J Neurophysiol 93: 2908–2921, 2005.
First published December 1, 2004; doi:10.1152/jn.00473.2004.

Motion Mechanisms in Macaque MT

Bart Krekelberg and Thomas D. Albright
Howard Hughes Medical Institute, Systems Neurobiology Laboratories, Salk Institute for Biological Studies, La Jolla, California

Submitted 6 May 2004; accepted in final form 26 November 2004

Krekelberg, Bart and Thomas D. Albright. Motion mechanisms in macaque MT. *J Neurophysiol* 93: 2908–2921, 2005. First published December 1, 2004; doi:10.1152/jn.00473.2004. The macaque middle temporal area (MT) is exquisitely sensitive to visual motion and there is a large amount of evidence that neural activity in MT is tightly correlated with the perception of motion. The mechanisms by which MT neurons achieve their directional selectivity, however, have received considerably less attention. We investigated the motion–energy model as a description of motion mechanisms in macaque MT. We first confirmed one of the predictions of the motion–energy model; macaques—just like humans—perceive a reversed direction of motion when a stimulus reverses contrast with every displacement (reverse-phi). This reversal of perceived direction had a clear correlate in the neural responses of MT cells, which were predictive of the monkey's behavioral decisions. Second, we investigated how multiple motion–energy components are combined. Psychophysical data have been used to argue that motion–energy components representing opposite directions are subtracted from each other. Our data show, however, that the interactions among motion–energy components are more complex. In particular, we found that the influence of a given component on the response to a stimulus consisting of multiple components depends on factors other than the response to that component alone. This suggests that there are subthreshold nonlinear interactions among multiple motion–energy components; these could take place within MT or in earlier stages of the motion network such as V1. We propose a model that captures the complexity of these component interactions by means of a competitive interaction among the components. This provides a better description of the MT responses than the subtractive motion opponency envisaged in the motion–energy model, even when the latter is combined with a gain-control mechanism. The competitive interaction extends the dynamic range of the cells and allows them to provide information on more subtle changes in motion patterns, including changes that are not purely directional.

INTRODUCTION

The detection of visual motion must follow the detection of spatial contrast in the visual image. Efforts to understand the mechanisms that underlie motion detection have thus focused on the types of spatial contrast signals available in the visual image. A pivotal development in such efforts was the discovery, made nearly 35 yr ago by Anstis (1970), that if one inverts the luminance contrast of a pattern as it is displaced, the pattern will appear to move in the direction opposite to the displacement. To imagine this phenomenon, which Anstis termed *reverse-phi motion*, consider a simple pattern of white stripes that is repeatedly displaced rightward by small increments over a gray background. Although the physical displacement is over discrete spatial and temporal intervals, observers viewing this pattern reliably perceive an illusion of smooth, or "apparent," rightward motion. This is the phenomenon commonly known as *phi motion* (Wertheimer 1961). If, however, the black stripes become white, and vice versa, with each subsequent displacement, the pattern will appear to move to the left, i.e., in the reverse-phi direction (Fig. 1).

The powerful feature of this simple manipulation that leads to reverse-phi motion is the dissociation of two types of spatiotemporal continuity: on the one hand, luminance contrast boundaries ("features") continue to exist in the image and move in one direction despite changes in contrast polarity; on the other hand, the contrast reversals yield patterns of image intensity that move, on average, in the direction opposite to the displacements (Adelson and Bergen 1985; Anstis and Rogers 1975). The fact that motion is perceived in this latter direction reveals that the motion system does not rely on feature tracking but detects spatiotemporal continuity of image intensity patterns. This property is readily accounted for by models of motion detection that use oriented space–time filters, which are capable of identifying specific components of spatiotemporal Fourier energy (Adelson and Bergen 1985), commonly known as "motion energy."

Reverse-phi motion is a potentially useful tool for investigating the neuronal basis of motion processing. Most important, the behavior of a directional selective neuron in response to the reverse-phi manipulation may reveal the mechanism used to detect motion, and can be used to test the motion–energy hypothesis. In addition, regardless of the underlying mechanism, the sharp transition between perceptual states associated with a unique pattern attribute (contrast polarity) offers a fine opportunity to explore neural correlates of perceived motion.

A number of physiological studies have used luminance contrast-reversing stimuli of various types, which might be expected to elicit reverse-phi motion, and found that the responses of the recorded neurons "reversed" in the predicted manner (Emerson et al. 1987; Ibbotson and Clifford 2001; Livingstone et al. 2001). In primates, this was first noted by Dobkins and Albright (1994), who observed that the directionally selective responses of macaque middle temporal area (MT) neurons qualitatively mirrored the reverse-phi perceptual reports of human observers under similar stimulus conditions (Dobkins and Albright 1994). Livingstone et al. (2001) observed a related effect in MT neurons using one-dimensional sparse white-noise stimuli.

Inspired by the potential utility of reverse-phi stimuli for understanding motion mechanisms and encouraged by the earlier reports of neuronal sensitivity, we devised the present experiments. Our first goal was to determine the extent to

Address for reprint requests and other correspondence: B. Krekelberg, Systems Neurobiology Laboratories, The Salk Institute for Biological Studies, La Jolla, CA 92037 (E-mail: bart@salk.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

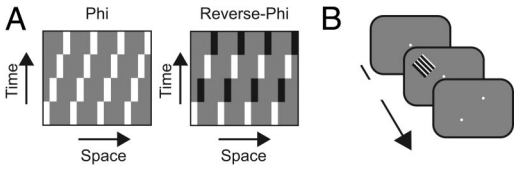

FIG. 1. Phi and reverse-phi stimuli and behavioral paradigm used to assess perceived direction of motion. *A*: space–time depiction of rightward drifting phi and reverse-phi stimuli. In each panel, the vertical axis represents the passage of time. The luminance profile of the grating stimulus along one axis of space is shown along the horizontal axis. The phi stimulus, illustrated in the *left panel*, consists of narrow white stripes on a gray background. Spatiotemporal flow of the white stripes and the Fourier energy is rightward. This stimulus is generally perceived to move rightward. The reverse-phi stimulus, illustrated in the *right panel*, is identical to the phi stimulus except that the narrow stripes undergo a contrast reversal (i.e., white to black, or black to white) with each displacement. Spatiotemporal flow of the changing stripes is rightward, but the Fourier energy is predominantly leftward. This stimulus is generally perceived to move leftward. *B*: behavioral paradigm. Each trial begins with the appearance of a single fixation spot (*rear panel*). After the spot has been fixated, the visual stimulus appears (*middle panel*), with the grating oriented such that it moves along the preferred axis of motion. After the stimulus is extinguished, 2 targets appear on either side of the fixation spot, aligned parallel to the axis of motion (*front panel*). The monkey must indicate perceived direction of motion by making a saccadic eye movement to the target that is displaced away from the fixation point in the direction of motion.

which MT neuronal responses and perceptual reports of reverse-phi motion are correlated. Thus we recorded neuronal responses in MT while the monkey performed a direction-discrimination task with phi and reverse-phi stimuli. The outcome of this experiment was consistent with the idea that the spatiotemporal Fourier components in the stimulus are the relevant signals for motion detection by monkeys and that the activity of MT cells is a neural correlate of this process.

The specification of a motion mechanism, however, is not complete with an understanding of what the relevant signals are. It needs to be supplemented with a description of how a multitude of such elementary signals is combined. According to the original motion–energy model, this operation is restricted to a linear subtraction of motion energy in opposite directions (motion opponency). In a second set of experiments, we tested this hypothesis and examined in some detail how MT neurons combine multiple spatiotemporal Fourier components. The data clearly show that a simple nonlinear summation of the component responses cannot explain the responses to multi-component stimuli. Instead, we propose that a competition among all Fourier components present in the stimulus provides an excellent description of the motion mechanisms in MT cells. Unlike the simple subtraction in the motion–energy model, this competition allows interactions among all Fourier components in a stimulus, not just the preferred and antipreferred directions. Moreover, the influence of any given component on the response varies greatly depending on the other components present in the stimulus. As a result, the response is not simply a vote as to whether the current direction is in the preferred or antipreferred direction, but rather a subtler signal that provides information on the total Fourier energy content of the stimulus.

METHODS

Overview

We conducted two experiments. *Experiment I* was designed to investigate the sensitivity of monkey observers and individual MT neurons to reverse-phi motion. Both behavioral and neuronal responses were recorded. As we show herein, the results of *Experiment I* are consistent with a model in which motion detection by MT neurons is accomplished through sensitivity to Fourier motion–energy components in the stimulus. *Experiment II* was designed to investigate this hypothesis more thoroughly and to determine how the multitude of Fourier components present in natural stimuli is combined. We begin with methods and procedures common to both experiments. Visual stimuli, behavioral paradigms, and electrophysiological recording procedures unique to each of the two experiments are identified below.

Subjects

Three adult male rhesus monkeys (*Macaca mulatta*; monkeys M, S, and T) were used in the neurophysiological part of *Experiment I*. Monkeys M and T served as subjects in the behavioral experiments. Simultaneous behavioral and neurophysiological measurements were obtained in monkey T. Two of the monkeys (monkeys S and M) were also used in *Experiment II*. The subjects had no significant refractive error. Experimental protocols were approved by the Salk Institute Animal Care and Use Committee, and conform to U.S. Department of Agriculture regulations and to the National Institutes of Health guidelines for humane care and use of laboratory animals.

Surgical preparation

Procedures for surgery and wound maintenance have been described in detail elsewhere (Dobkins and Albright 1994). In short, a head post and a recording cylinder were affixed to the skull using stainless steel rails, screws, and dental acrylic (monkeys M and T) or magnetic resonance (MR)–safe Cilux screws and dental acrylic (monkey S). Recording chambers were placed vertically above the anatomical location of area MT (typically 4 mm posterior to the interaural plane and 17 mm lateral to the midsagittal plane) to allow for a dorsoventral electrode trajectory. Chamber placement was guided by structural MR scans. In one animal (monkey T), a search coil for measuring eye position was surgically implanted in one eye using a variation of the method of Judge et al. (1980). After surgical recovery and attainment of criterion performance on the visual fixation task (see following text), a craniotomy was performed to allow for electrode passage into area MT. All surgical procedures were conducted under sterile conditions using isoflurane anesthesia.

