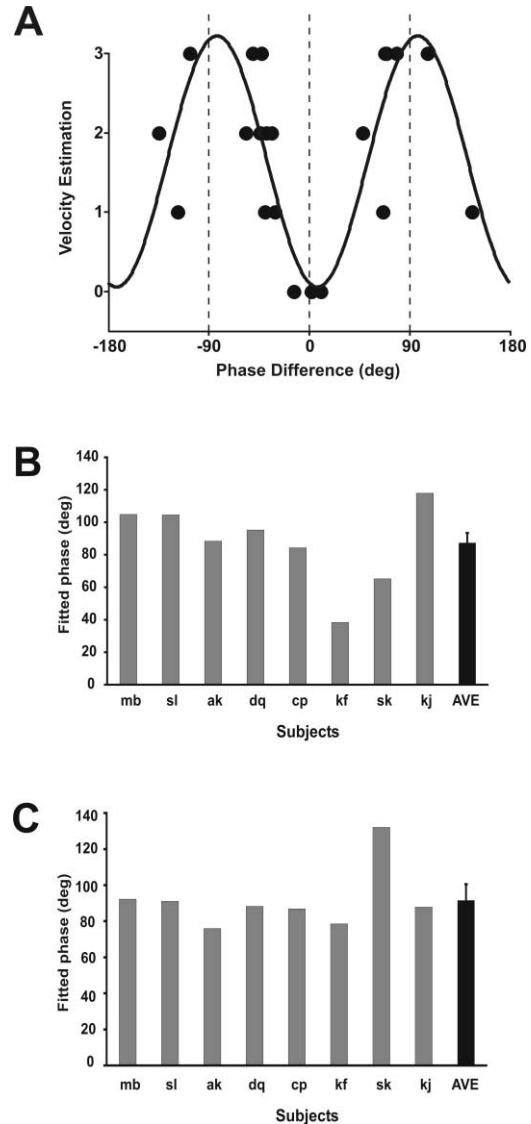

Figure 3. Perceived Velocity Magnitude of Afterimage Motion Depends on the Relative Orientation of the Afterimage's Lines and the Direction of Tracking Eye Movements

Subjects rated the perceived velocity on a 3-point scale (0, no movement).

(A) The perceived velocity magnitude as a function of relative orientation (line orientation − tracking orientation) in the auditory/finger tracking task for subject SL. Dots, individual data trials; solid line, the best fit sinusoid used to compute the maximum velocity angle.

(B) The angle of maximum velocity from the auditory/finger tracking experiment for the eight subjects. Error bars mark 1 standard error.

(C) The same for the dot afterimage tracking experiment.

