Behavioral/Systems/Cognitive

The Neural Representation of Speed in Macaque Area MT/V5

Nicholas J. Priebe, Carlos R. Cassanello, and Stephen G. Lisberger

Howard Hughes Medical Institute, Department of Physiology, W. M. Keck Foundation Center for Integrative Neuroscience, the Neuroscience Graduate Program, and the Sloan-Swartz Center for Theoretical Neurobiology, University of California, San Francisco, California 94143

Tuning for speed is one key feature of motion-selective neurons in the middle temporal visual area of the macaque cortex (MT, or V5). The present paper asks whether speed is coded in a way that is invariant to the shape of the moving stimulus, and if so, how. When tested with single sine-wave gratings of different spatial and temporal frequencies, MT neurons show a continuum in the degree to which preferred speed depends on spatial frequency. There is some dependence in 75% of MT neurons, and the other 25% maintain speed tuning despite changes in spatial frequency. When tested with stimuli constructed by adding two superimposed sine-wave gratings, the preferred speed of MT neurons becomes less dependent on spatial frequency. Analysis of these responses reveals a speed-tuning nonlinearity that selectively enhances the responses of the neuron when multiple spatial frequencies are present and moving at the same speed. Consistent with the presence of the nonlinearity, MT neurons show speed tuning that is close to form-invariant when the moving stimuli comprise square-wave gratings, which contain multiple spatial frequencies moving at the same speed. We conclude that the neural circuitry in and before MT makes no explicit attempt to render MT neurons speed-tuned for sine-wave gratings, which do not occur in natural scenes. Instead, MT neurons derive form-invariant speed tuning in a way that takes advantage of the multiple spatial frequencies that comprise moving objects in natural scenes.

Key words: direction tuning; speed tuning; MT; visual cortex; visual motion processing; spatial frequency

Introduction

Many of our behaviors depend on accurate information about changes in our environment. To make appropriate responses for a moving object, like catching a baseball in a glove, we need accurate information not only about the shape of the object, but also about its motion. Ideally, the shape of the object should not interfere with the estimation of its motion, and its location or motion should not cause errors in identifying the object. We have studied the influence of object form on visual motion processing in the extrastriate middle temporal visual area (MT).

In the visual system, it has been traditional to think of motion by characterizing the visual scene according to its spatial and temporal sine-wave components in Fourier space. For example, sine-wave gratings are characterized by "spatial frequency," defined in cycles per degree as the inverse of the width of a single cycle of the grating, and "temporal frequency," defined in cycles per second as the inverse of the time required for the intensity of a single pixel to go through a full cycle of sinusoidal modulation. The speed of a moving grating is the ratio of the temporal frequency and the spatial frequency. Although sine-wave gratings are commonly used in the laboratory setting to assess the response properties of neurons, moving real-world objects are

more complex than sine-wave gratings and contain multiple spatial and temporal frequencies. In the present paper, we have tested whether motion processing by the brain is different for real objects containing a broad spectrum of spatial and temporal frequencies, rather than for the unnatural grating stimuli used in the laboratory.

In principle, motion-sensitive neurons could be truly tuned for speed, meaning that the tuning is independent of the form of the moving stimulus (Movshon, 1975; Tolhurst and Movshon, 1975). Then, neurons would have the same preferred speed at different spatial frequencies, and temporal frequency tuning would vary as a function of spatial frequency. Alternatively, motion-sensitive neurons could have separate, independent tunings for spatial and temporal frequency. Most models of motion-selective neurons are based on separable responses to spatial and temporal frequency, and early data demonstrated such responses in the primary visual cortex (V1) of the cat (Tolhurst and Movshon, 1975; Holub and Morton-Gibson, 1981; Friend and Baker, 1993). In an earlier study of this question, Perrone and Thiele (2001) concluded that neurons in visual area MT are tuned for speed.

In the present paper, we show that the neural processing of speed is both more complex and more interesting than implied by either of the alternatives outlined above. First, by correcting a flaw in the data analysis of Perrone and Thiele (2001), we show that only a minority of MT neurons are speed-tuned in the sense that preferred speed is independent of spatial frequency. Second, we demonstrate that speed tuning depends less on spatial frequency for stimuli constructed by adding two sine-wave gratings. Speed tuning results from a nonlinearity that facilitates or sup-

presses responses when the two component gratings have the same or different speeds. We conclude that the absence of true speed tuning in area MT for sine-wave gratings does not pose a problem for representing the speed of real-world objects because they contain many spatial frequencies.