Apparatus for visual stimulation

All visual stimuli were generated with in-house OpenGL software using a high-resolution graphics display controller (Quadro Pro Graphics card, 1,024 × 768 pixels, 8 bits/pixel) operating in a Pentium class computer. Stimuli were displayed on a 21-in. analog RGB video monitor (Sony GDM-2000TC; 75 Hz, noninterlaced). The output of the video monitor was measured with a PR650 photometer (Photo-Research), and the voltage/luminance relationship was linearized independently for each of the 3 guns in the CRT. Stimuli were viewed from a distance of 57 cm in a dark room (<0.5 cd/m^2).

Behavioral paradigm

GENERAL. Monkeys were seated in a standard primate chair with the head post rigidly supported by the chair frame. Eye position was monitored using one of 2 standard methods. In one animal (monkey T), eye position was sampled at 500 Hz using the magnetic scleral search coil technique. In the 2 remaining animals (monkeys M and S), eye position was sampled at 60 Hz using an infrared video-based system (IScan). Eye position data were monitored and recorded with the CORTEX program (Laboratory of Neuropsychology, NIMH: www.cortex.salk.edu), which was also used to implement the behav-

ioral paradigm and control stimulus presentation. When in a given trial the monkey's behavior did not precisely follow the rules of the task (including accurate fixation), that trial was terminated and the behavioral and neural data from that trial were removed from the data set.

FIXATION TASK. This basic behavioral task, which was used in both experiments, required subjects to fixate a small (0.15°) centrally located red spot for the duration of the trial. Each trial began with the appearance of this fixation spot on the video display. After ocular fixation was achieved and held for 250 ms, the stimulus (see following text) appeared. After stimulus offset, the fixation spot remained visible for another 250 ms. Trials in which eye position was maintained inside a square window (1.7° wide) surrounding the fixation spot were concluded with a small (0.15 ml) juice reward. A trial was aborted immediately if eye position strayed outside the fixation window at any time. Each trial was followed by a 1-s intertrial interval.

Electrophysiological recording

We recorded the activity of single units in area MT using tungsten microelectrodes (FHC, 3–5 MΩ base impedance), which were driven into cortex using a hydraulic micropositioner (David Kopf, model 650). Neurophysiological signals were filtered, sorted, and stored using the Plexon system (Plexon, Dallas, TX). Off-line spike sorting based on principal components analysis of the waveforms was used to separate up to 3 cells from a single electrode.

We identified area MT physiologically by its characteristically high proportion of cells with directionally selective responses, receptive fields that were small relative to those of neighboring area MST, and its location on the posterior bank of the superior temporal sulcus. The typical recording depth agreed well with the expected anatomical location of MT that was determined from the structural MR scans.

INITIAL ASSESSMENT OF DIRECTIONAL SELECTIVITY AND RECEPTIVE FIELD MAPPING. Directional tuning was assessed rapidly using whole-field circular motion in the frontoparallel plane (Schoppmann and Hoffmann 1976). In each trial, a random-dot pattern moved either clockwise or counterclockwise on a circular pathway for 1.25 s. Response rates were quantified as a function of phase of circular motion and the preferred direction was defined as the vector average of these firing rates. Statistical significance was assessed with a Rayleigh test.

The receptive field (RF) location was then determined using an automated sequence of briefly moving dot patterns. These patterns each contained 50 randomly positioned dots within a 5 × 5° window. Each pattern appeared for 260 ms at random nonoverlapping locations in a 6 × 6-square grid (i.e., subtending a 30 × 30° region of visual space), which was centered at the center of gaze. Dots moved coherently in the preferred direction at 10°/s. Each of these pattern presentations was followed by a 130-ms pause before the appearance of the next dot pattern at a different location. Neuronal responses were assessed in trials of 6 consecutive pattern presentations, during which subjects fixated a central target. The general structure of these RF mapping trials and the fixation requirements were identical to those for the fixation task described above. Total trial duration was, however, about 2,700 ms (the sum of 6 pattern presentations and associated pauses, etc.). A 2D spatial map of neuronal responsivity was derived by this procedure, from which the RF center was determined. We placed the stimuli of the main experiment, which always had the same 10° width and height, in the center of the RF, but made no attempts to optimize spatial and temporal frequency properties to the preference of the neuron.

We analyzed neuronal responses with our own software, written in Matlab 6.5 (The MathWorks, Natick, MA). A number of conventional response metrics were computed. The measure of neuronal response used for comparison of the effects of different experimental conditions was the mean spike rate computed within a window of 1,000 ms after response onset. Response onset for a given stimulus condition was defined as the start of the first 50-ms bin in which the firing rate was more than 3 SDs from the baseline firing rate (computed during the initial 250-ms interval during which fixation was maintained but no stimulus was presented in the RF). We used the minimum difference across all conditions between onset of stimulus motion and response onset as an estimate of the response latency of a cell. This minimum response latency was used to analyze all conditions. Comparisons between conditions were based on rank-sum tests.

EXPERIMENT I: A NEURAL CORRELATE OF THE REVERSE-PHI ILLUSION. *Visual stimuli.* Moving stimuli used for *Experiment I* were of 2 general types: phi and reverse-phi (see Fig. 1). The stimulus configurations used were based on the Gamma (Γ) stimulus defined by Chubb and Sperling (1989). Both phi and reverse-phi stimuli were composed of quarter duty-cycle, rectangular-wave gratings [spatial frequency = 0.25 cycles per degree (cpd)], which moved within a 10°-square aperture for 1,000 ms on each trial. The space–time-averaged luminance of the gratings (15 cd/m^2) was identical to that of the uniform gray field that surrounded them. Grating contrast (Michelson) was 15%. On each trial, the grating was displaced one quarter of its wavelength on every 4th video frame, i.e., every 53.3 ms (temporal frequency = 4.7 Hz). Reverse-phi stimuli differed from their phi counterparts solely by the fact that the luminance contrast polarity of the moving grating inverted with every displacement. Phi and reverse-phi stimuli were presented in the preferred and antipreferred directions (rounded to the nearest 45°) of the cell under study.

Stimulus/response nomenclature and definitions. The terms phi, reverse-phi, and direction of motion have heretofore been used to refer to both stimuli and the percepts that such stimuli may elicit. To avoid such ambiguity, we adhere to the following definitions throughout: *1*) Phi and reverse-phi stimuli refer to the moving grating and its contrast-reversing counterpart, respectively. Direction of stimulus motion refers to the direction of the smallest physical displacement of a grating on the screen, independent of grating contrast polarity. For the reverse-phi stimulus, this stepping direction does not correspond to the generally perceived direction of motion. *2*) Phi and reverse-phi motion refer to the motion percepts that are commonly, but not inevitably, elicited by phi and reverse-phi stimuli, respectively. For example, a reverse-phi stimulus will commonly elicit a percept of reverse-phi motion, but occasionally and under some conditions it will elicit phi motion (Chubb and Sperling 1989; Maruya et al. 2003).

Direction-discrimination task. This psychophysical task placed the same fixation requirements on the subject that were imposed in the fixation task. In addition, subjects (monkeys M and T) were required to report perceived direction of motion after stimulus presentation on each trial. Thus the sequence and timing of events on discrimination trials were identical to those on fixation trials until the end of the 250-ms fixation interval that followed stimulus presentation. At that time, 2 small (0.15°) red targets appeared on the display at positions equidistant (5°) from the central fixation spot. One target was displaced from the fixation spot in the direction of stimulus motion and the other was displaced in the opposite direction. Subjects were trained to report perceived direction of motion by making a saccadic eye movement to the target that was displaced in the direction of stimulus motion. Subjects were required to execute this response within 2,000 ms after target onset and to maintain fixation on the chosen target for 500 ms.

Subjects were initially trained to perform the direction discrimination task using phi-motion stimuli. After subjects reached a performance level of ≥80% correct in all 8 directions, we introduced reverse-phi stimuli. Trials with phi and reverse-phi stimuli were then randomly interleaved. On trials in which phi motion was presented, selection of the correct target was rewarded with a small drop (0.15 ml) of juice on 70–100% of the completed trials (the fraction of trials

that was rewarded was titrated for each animal/session to simultaneously minimize fluid consumption rate and maximize performance level). On trials in which reverse-phi stimuli were presented, juice reward was given on a random schedule (60% of trials). This random reinforcement for reverse-phi stimuli was used to avoid the possibility that the subject would become trained to report a particular direction for reverse-phi stimuli regardless of what was actually perceived.

Subjects discriminated 2 opposite directions of motion during any given behavioral session, but the stimulus parameters that were used depended on whether neurophysiological data were being obtained concurrently. In the case of monkey M, the stimuli were always positioned 5° to the left of fixation. We recorded >600 trials and averaged over all 8 directions of motion. In the case of monkey T, for which behavioral and neuronal data were obtained concurrently, the stimuli were centered on the RF of the recorded neuron and the axis of stimulus motion was chosen to most closely approximate the neuronal preferred axis of motion. In this case, we found that the behavioral data were largely independent of stimulus position in the visual field (within the limited range of eccentricities tested: <15°), and the physical direction of grating displacement. Data from monkey T were therefore pooled over all trials performed during physiological recordings (>2,000 trials).

Neuronal response measures. Direction selectivity indices were computed for phi (DSI_p) and reverse-phi (DSI_{rp}) stimuli, and were defined as the contrast between the response to the stimulus moving in the preferred direction and the stimulus moving in the antipreferred direction: DSI = 100 × [(preferred − antipreferred)/(preferred + antipreferred)]. Significance of the directional selectivity was assessed using a rank-sum test on the firing rates for preferred versus antipreferred stimulation. Note that performing this rank-sum test is not the same as testing whether the DSI is significantly different from 0, which explains why some cells with low DSI values are nevertheless significantly tuned for direction (see Fig. 4).

Quantifying the relationship between neuronal and behavioral responses. One of the animals from which we obtained neuronal responses (monkey T) also performed the direction-discrimination task. The psychophysical and neuronal data were obtained concurrently. We thus performed a choice-probability analysis, which determines the extent to which firing rate variations that are not related to changes in the visual stimulus are related to the decisions the monkey makes. A correlation between these measures is an indication that the cell plays a role in the monkey's decision process (Britten et al. 1996).