area was first identified in the owl monkey (Zeki, 1980; Baker et al., 1981) in the medial temporal sulcus and has become known as area MT. Neurons in this region fire selectively in response to images that move in particular directions and at particular speeds (Maunsell and Van Essen, 1983). A comparable region, here called MT+, has been identified in humans at the posterior end of the inferior temporal sulcus using imaging techniques (Watson et al., 1993; Tootell et al., 1995; Dukelow et al., 2001). In patients with lesions of the human parieto-
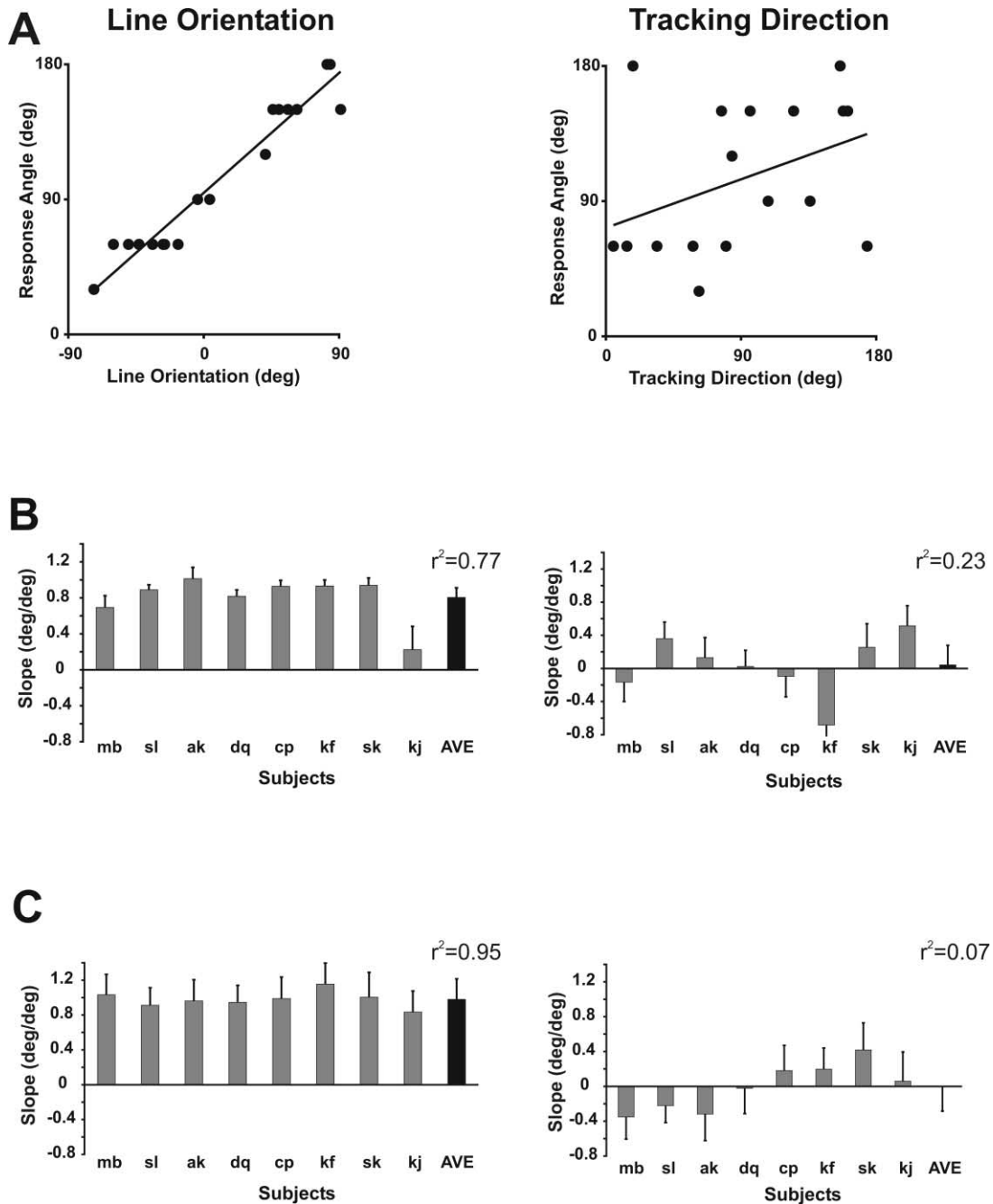
Figure 4. Perceived Direction of Afterimage Motion Depends on the Orientation of the Afterimage's Lines and not on the Direction of Tracking Eye Movements

(A) The perceived motion direction as a function line orientation (left) and instructed tracking direction (right) in the auditory/finger tracking task for subject SL. Filled circles, individual data trials; solid line, the least squares fit.

(B) The slopes and SEs from the auditory/finger tracking experiment for the eight subjects.

(C) The same for the dot afterimage tracking experiment.

occipital cortex that include MT+, eye movements produce a false perception of motion in objects that are still in the world (Haarmeier et al., 1997). Neurons in a subdivision of monkey MT+, MSTd, are activated by eye velocity even in the absence of retinal input (Newsome et al., 1988). Neurons here are also more strongly activated by retinal slip produced by motion of the object than when the same retinal slip is generated by movement of the eyes, an indication that these neurons signal not simply retinal slip but the combined influence of retinal slip and eye velocity (Erickson and Thier, 1991). Our neuroimaging data suggest that eye velocity plays a similar role in shaping activity in human MT+.