Materials and Methods

Physiological preparation. Extracellular single-unit microelectrode recordings were made in the MT of nine anesthetized, paralyzed monkeys (*Macaca fascicularis*). Anesthesia was induced with ketamine (5–15 mg/kg) and midazolam (0.7 mg/kg), and cannulae were inserted into the saphenous vein and the trachea. The animal's head was then fixed in a stereotaxic frame and the surgery was continued under an anesthetic regimen of isoflurane (2%) inhaled in oxygen. A small craniotomy was performed and the dura was reflected directly above the superior temporal sulcus (STS). The animal was maintained under anesthesia using an intravenous opiate, sufentanil citrate (8–24 $\mu g \cdot kg^{-1} \cdot hr^{-1}$), for the duration of the experiment. To minimize drift in eye position, paralysis was maintained with an infusion of vecuronium bromide (Norcuron, 0.1 $\mu g \cdot kg^{-1} \cdot hr^{-1}$; Oragnon, West Orange, NJ) and the animal was artificially ventilated with medical-grade air. The body temperature was kept at 37°C with a thermostatically controlled heating pad. The electrocardiogram, electroencephalogram, autonomic signs, and rectal temperature were monitored continuously to ensure the anesthetic and physiological state of the animal. The pupils were dilated using topical atropine and the corneas were protected with +2 diopters gas-permeable hard contact lenses. Supplementary lenses were selected by direct ophthalmoscopy to make the lens conjugate with the display. The locations of the foveae were recorded using a reversible ophthalmoscope.

Tungsten-in-glass electrodes (Merrill and Ainsworth, 1972) were introduced by a hydraulic microdrive into the anterior bank of the STS and were driven down through the cortex and across the lumen of the STS into the MT. The location of unit recordings in the MT was confirmed by histological examination of the brain after the experiment, using methods described by Lisberger and Movshon (1999). After the electrode was in place, agarose was placed over the craniotomy to protect the surface of the cortex and to reduce pulsations. Single units were isolated using a dual-time-window discriminator (DDIS-1, BAK Electronics, Germantown, MD) and action potentials were amplified conventionally and displayed on an oscilloscope. Both a filtered version of the neural signals and a tone indicating the acceptance of a waveform as an action potential were played over a stereo audio monitor, and the time of each accepted waveform was recorded to the nearest 10 μsec for subsequent analysis. The recording sessions lasted between 84 and 120 hr. The units included in this study are from recordings in nine monkeys.

All experiments followed protocols that had received prior approval by the Institutional Animal Care and Use Committee at University of California, San Francisco.

Stimulus presentation. After isolating a single unit in the MT, we mapped its receptive field on a tangent screen by hand. We recorded the spatial position of each receptive field and, for many of the neurons, the size of the minimum response field. All of the neurons reported in this paper had receptive field centers within 12° of the fovea. Visual stimuli were then presented on a video monitor (CCID 121; Barco, Poperinge, Belgium), and were generated by a video frame buffer (Cambridge Research, Kent, UK). The video system had a noninterlaced refresh rate of 100 Hz. The spatial resolution of the monitor was 1024 × 768 pixels and the screen subtended 33.6 cm horizontally and 25.2 cm vertically. Because the monitor was placed 65 cm from the monkey's eyes, there were at least 19 pixels per visual degree. The video monitor always had a mean luminance of 68 candelas per degree. We positioned a mirror so that stimuli presented on the video monitor fell within the receptive field of the isolated neuron.

Experiments consisted of a sequence of brief trials with an intertrial interval of ~700 msec. All trials began with the appearance of a stationary stimulus surrounded by a gray background of the mean luminance. For all trials, the stimulus appeared and was stationary for 250 msec before starting to move. After the motion was completed, the stimulus remained visible for an additional 250 msec. For each neuron, we first assayed the preferred direction by recording the responses to a 1000 msec motion of a 32% contrast sine-wave grating in 16 directions. We then assessed the preferences of the neuron for spatial and temporal frequency by measuring the response to gratings moving in the preferred direction for all combinations of six spatial frequencies (0.125, 0.25, 0.5, 1, 2, and 4 cycles per visual degree) and nine temporal frequencies (0, 0.25, 0.5, 1, 2, 4, 8, 16, and 32 cycles per second). For 5 of 104 neurons, the response to the lowest spatial frequency was not measured. For many neurons, gratings were presented at two contrasts, 32 and 8%, in randomly interleaved trials. Next, we tested the responses of 48 neurons with stimuli that contained two spatially overlapping gratings. Dual-grating stimuli were created by temporally interleaving frames in which each of the gratings were displayed individually. Because the refresh rate of the monitor was 100 Hz, the refresh rate to display both gratings was 50 Hz. To allow fair comparison to the dual gratings, single sine-wave gratings were displayed with a temporal refresh rate of 100 Hz, but frames containing the grating were temporally interleaved with a blank (gray) stimulus of the same mean luminance.