To determine the choice probability for a particular cell and a particular stimulus, responses were divided into 2 groups depending on the behavioral report on each trial. A receiver-operating characteristic (ROC) analysis of these 2 firing rate distributions gave the *choice probability* (CP). A CP >0.5 indicates that the cell fired more when the monkey made a decision that corresponded to the preferred direction of the cell. The reliability with which CP can be determined depends on stimulus-independent behavioral response variation and is thus tied to the fraction of incorrect responses. Because our monkey performed the task at high rates, we typically obtained few trials for one decision and many trials for the other decision. This situation prevents a meaningful analysis of the significance of the choice probability of individual cells (Britten et al. 1996). At the population level, however, the null hypothesis that the trial-to-trial response variability and the monkey's decision process are independent nevertheless predicts a distribution of CP values with a median of 0.5. We used a Wilcoxon signed-rank test to test this null hypothesis for the entire population of cells.

EXPERIMENT II: PROBING MOTION MECHANISMS IN THE FOURIER DOMAIN. Two of the 3 animals that participated in *Experiment I* (monkeys S and M) also served as subjects in *Experiment II*. The only behavioral requirement in this experiment was fixation of gaze.

Visual stimuli. The stimulus set used in *Experiment II* was based on the lower-order Fourier components (i.e., the components accounting for the greatest proportion of Fourier energy) of the phi stimulus used in *Experiment I*. In practice, this stimulus set was constructed for each neuron using the first 4 Fourier components for each direction along the neuronal preferred axis of motion. To illustrate, first consider the phi stimulus moving in the preferred direction. The first 4 Fourier components of this 0.25-cpd asymmetric rectangular-wave stimulus are sine-wave gratings of the following varieties (Lu and Sperling 1999):

Component	Spatial Frequency, cpd	Relative Contrast	Type	Abbreviation
1	0.25	1.0	Preferred	P
2	0.5	0.7	Static flicker	f
3	1.0	0.3	Antipreferred	a
4	2.0	0.2	Stationary	s

If we also consider the phi stimulus moving in the antipreferred direction, the set of components expands to include:

Component	Spatial Frequency, cpd	Relative Contrast	Type	Abbreviation
1	0.25	1.0	Antipreferred	A
2	0.5	0.7	Static flicker	f
3	1.0	0.3	Preferred	p
4	2.0	0.2	Stationary	s

Because the nonmoving components (2 and 4) are identical for each direction of motion, the basic stimulus set actually consists of 6 different Fourier components, which we designate by the following abbreviations: *P*, *a*, *p*, *A*, *f*, *s*.

We produced composite stimuli by adding together individual components. These were made from all combinations of the 4 components for each direction of motion: all 6 combinations of 2 components, all 4 combinations of 3 components, as well as the unique combination of 4 components. We stimulated each neuron using this composite stimulus set constructed for both preferred and antipreferred directions. In sum, our component stimulus set consisted of 6 different Fourier components and our composite stimulus set consisted of 21 different combinations of Fourier components [11 for each direction of motion, but one composite (*fs*) occurs twice]. Thus the full stimulus set contained 27 stimuli.

We stimulated each cell using the full stimulus set. As in *Experiment I*, all stimuli were presented within a square aperture centered on the RF. All component and composite stimuli were oriented perpendicularly to the preferred axis of motion (rounded to the nearest 45°).

Model fitting. We used a nonlinear least-squares curve-fitting algorithm in Matlab (lsqcurvefit.m) to determine the best-fitting parameters for each model. The initial conditions of the fitting procedure were randomly chosen and the best fit of 50 repetitions was used as the final answer. As an intuitive measure of the explanatory power of a model, we determined the percentage explained variance, given by: 100 × {1 − [variance (predicted − actual)/variance (actual)]}. This number is bounded above by 100%, but can become negative for extremely poor models.

The models we sought to compare differed both in their mathematical form and the number of free parameters. This invalidates comparisons based only on explained variance or nested log-likelihood. The Akaike Information Criterion (AIC) is a measure of model performance that allows one to compare the performance of arbitrary models while correcting for their different number of free parameters (Burnham and Anderson 1998). An intuitive (although not entirely correct) way to interpret the AIC is that it is the sum of a measure of the goodness-of-fit and a penalty term for the number of free parameters in the model

$$\mathrm{AIC} = -2\log(L) + 2\frac{kn}{n-k-1}$$

where L is the likelihood of the model, k is the number of free parameters in the model, and n is the number of samples. The difference in AIC values between two models can be used to select the better model. We followed the rule of thumb proposed by Burnham and Anderson (1998) and considered the model with the lowest AIC value to be better if its AIC value was ≥ 3 less than that of the alternative model. When the difference between the models was <3, the analysis was deemed inconclusive.

RESULTS

Experiment I: a neural correlate of the reverse-phi illusion

BEHAVIORAL DATA. Two of our 3 subjects (monkeys M and T) performed a direction-discrimination task using both phi and reverse-phi stimuli. The task and stimulus are detailed in METHODS and Fig. 1; the results are shown in Fig. 2. Behavioral reports elicited from both monkeys by the phi stimulus were predominantly of phi motion (monkey T, 91%; monkey M, 80% of trials). By contrast, when viewing the reverse-phi stimulus both monkeys reported reverse-phi motion on the majority of trials. Reports of phi motion were much less frequent (monkey T, 10%; monkey M, 32% of trials). This behavioral dichotomy mirrors that seen in human subjects, which suggests that monkeys experience phi and reverse-phi motion similarly to humans under these conditions. Interestingly, we found that, over the recording period of 4 mo, monkey T never appeared to recognize that rewards on reverse-phi trials were independent of his decision (i.e., he consistently reported a reverse-phi percept). This absence of extinction is testimony to the strength of the reverse-phi illusion. The demonstration that monkeys perceive the reverse-phi illusion shows that they are good animal models to investigate the neural correlates of this illusion and its implications for motion perception in general.

NEURONAL DATA. We recorded responses to phi and reverse-phi stimuli from 138 isolated MT neurons in 3 monkeys (monkey T, 82 neurons; monkey S, 27 neurons; monkey M, 29 neurons). These neurons showed statistically significant directional responses to the circular pathway stimulus (see METHODS). Of these neurons, 73 (53%) exhibited significant directional selectivity for either the phi or reverse-phi stimulus. The relatively low prevalence of significant direction selectivity for these stimuli can be attributed to at least two factors. First, our stimuli moved only every 53 ms; this stepping motion is beneficial to induce the reverse-phi percept, but it is known to reduce the direction selectivity of MT cells (Churchland and Lisberger 2001). Second, we did not optimize our stimuli to the spatiotemporal frequency preferences of the cells, but kept the stimuli the same for each cell. The responses of the neurons were indistinguishable across the 3 subjects and we have thus pooled data in the population analyses below. We begin, however, with an individual cell example taken from monkey T (Fig. 3), which enables us to illustrate the relationship between neuronal and perceptual sensitivity.

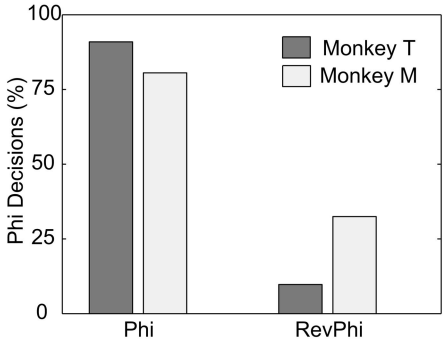

FIG. 2. Behavioral performance on a 2-alternative forced-choice direction-discrimination task using phi and reverse-phi stimuli. The vertical axis represents the percentage of trials on which the monkeys reported motion in the phi direction, for phi and for reverse-phi stimuli. For both monkeys, phi stimuli elicited predominantly phi motion reports. By contrast, reverse-phi stimuli elicited predominantly reverse-phi reports. In monkey T, we averaged the data over all (57) recording sessions (>2,000 trials); in monkey M we averaged 2 behavioral sessions (600 trials).

Example neuron. To illustrate the neural responses to phi and reverse-phi stimuli, we chose a single cell with relatively strong directional tuning. The RF center of the neuron shown in Fig. 3 was located about 8° from the center of gaze in the upper contralateral (*right*) quadrant (Fig. 3*A*). The assessment of directional tuning (Fig. 3*B*) revealed a preference for 315° (down and rightward) motion. As expected, when the phi stimulus moved in this direction, we observed a marked increase in the neuronal firing rate (Fig. 3*C*, *upper-right panel*). By contrast, when the stimulus was of the reverse-phi variety but moved in the same direction (315°), there was a very small onset transient, but no sustained response (Fig. 3*C*, *lower-right panel*). (Note that "direction of stimulus motion " here always refers to the direction of smallest grating displacement, independent of contrast polarity. See METHODS for details.) Importantly, when moving in the 135° direction, the phi stimulus elicited little sustained response (Fig. 3, *upper-left panel*), whereas the reverse-phi stimulus (Fig. 3, *lower-left panel*) elicited a response comparable to that obtained by the phi stimulus in the 315° direction. This is entirely consistent with the hypothesis that the neuron extracted the motion energy in the stimuli, because both phi stimuli moving in the 315° direction and reverse-phi stimuli moving in the 135° direction are associated with significant motion energy in the 315° direction.

In Fig. 3*D* we have plotted the average firing rate of our example MT neuron as a function of stimulus type and direction of motion. These data illustrate, once again, that responses were strongest when either phi stimuli moved in the 315° direction or reverse-phi stimuli moved in the 135° direction. Behavioral reports of perceived direction of motion are summarized in Fig. 3*E*. Here the vertical axis represents the fraction of trials on which the monkey reported motion in the 315° direction. Mirroring the neuronal data precisely, such reports were most common when either phi stimuli moved in the 315° direction or reverse-phi stimuli moved in the 135° direction.