In contrast, our experiments suggest that the activity

in V1 is not correlated with the perception of motion in a stationary afterimage. This agrees with studies in the monkey which show that V1 does not discriminate between retinal slip that is produced by motion of an object and the same retinal slip produced by motion of the eye (Ilg and Thier, 1996). In earlier studies, a small minority of V1 neurons did prefer real retinal slip over that induced by pursuit eye movement, but these differences may be related to differences in the visual input (Galletti et al., 1984). A similar preference has also been reported in monkey areas V2 (Galletti et al., 1988) and V3a (Galletti et al., 1990). We were not able to test these latter areas because in our study we acquired only 8 slices, each 2.5 mm thick, to increase the spatial resolution, often leaving much of these areas outside our field of view.

The fact that fMRI activation is related to the perception of motion of afterimages in MT+ but not in V1 suggests that attention is not the causal factor. Recent fMRI studies have re-emphasized the importance of attention in modulating the activity of MT+ (Huk et al., 2001). This appears to be especially true for studies of the motion aftereffect in which prolonged viewing of motion in one direction results in the perception of oppositely directed motion of a stationary stimulus. The motion aftereffect likely involves a shift in the balance of activity between pools of neurons, the adapted pool becoming less active while the unadapted pool becomes more active. There may thus be little change in the overall level of activity as measured by fMRI. The situation is very different for the perception of motion of a retinally stable afterimage. Single-unit studies involving actual retinal motion demonstrate that activation in monkey MT+ depends on two factors: a visual object that moves in the neuron's on-direction and attention directed to this object (Treue and Maunsell, 1999). If either factor is absent, there is little activation. In the case of the perception of motion of a retinally stable afterimage, the percept of motion no doubt also requires these two factors: the presence of the afterimage and attention. In Experiment 2, the subject's task was to attend to the presence of the afterimage, not its motion, and the afterimage duration was not significantly different whether pursuit was perpendicular or parallel to the line orientation.

The differences in MT+ activation during tracking do not appear to be correlated to any differences in the velocity of the pursuit eye movement. It is important to exclude this possibility because neurons in monkey MST receive an extraretinal input related to the execution of pursuit eye movements (Newsome et al., 1988). In Experiment 2, the pursuit eye movements had the same average velocity in the test and control conditions and thus the resulting activation should subtract out.

Equation 2 did a good job of predicting perceived speeds and directions of motion in our experiments, but of course it is not a complete account of motion processing, because it describes the interaction of eye velocity and visual information only over patches of retina where the illumination gradient is unidirectional. When the brain pieces together information from different retinal sites where gradients differ, it very likely applies some additional, more global optimization principle. In the case of purely retinal motion data, without eye motion, theorists have proposed that the visual system chooses an interpretation that maximizes the smooth-

ness of the velocity field (Hildreth and Koch, 1987). The brain may well apply an analogous principle when it deduces spatial motion from eye velocity in the presence of more complex visual stimuli than our patterns of parallel lines.

Our experiments do not address the important question of how eye velocity, **e**, is derived. The extraretinal information derived from the efference copy of the signal generated by the oculomotor system (von Helmholtz, 1867) is no doubt not the only contributor. Nor is it likely that it is derived solely from the structural features of the patterns detected by the eye (Gibson, 1968). Rather, as proposed by Wertheim (1994), eye velocity appears to be derived from a combination of sources. Each source has inherent errors (Freeman and Banks, 1998) and is thus subject to recalibration. For example, eye velocity is continuously updated by the pattern of rotational flow detected by the eye (Haarmeier et al., 2001).

In summary, a critical flaw in the summation theory of motion perception is that it includes no interaction between the eye velocity signal and the spatial structure of the visual image. Our Equation 2 specifies an interaction based on the principle that the brain computes the smallest object velocity in space consistent with its sensory and motor data. This interaction correctly predicts the motion percepts in our experiments, and in other situations as well. For instance, consider the case where the spatial structure of the visual image is as simple as possible: if you move your eyes in pitch darkness when you see nothing at all—no afterimages—then the summation theory predicts that you should nevertheless see some sort of movement equal to your eyes' motion, but Equation 2 correctly predicts no such sensation, because the spatial gradient dI/dx is zero (and so is its Moore-Penrose inverse), and therefore **v** equals zero as well. In Equation 2, the spatial gradient interacts multiplicatively with eye velocity and with the temporal rate of change of retinal activity to determine the perceived speed and direction of object motion. Our imaging data suggest that this interaction is reflected in the activity of the motion complex MT+, but not in primary visual cortex V1, and may therefore occur between these areas or in MT+ itself.