For a subset of 71 neurons, we measured the speed tuning for random dots that moved
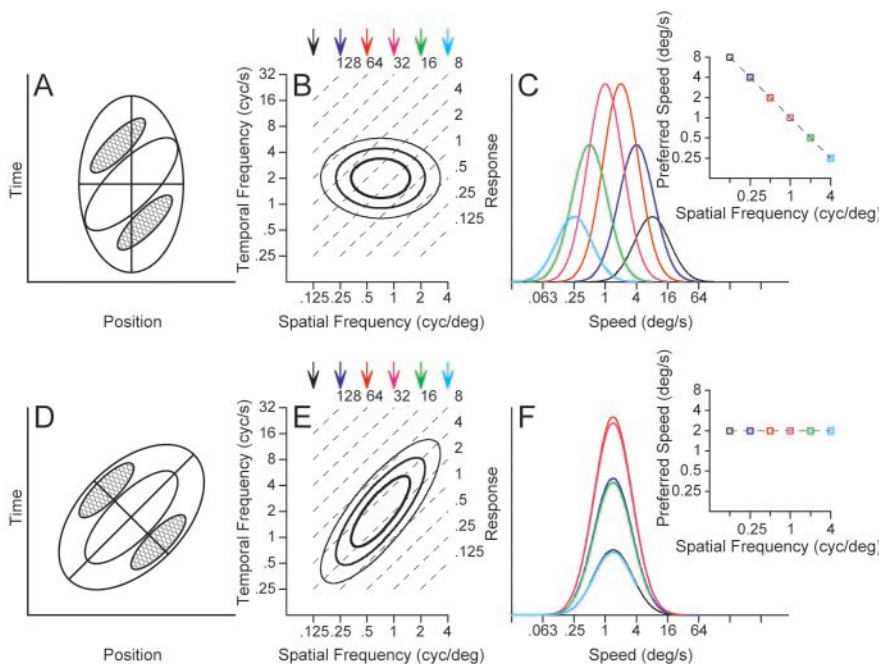
Figure 1. Two models for the creation of motion-sensitive neural responses *A, D,* Response contours indicating the strength of responses as a function of the position of the stimulus and time. Red contours indicate regions of increasing response to dark stimuli, whereas blue contours indicate the response to light stimuli. The response fields in *A* and *D* are diagrams contrived to represent motion filters that would and would not have a dependence of speed tuning on the spatial frequency of sine-wave gratings. *B, E,* The contour lines show the receptive field of each model after transformation into Fourier space. The parallel, dashed lines indicate isospeed contours, with speed indicated by the number at the top right of each line. *C, F,* Each curve shows the speed tuning of the model at one spatial frequency; different colors indicate different spatial frequencies and the colors of the curves are coordinated with those of the arrows above *B* and *E,* The graphs inset at the top right of *C* and *F* plot preferred speed as a function of spatial frequency. *C* and *F* were derived exactly from the contour plots in *B* and *E*.

within a window placed over the receptive field of the neuron (Priebe et al., 2002). Random dots were presented on an analog oscilloscope (models 1304A and 1321B, P4 phosphor, Hewlett-Packard, Palo Alto, CA), using signals provided by digital-to-analog converter outputs from a personal computer (PC)-based digital signal processing board (Spectrum Signal Processing, Burnaby, Canada). This system allows a temporal refresh rate of 500 or 250 Hz and a nominal spatial resolution of 64,000 by 64,000 pixels. The display was positioned 65 cm from the animal and subtended 20° horizontally by 20° vertically. Because of the dark screen of the display, background luminance was beneath the threshold of the photometer, less than 1 mcd/m^2. For all random-dot trials, textures appeared and were stationary for 256 msec before moving for 512 msec. After the motion was completed, the dots remained visible for an additional 256 msec. After identifying the preferred direction of each neuron, we determined its preferred speed by moving the random-dot texture in the preferred direction at 11 speeds: 0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, and 128°/sec.