Population data. The reverse-phi behavior of the cell highlighted in Fig. 3 was typical of the population. To quantify the effect across the population, we computed 2 directional selectivity indices for each neuron: one for phi stimuli (DSI_p) and another for reverse-phi stimuli (DSI_{rp}) (see METHODS). The sign of DSI_p was constrained to be positive. However, because we defined "direction of stimulus motion " to be the direction of smallest grating displacement, regardless of contrast polarity, a

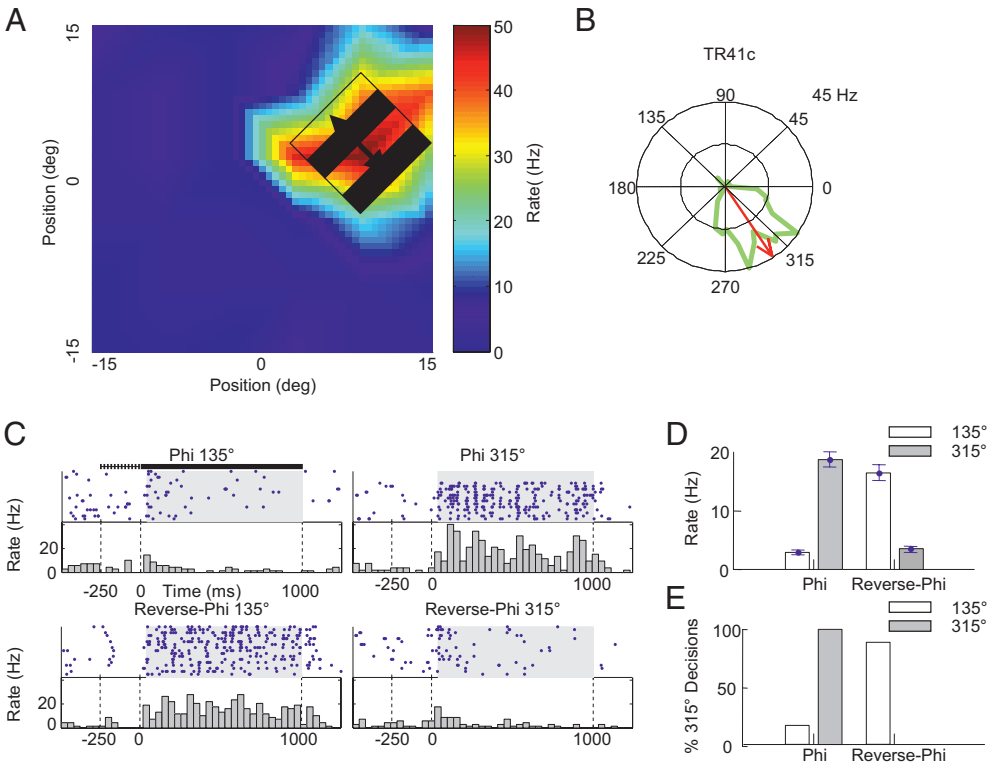

FIG. 3. Responses of a representative middle temporal (MT) neuron to phi and reverse-phi stimuli. *A*: receptive field map. The color code represents the firing rate evoked by a 5° × 5° patch of dots moving coherently in the preferred direction. The cartoon stimulus shows the position and axis of motion for phi and reverse-phi stimuli that were used to assess responses of this neuron. *B*: directional tuning curve plotted in polar format. The green curve shows the average firing rate of the cell to a coherently moving dot pattern, as a function of stimulus direction. Red arrow indicates the preferred direction, computed as the vector average. For assessment of phi and reverse-phi responses, the preferred and antipreferred directions were rounded to 315 and 135°, respectively. *C*: spike raster plots and peristimulus time histograms (PSTHs) illustrating responses to phi and reverse-phi stimuli moving in the 315 and 135° directions. Stimuli were stationary for 250 ms before motion onset, as indicated by the dashed line above *top-left* PSTH. Stimuli moved from 0 to 1,000 ms, as indicated by the solid black line above *top-left* PSTH. *D*: summary of neuronal firing rate as a function of stimulus condition. (Spikes for each condition were averaged within the gray windows, shown in *C*, which were corrected for neuronal response latency.) Largest responses were elicited by phi stimuli moving in the 315° direction and reverse-phi stimuli moving in the 135° direction. Both of these conditions yield motion energy predominantly in the 315° (preferred) direction. *E*: summary of behavioral reports as a function of stimulus condition. These behavioral data were collected simultaneously with the neurophysiological data presented in *C* and *D*. In accordance with the magnitude of neuronal response (*D*), reports of perceived motion in the 315° direction were most common for phi stimuli moving in the 315° direction and reverse-phi stimuli moving in the 135° direction. Note that in this experiment, the monkey never reported motion in the 315° direction for a reverse-phi stimulus stepping in the 315° direction.

cell that detects reverse-phi motion—a cell with a selectivity pattern like the one shown in Fig. 3—will exhibit a negative DSI_{rp}. For the cell in Fig. 3 $DSI_p = 76\%$ and $DSI_{rp} = -71\%$. In Fig. 4 we have cross-plotted values of DSI_{rp} and DSI_p for the cells that were directionally selective ($n = 73$) for either phi or reverse-phi stimuli. As can be seen from the marginal distribution, DSI_{rp} was negative for the majority (58 out of 73 cells; 80%) of cells and the median (−26%) was significantly different from zero ($P < 0.01$). Moreover, values of DSI_p and DSI_{rp} were highly correlated (Spearman $R = -0.63$; $P < 0.001$) for those neurons that exhibited significant selectivity for both stimulus types (red dots). Even for the subpopulation of neurons that were not significantly directionally selective (and thus not shown in Fig. 4), the median $DSI_p = 7.6\%$, whereas $DSI_{rp} = -4\%$ (both significantly different from zero; $P < 0.05$). This confirms that, even though the strong direction selectivity for phi motion of the example neuron in Fig. 3 was atypical for the population, the reversal of the directional response for the reverse-phi stimulus was typical.

Choice probability. As Fig. 3*E* reveals, monkey T did not always give the same answer to the same stimulus. To quantify the neuron's involvement in the monkey's decision process, we determined the choice probability (CP; see METHODS). This analysis was performed for every recording in which the monkey made at least one mistake; this resulted in a database of 94 neurons. These neurons all had statistically significant direction-selective responses to smoothly moving random-dot patterns. A CP value >0.5 indicates that, on trials in which the cell fired more than usual, the monkey was more likely to report motion in the preferred direction of the cell. The distribution of CP values for the phi stimulus is plotted in Fig. 5*A*. Given the small fraction of error trials (10%), an analysis of significance on individual cells is not meaningful (Britten et al. 1996). The overall distribution of CPs, however, possessed a median (0.60) that was significantly greater than 0.5 ($P < 0.05$). The corresponding distribution of CP values for the reverse-phi stimulus is plotted in Fig. 5*B*. For the reverse-phi stimuli, the median of the distribution of CP values (0.55) was also significantly larger than 0.5 ($P < 0.05$). These MT choice probabilities were comparable with those found in other direction-of-motion tasks (0.55; Britten et al. 1996) but smaller than those found for a stereoscopic depth task (0.67; Dodd et al. 2001). Note also that for phi and reverse-phi stimuli, the effect was somewhat larger for stimuli moving in the preferred

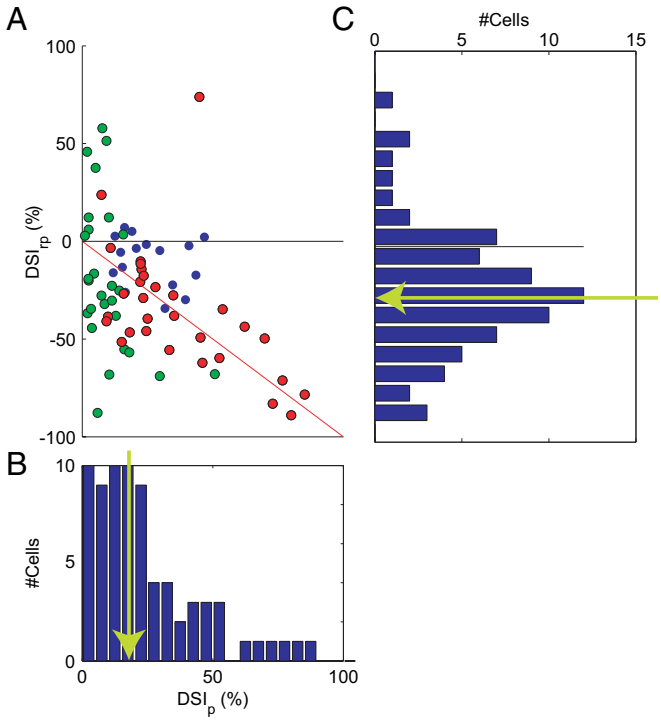

FIG. 4. Relationship between directional tuning for phi and reverse-phi stimuli at the neuronal population level. *A*: scatter plot of directional selectivity indices for phi (DSI_p) vs. reverse-phi (DSI_{rp}) stimuli. Each data point represents data from a single MT neuron. Dot color conveys significance of directional tuning per stimulus type: phi stimulus only, blue (16 cells); reverse-phi stimulus only, green (27 cells); directionally selective for both stimuli, red (30 cells) (significance based on rank-sum test, $P < 0.05$; see METHODS). *B*: marginal distribution of DSI_p. By definition, all values were positive. Green arrow indicates the median (18%) of the distribution. *C*: marginal distribution of DSI_{rp}. Negative values indicate that the neuronal response to reverse-phi stimuli was largest when the stimuli moved in the antipreferred direction, which is the condition that should elicit a percept of motion in the preferred direction. Most cells possessed a negative DSI_{rp}. Median of the distribution (−26%) is indicated by the green arrow.

direction than for stimuli moving in the antipreferred direction. However, because the trends in the separate data sets were the same, they were pooled to construct the histogram in Fig. 5. These data show that the monkey's behavioral report of direction was on average predicted by MT responses both for phi and for reverse-phi stimuli, which suggests, in turn, that the two stimulus types engaged a common decision mechanism.