Experimental Procedures

Subjects viewed images in high-resolution fiber-optic LCD goggles (Silent Vision 4021 series, Avotec Inc.) where each eye's screen subtended 30° horizontally and 23° vertically. We sampled the position of the right eye at 60 Hz using the Visible Eye™ Integrated Eye Tracking and Visual Stimulation System (Avotec Inc.), which includes an SMI iView-fMRI infrared video-based eye tracker (resolution 0.1°).

For neuroimaging, we used a 4.0 Tesla Varian whole-body system with Siemens Sonata gradients. We centered a 15.5 by 11.5 cm quadrature radio frequency surface coil over the subject's occipital pole. Functional data were collected using a volume acquisition time of 2.0 s. To locate MT+, we used voxels of 3 × 3 × 5 mm (11 slices); to examine the time course of its activation, 1.5 × 1.5 ×

2.5 mm (8 slices). Functional data were superimposed on high-resolution ($0.8 \times 0.8 \times 1.25$ mm) inversion-prepared 3D T1-weighted anatomical images of the brain collected immediately after the functional images using the same in-plane field of view. To visualize the anatomic location of MT+, we aligned the anatomic image obtained with the surface coil to one obtained with a full-head coil. We analyzed the data using BrainVoyager 4.6 software. Using general linear model analysis with a square-wave function convolved with the hemodynamic response, we identified those voxels whose significance exceeded $p < 0.00001$ and which lay adjacent to six or more similarly identified voxels.

To reveal regions of cortex activated by visual motion, the display alternated between a moving texture of randomly oriented lines that filled the display and, as a control, a stationary display of the same texture. For each moving epoch, the texture was rotated, translated horizontally or vertically, or contracted and expanded while subjects fixated a small central stationary red dot. These moving and stationary stimuli were each viewed for 12 s and repeated 24 times. MT+ was defined anatomically as the activation observed at the junction of the inferior temporal and lateral occipital sulci. V1 was defined as activity within the calcarine sulcus. In 3 of the 6 subjects, this location was confirmed by comparing the activation produced by video movies presented along the horizontal and vertical meridian.

Experiment 1

In this experiment, we compared the fMRI BOLD signal produced by an afterimage that was perceived as moving by the subject to that when little or no afterimage was present. Afterimages were created by instructing subjects to fixate on a small stationary red dot for 12 s against a large white stationary grid flashed (for .8 s every 1 s) on a black background. After this "burn" period, the grid was removed and subjects tracked the fixation dot as it oscillated slowly for 24 s ($\pm 1°$/s horizontally, maximum displacement $1°$) against a black background while perceiving a negative afterimage. During the control epoch, the afterimage was prevented from developing by periodically (1 Hz) blanking the grid and then shifting it randomly to evenly distributed locations to the right and left. After this "no burn" period, the grid was extinguished and a weak afterimage was perceived while the subjects tracked the motion of the red dot. Eye movements were measured as described above. The order of the epochs for one complete scan was blank, burn, test AI, no burn, test AI, no burn, test AI, burn, test AI, burn, test AI, no burn, test AI, no burn, test AI, burn, test AI, burn, test AI, blank. Each scan was repeated 3 times. To control the level of attention, the subject indicated, with a button press, when this weak afterimage disappeared. Seven subjects were studied.