The results from *Experiment I* provide evidence that MT neurons play a significant functional role in reverse-phi motion perception. This suggests that the mechanisms underlying motion detection by many MT neurons are not based on the tracking of feature boundaries. By contrast, a model based on the detection of motion energy offers an elegant explanation for the reverse-phi phenomenon. Specifically, the motion–energy model posits the existence of filters for patterns of image intensity that are oriented in space–time and thus detect spatiotemporal Fourier components in a moving stimulus. When contrast is reversed with every spatial displacement, some of these Fourier components reverse direction of motion. Thus if the spatiotemporal Fourier components do indeed provide the relevant signals for motion detection mechanisms, one would expect the observed reversal of detected motion for reverse-phi stimuli. Although the physiological results of *Experiment I* are thus generally supportive of the motion energy model, they do not specify how the multitude of Fourier components present in natural stimuli is combined. This issue was addressed in *Experiment II*.

Experiment II: probing motion mechanisms in the Fourier domain

In this experiment we investigated how neuronal responses to single Fourier components are related to the neural responses to a stimulus that contains multiple Fourier components. Specifically, we recorded responses of 54 MT neurons to 6 single Fourier components (sine-wave gratings) as well as 21 composite stimuli that were created by combining these components (see METHODS). Comparing responses to the individual Fourier components with those elicited by the composite stimuli allowed us to determine how MT neurons combine information in Fourier space.

ILLUSTRATIVE NEURONAL RESPONSES. Data from one MT neuron are shown in Fig. 6. The inner circle of the figure shows peristimulus histograms of the neuronal responses to 6 individual Fourier components of the phi stimulus (*P*, *p*: preferred direction; *f*, flicker; *s*, stationary; *A*, *a*, antipreferred; see METHODS). As expected, the largest response was elicited by the components moving in the preferred direction (stimulus *P*, *p*). The cell also responded strongly to the flickering sinusoid (stimulus *f*), which corroborates previous evidence that many MT neurons respond to nonmoving contrast-modulated patterns (Churan and Ilg 2002; Thiele et al. 2000). The components moving in the antipreferred direction (stimuli *A*, *a*) induced a small above-baseline response. Finally, the stationary component (*s*) reduced firing somewhat below baseline. We found such a suppressive effect of the stationary component in a minority (22) of the 54 cells. Because baseline activity is low in MT, suppression below baseline is also

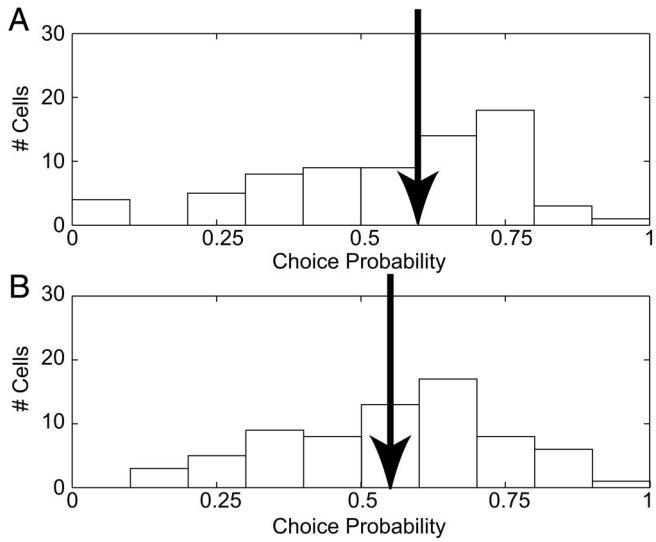

FIG. 5. Choice probability (CP) analysis in monkey T. *A*: distribution of CPs for the phi stimulus. Distribution is significantly biased toward values >0.5, which indicates that, on average in this population, the neuronal response magnitude on individual trials was predictive of the monkey's decisions regarding the perceived direction of motion. Arrow indicates the median choice probability (0.60). *B*: distribution of CPs for the reverse-phi stimulus. As for the phi stimulus, this distribution is biased toward values >0.5, which indicates that neuronal responses are predictive (median choice probability = 0.55) of the report of reverse-phi motion.

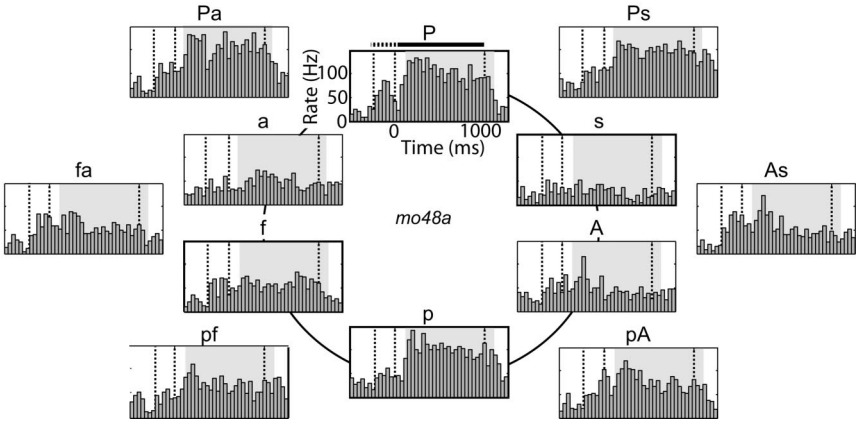

FIG. 6. MT responses to individual Fourier components and to superpositions of those components. *A*: responses of a representative MT neuron to each of the 6 distinct Fourier components (denoted *P*, *f*, *a*, *s*, *A*, and *p* and defined in METHODS) are shown in the form of peristimulus histograms on the inner circle. Time and response scales are shown in the histogram corresponding to the *P* stimulus. Stimuli appeared static within the response field (RF) for 250 ms before motion onset, as indicated by dashed line above the PSTH for the *P* component. Stimuli moved within the RF during the period indicated by the solid black line. Outer circle of histograms shows responses to composite stimuli that are a superposition of the 2 components whose responses are shown adjacent on the inner circle.

necessarily small. Thus when averaging the response of the MT population to stationary stimuli it was still above baseline. Consistent with previous reports (Albright 1984), the population average response to a stationary stimulus was about 24% of the response to the preferred stimulus.

Between every 2 components, a PSTH shows the cell's response to a stimulus composed of those 2 components. For example, the response to the stimulus produced by superimposing an antipreferred direction sinusoid (*a*) on the preferred direction sinusoid (*P*) is shown in the *top left* PSTH (stimulus *Pa*). The cell's response to *Pa* is not substantially different from the response to *P* alone. The composite stimulus *pf*, on the other hand, elicited a response that is the average of the responses to the components *p* and *f*. Thus cursory comparison of responses to individual versus composite stimuli does not point to a simple interaction rule, such as linear summation, but suggests that there are complex interactions between Fourier components. In an effort to identify consistent properties of the underlying mechanism, we developed 3 mathematical descriptions of these interactions.

MODELING COMPONENT INTERACTIONS. *1) Power-law summation.* In the power-law summation model the response to a composite stimulus is a power-law function of the responses to the components. Power-law summation accurately describes how responses to multiple objects in an MT receptive field depend on responses to single objects in the RF (Britten and Heuer 1999; Heuer and Britten 2002). Because this model works well for spatial summation, it seems worthwhile investigating whether it is an accurate description of Fourier space summation. The mathematical form we used is as follows

$$R = \text{gain}\left\lceil \sum_i \text{sign}\,(r_i)*(r_i)^{\text{exp}} \right\rceil^{1/\text{exp}} + \text{offset}$$

where R is an MT neuron's response to a composite stimulus, and the r_i represent that neuron's responses to the component stimuli. The sign of the component (+1 for a component that drives the cell above baseline; −1 for a component that inhibits the cell below baseline) is denoted by sign (r_i). The summation Σ runs over all components. Finally, the $\lceil\ \rceil$ operator represents a linear threshold that sets all negative values to zero and leaves positive values unaltered.

This model includes both excitation and (subtractive) inhibition through the sign (r_i) term: if a component reduces firing below baseline, its response will be subtracted to determine the response to the composite stimulus. The appealing property of this model is that it allows one to study qualitatively different interactions within a single mathematical framework. For instance, with gain = 1 and exp = 1, the model performs simple weighted linear summation. With large values of exp, on the other hand, the model implements a highly nonlinear winner-take-all algorithm, in which the response to the composite is essentially determined by the response to the strongest component. With exp = 0.5, this model implements a half-squaring nonlinearity, similar to the nonlinearity used by Simoncelli and Heeger (1989) in their model of pattern MT cells. To systematically investigate these qualitatively different models, we followed Britten and Heuer (1999) and determined best-fitting parameters per cell, while constraining one or 2 of the 3 model parameters. The offset parameter was always free. Table 1 summarizes the results of this analysis in terms of the average percentage of the variance in the data that the model could explain.

The performance of none of these models is impressive. Power-law summation, however, does outperform the other models. Even after correcting for the extra free parameter, the full power-law summation model fits the data better than scaled linear summation for 45% of the cells and better than the Simoncelli and Heeger model for 63% of the cells (nested log-likelihood tests, $P < 0.05$). In other words, allowing the summation to be nonlinear and different for each cell provides a better fit. The importance of strong nonlinear interactions is further confirmed by the observation that cells for which the power-law model explained more than half of the variance typically had a high exponent (median 2.8).

Because the median-explained variance is quite low in these simple summation models, it is perhaps more interesting to determine why none of the models worked than determining which of these models worked best. The example cell in Fig. 6 provides some clues. For this cell, the average responses to the

TABLE 1. *Comparison of simple linear and nonlinear Fourier space summation models*

Model	Gain	Exponent	Variance, %
Winner-take-all	1	100	28
Unscaled power-law summation	1	Free	34
Simoncelli and Heeger	Free	0.5	35
Scaled linear summation	Free	1	43
Power-law summation	Free	Free	53

Values based on $n = 54$ cells. The explained variance refers to the median percentage of the variance that could be explained by the model.

p and *P* components were nearly identical and so were the responses to the *a* and *A* components. A comparison of the *pA* and *Pa* composites, however, shows that even though the components of these stimuli evoked very similar firing rates in this cell, the composites constructed by combining those evoked very different responses.

This is an important observation because it shows that the rates that the components evoked are not by themselves sufficient to explain how the cell will respond to a composite stimulus consisting of only those components. In other words, any model, regardless of its precise mathematical form, in which the response to a composite stimulus is determined only by the response to the components, cannot explain the full Fourier space summation properties of this MT cell. To investigate how prevalent these responses properties were in our database, we searched, for each cell, for pairs of composite stimuli that contained 2 components that evoked similar rates (e.g., the *Pa* and *pA* stimuli in Fig. 6). The difference between the component responses was allowed to be at most equal to the average SE in the component responses. Over the whole database, 210 pairs of composite stimuli with 2 components satisfied this constraint. Any model relying on only the component responses to explain the composite response would predict composite responses that were very similar. To quantify how often such a model would fail, we counted how often the difference in the composite response was >2 SE; this was the case in 127 of the 210 pairs (60%). Thus even using a conservative criterion, and looking only at one way in which any purely response-based model would fail, we have strong evidence against such models for more than half of the cells.

We conclude that one reason why the simple summation models of Table 1 fail is that the influence of a component on a composite stimulus depends not only on its firing rate but also on other factors. One factor could be that a component has an influence on a composite that is stronger or weaker than the response to the component alone would suggest. We investigated such first-order factors in the weighted power-law model.

2) Weighted power-law summation. The weighted power-law model for Fourier domain summation extends the standard power-law model by introducing a separate gain parameter (W_i) for each component. The mathematical form is

$$R = \left\lceil \sum_i \text{sign}\,(r_i)*(W_i*C_i)^{\text{exp}} + \text{offset} \right\rceil^{1/\text{exp}}$$

where C_i is an indicator variable representing the absence (0) or presence (1) of component *i* in the stimulus. The summation Σ still runs over all components, each of which is now weighted by a different weight that is constrained to be positive (W_i). Through these weights, the components can have an influence on the response *R* that is not directly proportional to the response that the component evoked when presented alone. Note, however, that these weights were also constrained by the fact that when all C_i were zero except one, the response of the model should predict the response to the single component. Thus this model is qualitatively different from the power-law model in that it aims to describe not only the composite responses but also the responses to the individual components.

We again determined the best-fitting parameters separately for each cell. The median best-fitting gain parameter for all cell models was 2.2 (quartile range: [1.0; 4.0]) and the median exponent was 16 (quartile range: [4; 261]; note that values >10 essentially represent winner-take-all behavior). On average the power-law model explained 75% of the variance (quartile range: [62%; 87%]). The component weights (W_i) in these weighted power-law models show, not surprisingly, that the preferred direction components (*P*, *p*) typically have the greatest influence on the composite response, followed by the flicker and antipreferred components (*f*, *N*, *n*). The stationary (*s*) component weight was typically small. For most cells, the best-fitting power-law model was a winner-take-all mechanism. In other words, the response to a composite was primarily determined by the response to the component that best drove the cell. Although this model does explain a significant part of the variance in the data set, it also leaves some structure in the responses to composite stimuli unexplained.

To illustrate a weakness of the weighted power-law model, we have presented a model fit for an illustrative MT neuron in Fig. 7*A*. The model response (vertical axis) was limited to a small set of responses, whereas the actual firing rate (horizontal axis) varied more or less continuously over a large response range. The cluster of data points with model responses at about 27 Hz corresponds to stimuli that contained a preferred component (stimulus *P*). Every composite stimulus that contained this component yielded an identical output from the power-law model in winner-take-all mode. Indeed, even when the pre-

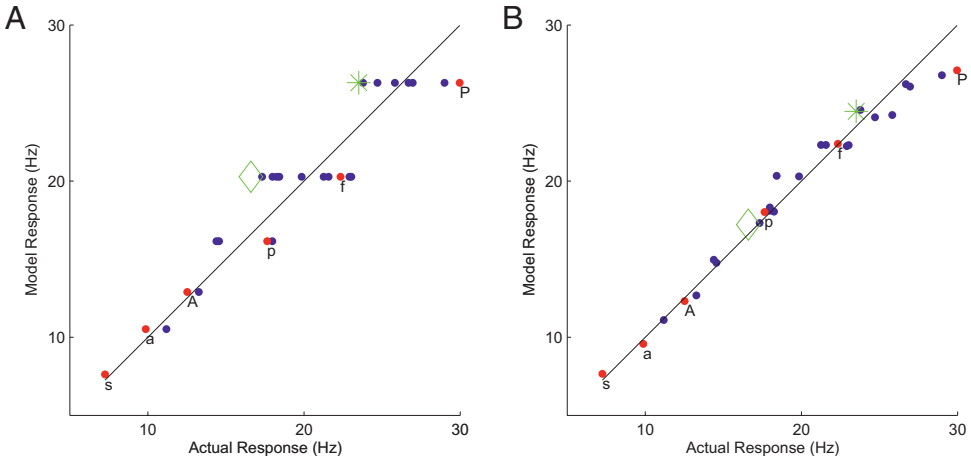

FIG. 7. Model fits for an example MT neuron. *A*: comparison of the actual responses of an example cell with the responses predicted by the best-fitting power-law model. The model resembles a winner-take-all algorithm and captures 89% of the variance in the data. *B*: performance of the best-fitting competition model of the same cell. This model captures 97% of the variance in the data. In red are responses to single-component stimuli. The green diamond represents a stimulus in which a flicker component is combined with an antipreferred component (*Af*). The green asterisk corresponds to a stimulus that is a combination of the preferred component and 3 nonoptimal components (*Pfas*). Responses to these stimuli are discussed in the text. Blue dots correspond to other composite stimuli.

ferred component was combined with 3 nonoptimal components (yielding composite stimulus *Pfas*, indicated by asterisk), the model response did not change and clearly overestimated the actual response. Similarly, the cluster of data points with model responses at about 20 Hz corresponds to stimuli that contained the flicker component (stimulus *f*). When this flicker component was combined with the antipreferred component (yielding composite stimulus *Af*, indicated by a diamond), the response was again overestimated by the model. In this case, the suppressive influence of the antipreferred component (when presented alone this component inhibited the cell below baseline) was ineffective when combined with the excitation elicited by the flicker component.

Because the weighted power-law model typically converged to winner-take-all mode, the influence of a component depended on the other components present in the stimulus in an all-or-nothing fashion. In the neurons, however, adding a given component had a much more graded effect that varied depending on which other components were present in the composite stimulus. The competition model, investigated next, allows for such flexibility.

3) Competition. To account for graded and complex interactions among Fourier components, we considered a feedforward network model that uses tuned excitatory input and tuned divisive inhibition to implement a competition among the component inputs (Grossberg 1973; Reynolds et al. 1999). Unlike the winner-take-all kind of competition implicit in the power-law model, this competition does not necessarily lead to a single "winner. " In the competition model, MT neurons receive both excitatory and inhibitory inputs from the same components, but with different weights. Mathematically

$$E = \sum_j W_j^+ C_j \quad I = \sum_j W_j^- C_j \quad R = \frac{K*E}{E + I + \sigma}$$

where E and I represent the excitatory and inhibitory drive to the modeled neuron, respectively, and R is the neuronal response. We fixed the K parameter to the maximum firing rate of the cell and the σ parameter (arbitrarily) to 1. This leaves the 6 excitatory (W_j^+) and 6 inhibitory weights (W_j^-) to be found by parameter fitting. All weights are constrained to be positive. Just as in the model proposed by Simoncelli and Heeger (1998), divisive inhibition is an important aspect of this model. An important difference, however, is that the (W_j^-) weights make this divisive inhibition tuned in Fourier space.

The competition model worked extremely well. On average, it explained 94% of the variance (quartile range: [86%, 97%]). An example fit for direct comparison with the weighted power-law model of Fig. 7*A* is shown in Fig. 7*B*. For this cell, the highest firing rates were somewhat underestimated by the competition model, but this was not typical across the population. Importantly, the competition model captured the influence of components rendered ineffective by the power-law model. This occurred because all components in a stimulus were in competition with one another to determine the firing rate of the cell. Thus for example, when the 3 nonoptimal components (*fas*) were added to the preferred component (*P*), the model response was reduced, mimicking the behavior of the MT cell (asterisk). Similarly, the modeled response (diamond) to the stimulus composed of flicker (*f*) and antipreferred (*A*) components was correctly intermediate between the responses to the components alone.

The competition model explained more of the variance in the data, but compared with the weighted power-law model, it had 4 extra free parameters. This may have accounted for some of the increase in explained variance. We therefore used the Akaike Information Criterion (AIC) to determine the better model; this criterion corrects for the different number of free parameters. We used an AIC threshold that ensures that a model with more free parameters is considered better only if the increase in information the model provides is warranted by the increase in free parameters (see METHODS; Burnham and Anderson 1998).

Figure 8 compares the performance of the two models on a cell-by-cell basis. For 25 cells (46%), the competitive interaction model was a better description than the power-law model. The power-law model was a more parsimonious description for only 15 cells. The AIC analysis was inconclusive for the remaining 14 cells. This confirms that for many cells, the competition model is truly a better way to describe the data.

More important, we wish to illustrate *why* the competition model is capable of capturing the complexity that we observed in Fourier space summation. Figure 9 illustrates the competition model with the 2-component composite stimulus responses of the cells we have used as examples in Figs. 6 and 7. Figure 9, *A* and *C* show the actual responses of the example cells to each of its 2-component composite stimuli, and compares these to the responses to the individual components as well as the model responses. Figure 9, *B* and *D* show the weights of the best-fitting competition model for the 2 cells.