Experiment 2

In this experiment, we compared the fMRI BOLD signal produced by an afterimage that was perceived as moving to one that produced a weaker perception of motion. This condition occurs when the afterimage is that of parallel lines. A stronger percept of motion is produced when one pursues in a direction perpendicular to the afterimage lines rather than parallel to the lines. To minimize the gradient at the ends of the lines, the lines used to "burn in" this afterimage were faded at their ends. Four conditions were compared. In one set of scans, the afterimage of horizontal lines was burnt in and we compared the fMRI response while the subject pursued horizontally and vertically ($\pm 2°$/s, maximum displacement $2°$). In another set of scans, responses during horizontal and vertical were compared after an afterimage of vertical lines was burnt in. As in Experiment 1, the subject indicated by a button press when this afterimage disappeared. Eye movements were measured as described above. The order of the epochs for one complete scan was blank, burn lines, horizontal pursuit, burn lines, vertical pursuit, burn lines, horizontal pursuit, burn lines, vertical pursuit, blank. The scan was repeated 3 times with horizontal lines and 3 times with vertical lines. Six subjects were studied.

Experiment 3

To test the predictions of Equation 3, we performed two behavioral experiments outside the magnet.

Tracking a Sound/Finger in the Dark

The same pattern of parallel lines, faded at the periphery, was displayed for 12 s to create an afterimage (pattern extended $100°$ horizontally and $75°$ vertically). Then, a pair of dots was flashed, on opposite sides of the fixation point: $6°$ on either side at random orientations. Then, in complete darkness, the subject slowly and smoothly moved his/her finger back and forth (.25 Hz as indicated by the sound of a metronome) between the remembered locations of the two dots. A small speaker, taped to the finger, provided an auditory cue to its location. Gauthier and Hofferer (1976) have shown that smooth pursuit eye movements are generated when subjects point to a moving acoustic target. After tracking the finger for 9 s, the subject was shown a clock face and asked to indicate the direction of motion perceived in the afterimage. Also, subjects rated the speed of the perceived motion on a 4-point scale (0 for no movement, 3 for fastest). This was repeated for 20 trials, each with a random orientation of both the displayed lines and the finger movement direction. Eight subjects were tested.

The subject's rating of *amount* of perceived motion (scale 0, 1, 2, or 3) was plotted as a function of the angular difference between line orientation and the direction of tracking. A least squares fit of these data to a sine wave of fixed frequency was used to compute angle (phase) of maximum velocity. The subject's rating of *direction* of perceived motion was plotted as a function of line orientation and the direction of tracking. Linear least squares regression was used to compute and compare the slope of each relation.

Tracking an Afterimage Target

As above, but with a different target. Two small white circles ($6°$) were added to the display of the parallel lines on opposite sides of the fixation point ($4°$ off center). This created an afterimage of the lines and the two small circles, like floaters, on either side of the fovea. Subjects tracked one circle and then the other, switching in time with the metronome (.25 Hz). The orientation of the pair of circles was randomized between trials, as was the orientation of the parallel lines. Eight subjects were tested.

Appendix: Derivation of the Visual-Motor Interaction Equation

First, we review how the brain might deduce motion from visual images when the eyes are stationary. Consider a small patch of retina, in which the spatial gradient of illumination, at the current instant, is $dI/d\mathbf{x}$ ($\mathbf{x}$ is a two-dimensional variable representing retinal location). And suppose the pattern of illumination is moving across the retina with velocity $\mathbf{v} = d\mathbf{x}/dt$. Then, the temporal rate of change of illumination in this patch is the product:

$$dI/dt = (dI/d\mathbf{x})d\mathbf{x}/dt. \qquad (4)$$

This means that the brain, knowing dI/dt and $dI/d\mathbf{x}$, could compute the image velocity, $d\mathbf{x}/dt$, by division: $d\mathbf{x}/dt = (dI/d\mathbf{x})^{-1}dI/dt$ (for a review of this gradient model, see Hildreth and Koch, 1987). But there is a problem: $dI/d\mathbf{x}$ is a kind of vector, and there is no unique way to divide by a vector; in other words, the inverse $(dI/d\mathbf{x})^{-1}$ is not uniquely defined, and so $d\mathbf{x}/dt$ cannot be uniquely determined. Physically, this reflects the fact that many different $d\mathbf{x}/dt$ vectors are consistent with any values of dI/dt and $dI/d\mathbf{x}$. The smallest of these $d\mathbf{x}/dt$ vectors can be computed using the Moore-Penrose pseudoinverse of the vector $dI/d\mathbf{x}$, $(dI/d\mathbf{x})^{+}$:

$$\mathbf{v} = (dI/d\mathbf{x})^{+}dI/dt. \qquad (5)$$

What happens when the eye moves? Some theorists, including J.J. Gibson (1968), have emphasized how much information about self-motion can be deduced from patterns of retinal flow, but it is clear that in many cases, as in our afterimage experiments, motion perception also uses nonvisual information about eye position. How do the visual and nonvisual data interact? The usual view, which originated with von Helmholtz (1867) and was formalized by Von Holst and Mittelstaedt (1950), is that perceived velocity is the sum of retinal motion and eye motion. That theory can be expressed by adding a term to Equation 5:

$$\mathbf{v} = (dI/d\mathbf{x})^{+}(dI/dt) + \mathbf{e} = d\mathbf{x}/dt + \mathbf{e}, \qquad (6)$$

where $\mathbf{v}$ is the perceived velocity of the object in space, $d\mathbf{x}/dt$ is

the velocity of the image across the retina, and **e** is eye velocity in space, which could be computed from sensory or efference-copy signals coding the motions of the eye in head and head in space (Wertheim, 1994).

In the case of an afterimage, d**x**/dt is zero and therefore **v** = **e**. Thus, according to Equation 6, the perceived motion should be determined solely by the motion of the eye. That this is not in general true is shown by our experiments in which the perceived motion is perpendicular to afterimage lines. These perceptions are explicable if we propose that the brain is computing the minimal object velocity in space that is compatible with its sensory and motor information, but that computation is very different from the one in Equation 6. Instead of adding eye velocity to a visual estimate of image velocity, **e** interacts with the spatial structure of the visual image, and

$$\mathbf{v} = (d\mathbf{I}/d\mathbf{x})^{+}(d\mathbf{I}/dt + (d\mathbf{I}/d\mathbf{x}) \bullet \mathbf{e}), \quad (7)$$

where • is the dot product of two vectors. This product determines the speed of the perceived motion of the stationary afterimage. It becomes zero when the tracking direction is perpendicular to the gradient orientation and reaches a maximum when the two are parallel. The direction of the perceived motion is determined by $(d\mathbf{I}/d\mathbf{x})^{+}$.

References

Baker, J.F., Petersen, S.E., Newsome, W.T., and Allman, J.M. (1981). Visual response properties of neurons in four extrastriate visual areas of the owl monkey (*Aotus trivirgatus*): a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. J. Neurophysiol. *45*, 397–416.

Brown, J.L. (1966). Afterimages. In Vision and Visual Perception, C.H. Graham, ed. (New York: Wiley), pp. 479–503.

Carpenter, R.H.S. (1988). Movements of the Eyes (London: Pion).

Culham, J.C., He, S., Dukelow, S.P., and Verstraten, F.A. (2001). Visual motion and the human brain: what has neuroimaging told us? Acta Psychol. *107*, 69–94.

Descartes, R. (1664). Treatise of Man. English translation by T.S. Hall (Cambridge, MA: Harvard University Press).

DeSouza, J.F.X., Dukelow, S.P., Gati, J.S., Menon, R.S., Andersen, R.A., and Vilis, T. (2000). Eye position signal modulates a human parietal pointing region during memory-guided movements. J. Neurosci. *20*, 5835–5840.

Dukelow, S.P., DeSouza, J.F.X., Culham, J.C., van den Berg, A.V., Menon, R.S., and Vilis, T. (2001). Localization of putative human area MST using fMRI. J. Neurophysiol. *86*, 1991–2000.

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

Erickson, R.G., and Thier, P. (1991). A neuronal correlate of spatial stability during periods of self-induced visual motion. Exp. Brain Res. *86*, 608–616.

Freeman, T.C., and Banks, M.S. (1998). Perceived head-centric speed is affected by both extra-retinal and retinal errors. Vision Res. *38*, 941–945.