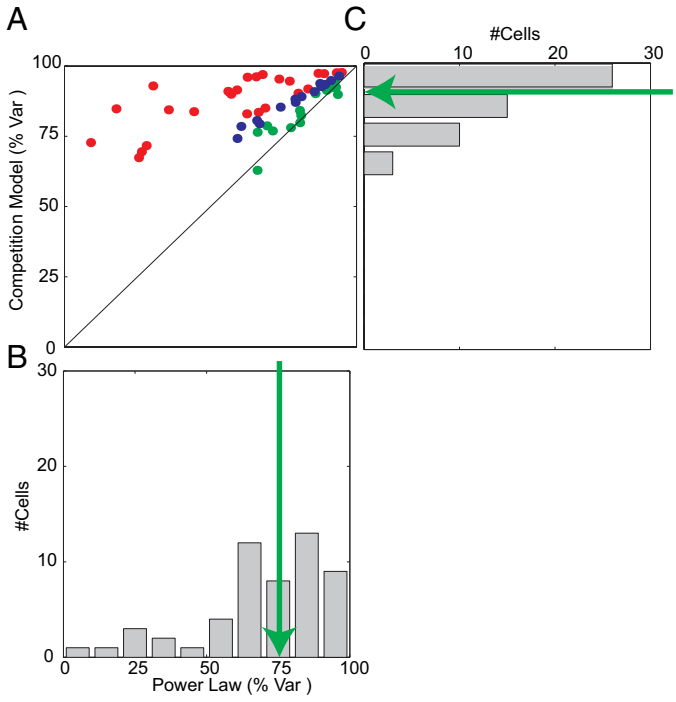

FIG. 8. Population comparison of the models. *A*: comparison of explained variance in the power-law and competition models. For nearly all cells, the competition model explains more variance. Colors represent the outcome of the Akaike Information Criterion (AIC) analysis. Red dots represent cells for which the competition model was truly better according to the AIC. Green dots represent cells where the power-law model was a more parsimonious description of the data. Blue dots represent cells for which the AIC analysis was inconclusive. *B*: histogram of the explained variance in the power-law model. *C*: histogram of the explained variance in the competition model.

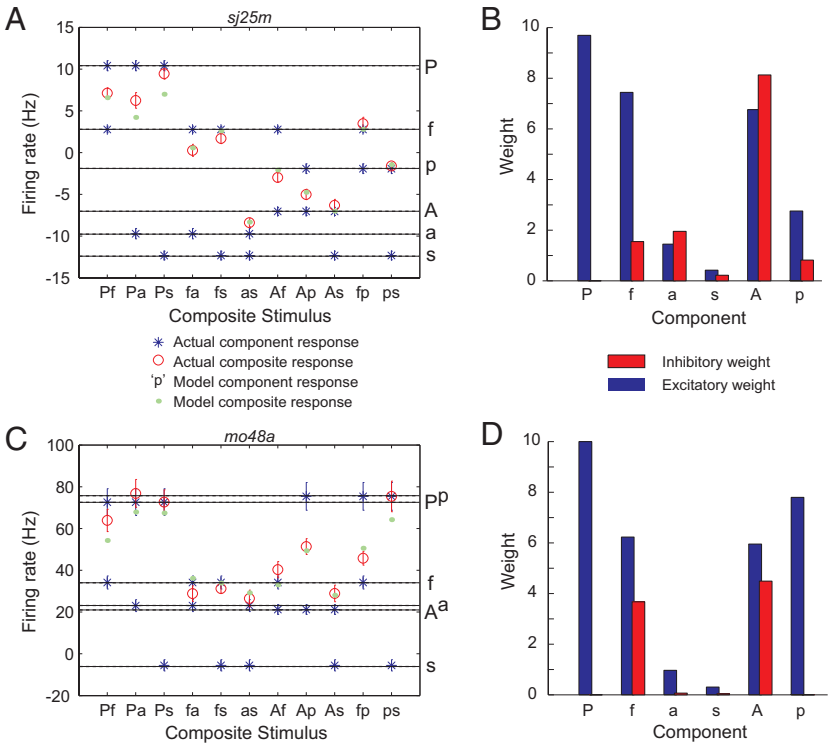

FIG. 9. The competition model. *A*: A single cell's average responses (corrected for baseline activity) are shown on the vertical axis. Red circles refer to the *actual* response to a composite stimulus. Green dot shows the *model* response. Blue asterisks represent the *actual* responses to either of the components present in any given composite stimulus. Dashed lines (and the corresponding labels on the *right side* of the panel) identify the *model* responses to individual components. Error bars are SEs in the mean and are sometimes smaller than the data points themselves. *B*: excitatory (W_j^+; blue) and inhibitory (W_j^-; red) weights in the best-fitting competition model for the cell shown in *A*. *C*: second example, based on the cell shown in Fig. 6. *D*: competition weights for the cell shown in *C*. See text for discussion.

The first example cell (shown in Fig. 9, *A* and *B*), also shown in Fig. 7*B*, shows that for this cell, the component *P* has a large excitatory weight but a negligible inhibitory weight; thus this component will tend to evoke the maximal response in this cell. This is confirmed in *A*; the dashed line representing the component response of the model to component *P* is the maximum response and it is close to the actual *P* component response (blue asterisk). The *f* component has significant excitatory and inhibitory weights; the balance between the two determines the response when this component is presented in isolation. When a component's inhibitory weight is sufficiently large, the response to the component alone will be reduced relative to the baseline. In this cell, this is the case for both the *a* and the *A* components. As can be seen in Fig. 9*A*, the reduction from baseline, however, is larger for *a* than for *A*. Naively, one would expect this larger reduction from baseline in isolation to cause a larger reduction from baseline when combined with another component. In this cell this is clearly not the case: when *a* is added to *f* to create the composite *af*, the rate is reduced somewhat, but when *A* is added to *f* to create composite *Af*, the rate is reduced well below baseline (Fig. 9*A*). The model captures this behavior. The reason that the model can do so is that the response to a single component is determined essentially by the *ratio* between the excitatory and inhibitory weights and not their absolute values. When other components are added to the stimulus, however, the *absolute* weight values do play a role, and components with large weights (in this case *A*) have greater influence. This also explains the role of the stationary component (*s*). Presented in isolation, the combination of a small excitatory and inhibitory weight induces a very small response that is significantly below baseline firing. When the *s* component is combined with other components, however, it hardly has any influence at all.

Figure 9, *C* and *D* illustrate the competition model for the example cell in Fig. 6. The most interesting comparison here is between composites *Pa* and *Ap*. This cell's responses to the components *P* and *p* were very similar and so were the responses to *a* and *A*. In Fig. 9*C* this is shown by the location of the blue asterisks for these components (actual component responses) as well as the dashed lines (model component responses). As pointed out earlier, any summation model that uses only the component responses to predict the composite responses would predict almost equal responses to *Pa* and *Ap*. The cell's actual responses, however, were very different and the model captured this. The ratio of the excitatory to inhibitory weights was approximately the same for the *A* and *a* components, but because the absolute value of the *A* component weights were much larger, they had a much stronger suppressive influence in a composite stimulus.

DISCUSSION

Our data show that macaques, just like humans, perceive the reverse-phi illusion and that a neural correlate of the illusion can be observed in the responses of neurons in cortical visual area MT. This is compatible with the interpretation that Fourier motion energy—not the tracking of features—drives the responses of MT cells. Because most natural stimuli include multiple Fourier components, however, a motion mechanism must specify not only what the relevant signals are, but also how these are combined. Our second set of experiments showed that the responses of an MT cell to a stimulus consisting of multiple Fourier components cannot be explained solely on the basis of the *responses* to those components. In other words, two components that evoke the same response when presented alone can have very different effects when either of them is combined with a third component. Such complex

interactions are well described by a competitive mechanism operating among the Fourier components. This competition can be seen as a natural extension of the subtractive motion opponency that Adelson and Bergen (1985) proposed on the basis of psychophysical data.

Here we will first discuss the neural correlates of reverse-phi, and then what the neural circuitry outside MT may contribute to the pattern of responses we found. Last, we will discuss what the phenomenon of reverse-phi tells us about the mechanisms underlying motion perception.

Neural correlates of reverse-phi

Previous reports have demonstrated that directional neural responses reverse with contrast reversals in cat V1 (Emerson et al. 1987), the nucleus of the optic tract of the wallaby (Ibbotson and Clifford 2001) and macaque V1 (Livingstone and Conway 2003), and MT (Livingstone et al. 2001). Given the nature of the animal subjects or the stimuli used in these studies, however, a direct link between the neural responses and the perceptual phenomenon of reverse-phi could not be established. The studies of Emerson et al. and Livingstone et al. used one-dimensional random white-noise stimuli in combination with a reverse-correlation technique to determine the underlying spatiotemporal structure ("kernels") of the receptive fields. They showed that these kernels are oriented in space–time, which is consistent with the motion–energy model, and that these kernels are inverted when the contrast-sign is inverted, consistent with the reverse-phi phenomenon. The white-noise reverse-correlation technique provides a detailed look at the structure of the spatiotemporal receptive field but, because white-noise stimuli do not lead to unequivocal phi or reverse-phi motion percepts, the link with the perceptual phenomenon remains tentative. The stimuli in our study do lead to a clear percept of motion, which allowed us to record both behavioral and neural responses and establish a direct link between them.

How is it implemented?

By themselves our data tell us something about only the end result of all processing that leads to the response of an MT cell. This surely includes considerable processing in the retina, the LGN, motion processing in V1 and within MT, and could in principle include signals from many other areas. Because we recorded only the stimulus and the MT response, the weights in the models describe the weights between the stimulus space and the MT cell, not a synaptic weight connecting a V1 cell to an MT cell. In other words, it describes the computation that takes place in the whole network, not its implementation. The extent to which the competitive interactions arise from earlier areas or de novo in MT can be investigated in future work by measuring competitive interactions in V1 as well as simultaneous recordings in MT and V1.

One well-known model of the circuitry needed to achieve directional selectivity makes use of shunting inhibition to veto (Barlow and Levick 1965) the dendritic propagation of depolarization in the antipreferred direction (Koch and Poggio 1985). Recently Mo and Koch (2003) extended this model to include reverse-phi motion selectivity. This model combines same-contrast shunting inhibition with an opposite-contrast shunting inhibition of a stimulus moving in the preferred direction. This antisymmetric coupling of ON and OFF pathways results in directional responses that reverse with contrast reversals. Mo and Koch did not report how their model cells respond when more complex stimuli impinge on the dendritic tree. Intuitively, however, it seems possible that a complex spatiotemporal distribution of luminance increments and decrements could result in an equally complex pattern of facilitation and shunting inhibition. Part of the complex interactions among Fourier components that we observed in MT may therefore already arise at an earlier level that implements directional selectivity with the veto mechanism.

A competitive motion energy model

The first stage of the classic motion–energy model consists of space–time-inseparable filters that estimate motion–energy in a particular direction (Adelson and Bergen 1985). Such filters have been demonstrated psychophysically in humans (Burr et al. 1986) and the properties of complex cells in V1 match this stage of the motion–energy model (Conway and Livingstone 2003; Emerson 1997; Emerson et al. 1992). Psychophysical evidence, however, suggests that human motion perception also relies on opponent mechanisms (Stromeyer et al. 1984). Adelson and Bergen (1985) suggested that such mechanisms could simply result from a later stage of processing—the opponent stage—in which the responses of two filters that measure motion energy in opposite directions are subtracted. In the macaque, motion opponency is a prominent feature of many MT cells (Snowden et al. 1991). This can also be seen in our data; the cell in Fig. 9*A*, for instance, reduces its firing rate below baseline when stimulated in its antipreferred direction. Functional imaging data show that human MT+ also has strong motion opponency (Heeger et al. 1999) and this property has been incorporated into neural models of MT (Simoncelli and Heeger 1998).

However, our data as well as previous reports on single-cell responses in area MT (Snowden et al. 1991), and human psychophysical performance on motion discrimination (Georgeson and Scott-Samuel 1999; Mather and Molden 1983), are not compatible with opponency as a simple subtraction of opposite-motion directions. To resolve this, previous models have hypothesized competitive interactions among multiple-direction detectors. This can explain the existence of the orthogonal motion aftereffect (Grunewald and Lankheet 1996), and it has been used in network models to enhance directional responses around the true motion direction, while nevertheless allowing activity representing other directions to remain in the network (Nowlan and Sejnowski 1995; Wilson et al. 1992). With the competition model, we introduced a soft competition for very similar reasons. The soft competition allows the model to respond in a fine-grained manner to small changes in the spatiotemporal visual pattern, just like MT cells do (Fig. 7). The competitive interaction among multiple Fourier components that we propose can therefore be interpreted as a further generalization of multiple-direction opponency (Grunewald and Lankheet 1996; Snowden et al. 1991) as well as motion contrast (Georgeson and Scott-Samuel 1999) to include competition among all Fourier components, even those that carry no directional motion energy (stationary or flicker components).

Figure 9 reveals properties of the model that go beyond soft competition. The data revealed that that, even if one knows an MT cell's response to every single Fourier component in a complex stimulus, one still cannot predict the response to the complex stimulus. The competition model captures this; a component can have a strong influence when presented alone, but nevertheless a negligible influence when combined with other components. The particular components for which we found this behavior differed per cell and we have not been able to find a general competition rule to summarize the population. This may have been expected, given the fact that we used the same set of 6 Fourier components in each experiment, whereas the cells presumably had different spatiotemporal frequency tunings. Sampling only 6 of the relevant points in a large space makes finding a common pattern of competitive interactions unlikely. To uncover a common pattern, we believe it to be necessary to measure interactions across a large part of Fourier space and additionally for components moving in other than the preferred or antipreferred directions. The combinatorial explosion inherent in a factorial design that extends *Experiment II* along these lines, however, has so far prevented us from attempting this.

Other models

The flexibility of the power-law model allowed us to cover a wide range of models from simple linear summation to winner-take-all interactions. None of these models, however, worked satisfactorily. Although it is tempting to believe that a slightly different nonlinear summation model could have worked, we believe this not to be the case. The most convincing evidence for this is presented in Fig. 6 and Fig. 9, *C* and *D*. Any model that relies only on the responses to the components will fail to explain the large difference in the response to 2 stimuli (such as *Pa* and *pA*) that contain only components that evoke identical responses.

A minimum prerequisite for a successful model therefore is that it allows for first-order effects (the component influences the response) but also second-order effects (the influence of one component depends on which other components are present). The competition model we propose is a model that meets this prerequisite. Interestingly, it makes use of divisive inhibition to do so. This is reminiscent of the Simoncelli and Heeger model of MT cells, in which divisive inhibition serves to normalize the response of the cells to the overall contrast in the stimulus. The main difference with the divisive inhibition we used is that, in the competitive motion–energy model, inhibition is tuned in Fourier space. As such, the divisive inhibition in our model is part of the computation that leads to direction selectivity and not a gain control or normalization mechanism that corrects for the total amount of contrast in the stimulus. The similarities in the descriptive models, however, provide further evidence that divisive inhibition may be a general mechanism underlying many computations in the visual system (Heeger 1992).

ACKNOWLEDGMENTS

We thank S. Anstis, K. Britten, K. Dobkins, G. Horwitz, J. Reynolds, and A. Schlack for useful discussions and/or comments on the manuscript. J. Costanza, D. Diep, and L. Abavare provided superb technical assistance.

GRANTS

B. Krekelberg was supported by the Human Frontiers Science Program (LT-50/2001-B). T. D. Albright is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

Adelson EH and Bergen JR. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2: 284–299, 1985.

Albright TD. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52: 1106–1130, 1984.

Anstis SM. Phi movement as a subtraction process. *Vision Res* 10: 1411–1430, 1970.

Anstis SM and Rogers BJ. Illusory reversal of visual depth and movement during changes of contrast. *Vision Res* 15: 957–961, 1975.

Barlow HB and Levick WR. The mechanism of directionally selective units in rabbit's retina. *J Physiol* 178: 477–504, 1965.

Britten KH and Heuer HW. Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19: 5074–5084, 1999.

Britten KH, Newsome WT, Shadlen MN, Celebrini S, and Movshon JA. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci* 13: 87–100, 1996.

Burnham KP and Anderson DR. *Model Selection and Inference—A Practical Information-Theoretic Approach.* New York: Springer-Verlag, 1998.

Burr DC, Ross J, and Morrone MC. Seeing objects in motion. *Proc R Soc Lond B Biol Sci* 227: 249–265, 1986.

Chubb C and Sperling G. Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proc Natl Acad Sci USA* 86: 2985–2989, 1989.

Churchland MM and Lisberger SG. Shifts in the population response in the middle temporal visual area parallel perceptual and motor illusions produced by apparent motion. *J Neurosci* 21: 9387–9402, 2001.

Conway BR and Livingstone MS. Space–time maps and two-bar interactions of different classes of direction-selective cells in macaque V-1. *J Neurophysiol* 89: 2726–2742, 2003.

Dobkins KR and Albright TD. What happens if it changes color when it moves?: the nature of chromatic input to macaque visual area MT. *J Neurosci* 14: 4854–4870, 1994.

Dodd JV, Krug K, Cumming BG, and Parker AJ. Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *J Neurosci* 21: 4809–4821, 2001.

Emerson RC. Quadrature subunits in directionally selective simple cells: spatiotemporal interactions. *Vis Neurosci* 14: 357–371, 1997.

Emerson RC, Bergen JR, and Adelson EH. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Res* 32: 203–218, 1992.

Emerson RC, Citron MC, Vaughn WJ, and Klein SA. Nonlinear directionally selective subunits in complex cells of cat striate cortex. *J Neurophysiol* 58: 33–65, 1987.

Georgeson MA and Scott-Samuel NE. Motion contrast: a new metric for direction discrimination. *Vision Res* 39: 4393–4402, 1999.

Grossberg S. Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Stud Appl Math* 52: 213–257, 1973.

Grunewald A and Lankheet MJM. Orthogonal motion after-effect illusion predicted by a model of cortical motion processing. *Nature* 384: 358–361, 1996.

Heeger DJ. Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9: 181–197, 1992.

Heeger DJ, Boynton GM, Demb JB, Seidemann E, and Newsome WT. Motion opponency in visual cortex. *J Neurosci* 19: 7162–7174, 1999.

Heuer HW and Britten KH. Contrast dependence of response normalization in area MT of the rhesus macaque. *J Neurophysiol* 88: 3398–3408, 2002.

Ibbotson MR and Clifford CW. Interactions between ON and OFF signals in directional motion detectors feeding the NOT of the wallaby. *J Neurophysiol* 86: 997–1005, 2001.

Judge SJ, Richmond BJ, and Chu FC. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res* 20: 535–538, 1980.

Koch C and Poggio T. The synaptic veto mechanism: does it underlie direction and orientation selectivity in visual cortex? In: *Models of the Visual Cortex*, edited by Rose D and Dobson VG. Chichester, UK: Wiley, 1985, p. 408–419.

Livingstone MS and Conway BR. Substructure of direction-selective receptive fields in macaque V1. *J Neurophysiol* 89: 2743–2759, 2003.

Livingstone MS, Pack CC, and Born RT. Two-dimensional substructure of MT receptive fields. *Neuron* 30: 781–793, 2001.

Lu ZL and Sperling G. Second-order reversed phi. *Percept Psychophys* 61: 1075–1088, 1999.

Maruya K, Mugishima Y, and Sato T. Reversed-phi perception with motion-defined motion stimuli. *Vision Res* 43: 2517–2526, 2003.

Mather G and Moulden B. Thresholds for movement direction: two directions are less detectable than one. *Q J Exp Psychol A* 35: 513–518, 1983.

Mo CH and Koch C. Modeling reverse-phi motion-selective neurons in cortex: double synaptic-veto mechanism. *Neural Comput* 15: 735–759, 2003.

Nowlan SJ and Sejnowski TJ. A selection model for motion processing in area MT of primates. *J Neurosci* 15: 1195–1214, 1995.

Reynolds JH, Chelazzi L, and Desimone R. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19: 1736–1753, 1999.

Schoppmann A and Hoffmann K-P. Continuous mapping of direction selectivity in the cat's visual cortex. *Neurosci Lett* 2: 177–181, 1976.

Simoncelli EP and Heeger DJ. A model of neuronal responses in visual area MT. *Vision Res* 38: 743–761, 1998.

Snowden RJ, Treue S, Erickson RG, and Andersen RA. The response of area MT and V1 neurons to transparent motion. *J Neurosci* 11: 2768–2785, 1991.

Stromeyer CF 3rd, Kronauer RE, Madsen JC, and Klein SA. Opponent-movement mechanisms in human vision. *J Opt Soc Am A* 1: 876–884, 1984.

Thiele A, Dobkins KR, and Albright TD. Neural correlates of contrast detection at threshold. *Neuron* 26: 715–724, 2000.

Wertheimer M. *Classics in Psychology*, edited by Shipley T. New York: Philosophical Library (originally published in 1912), 1961, p. 1032–1089.

Wilson HR, Ferrera VP, and Yo C. A psychophysically motivated model for two-dimensional motion perception. *Vis Neurosci* 9: 79–97, 1992.