Galletti, C., Squatrito, S., Battaglini, P.P., and Grazia Maioli, M. (1984). 'Real-motion' cells in the primary visual cortex of macaque monkeys. Brain Res. *301*, 95–110.

Galletti, C., Battaglini, P.P., and Aicardi, G. (1988). 'Real-motion' cells in visual area V2 of behaving macaque monkeys. Exp. Brain Res. *69*, 279–288.

Galletti, C., Battaglini, P.P., and Fattori, P. (1990). 'Real-motion' cells in area V3A of macaque visual cortex. Exp. Brain Res. *82*, 67–76.

Gauthier, G.M., and Hofferer, J.M. (1976). Eye movements in response to real and apparent motions of acoustic targets. Percept. Mot. Skills *42*, 963–971.

Gibson, J.J. (1968). What gives rise to the perception of motion? Psychol. Rev. *75*, 335–346.

Haarmeier, T., Thier, P., Repnow, M., and Petersen, D. (1997). False perception of motion in a patient who cannot compensate for eye movements. Nature *389*, 849–852.

Haarmeier, T., Bunjes, F., Lindner, A., Berret, E., and Thier, P. (2001). Optimizing visual motion perception during eye movements. Neuron *32*, 527–535.

Hildreth, E.C., and Koch, C. (1987). The analysis of visual motion: from computational theory to neuronal mechanisms. Annu. Rev. Neurosci. *10*, 477–533.

Huk, A.C., Ress, D., and Heeger, D.J. (2001). Neuronal basis of the motion aftereffect reconsidered. Neuron *32*, 161–172.

Ilg, U.J., and Thier, P. (1996). Inability of rhesus monkey area V1 to discriminate between self-induced and externally induced retinal image slip. Eur. J. Neurosci. *8*, 1156–1166.

Marr, D. (1982). Vision (New York: W.H. Freeman).

Maunsell, J.H., and Van Essen, D.C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J. Neurophysiol. *49*, 1127–1147.

Newsome, W.T., Wurtz, R.H., and Komatsu, H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. J. Neurophysiol. *60*, 604–620.

Shebilske, W.L. (1977). Visuomotor coordination in visual direction and position constancies. In Stability and Constancy in Visual Perception, W. Epstein, ed. (New York: Wiley), pp. 23–69.

Sherrington, C.S. (1918). Observations on the sensual role of the proprioceptive nerve-supply of the extrinsic ocular muscles. Brain *41*, 332–343.

Tootell, R.B., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., and Belliveau, J.W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J. Neurosci. *15*, 3215–3230.

Treue, S., and Maunsell, J.H.R. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. J. Neurosci. *19*, 7591–7602.

Ullman, S. (1983). The measurement of visual motion. Computational considerations and some neurophysiological implications. Trends Neurosci. *6*, 177–179.

von Helmholtz, H. (1867). Handbuch der Physiologischen Optik, Volume 2 (Hamburg: Voss), p. 234 in English translation by P.C. Southall (New York: Optical Society of America, 1925).

von Holst, E., and Mittelstaedt, H. (1950). The reafference principle: interaction between the central nervous system and the periphery, translation by R.D. Martin, published in The Behavioural Physiology of Animals and Man: Selected Papers of E. von Holst, Volume 1 (Coral Gables: University of Miami Press, 1973).

Wallach, H. (1976). On perceived identity. 1. The direction of motion of straight lines. In On Perception, H. Wallach, ed. (New York: Quadrangle).

Watson, J.D.G., Myers, R., Frackowiak, R.S., Hajnal, J.V., Woods, R.P., Mazziotta, J.C., Shipps, S., and Zeki, S. (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. Cereb. Cortex *3*, 79–94.

Wertheim, A.H. (1994). Motion perception during self-motion: the direct versus inferential controversy revisited. Behav. Brain Sci. *17*, 293–355.

Zeki, S. (1980). The response properties of cells in the middle temporal area (area MT) of owl monkey visual cortex. Proc. R. Soc. Lond. B. Biol. Sci. *207*, 239–248.