Data acquisition and analysis. Experiments were controlled by a computer program running on a UNIX workstation and a Windows NT PC running the real-time extension RTX (VenturCom, Waltham, MA). The two computers were networked together: the UNIX workstation provided an interface for programming target motion and customizing it during recording from a neuron; the PC provided real-time control of target motion and data acquisition. The times of spikes were recorded by the PC and sent over the network to the UNIX workstation, which stored them for subsequent analysis, along with codes indicating the target motion that had been commanded. Each experiment consisted of a list of trials, in which each trial presented a different stimulus motion. Trials were sequenced by shuffling the list and presenting the trials in a random order until all the trials in the list had been presented. The list was then shuffled and repeated again until enough repetitions of each stimulus had been obtained. The mean number of repetitions for each trial condition was 15.3 ± 5.27 (mean ± SD).

Data were analyzed by aligning all the responses to identical trajectories of grating motion on the onset of motion and accumulating poststimulus time histograms with a bin width of 1 msec. For presentation, the histograms were built with a bin width of 50 msec. Background responses were eliminated by creating a histogram for trials that presented a stationary stimulus, computing the mean firing rate during the interval when stimuli for other trials were moving, and subtracting this scalar from the firing rate in every bin of every other histogram. We then measured the firing rate from the background-corrected histograms to determine the stimulus selectivity of each neuron and to quantify the results of each experiment. Fits to the data were made by extracting the firing rate from each cell on a trial-by-trial basis. The extracted firing rates were then passed to a Matlab (MathWorks, Natick, MA) function ("nlinfit") that fit the data using the equations presented in the Results section of the paper. Confidence intervals for parameter estimates were computed from the Jacobian matrix and the residuals using the Matlab function "nlparci." Specific analyses are presented in the relevant places in the Results.

Results

Theoretical basis for the experiments

Most models of motion selectivity are based on a comparison of luminance or contrast signals across time and space. Although many models have been shown to extract the direction of motion accurately, the details of the filters in the models determine whether they signal the speed of motion independently of object shape or object contrast. For example, one of the most influential models of motion selectivity, the motion energy model originally developed independently by Adelson and Bergen (1985) and Watson and Ahumada (1985), can take forms in which speed tuning does (Fig. 1, top) or does not (Fig. 1, bottom) depend on spatial frequency. Figure 1, *A* and *D*, shows diagrams of the response field, in space and time, of two possible configurations of filters that might arise from the motion–energy model. Both configurations are oriented in the space–time coordinate system and
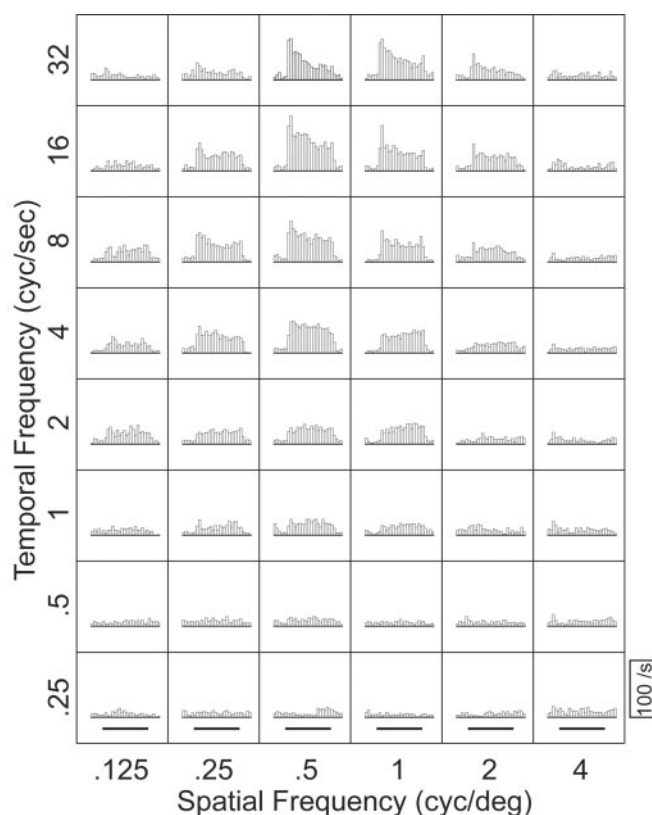
Figure 2. Responses of a representative MT neuron to moving sine-wave gratings with different combinations of temporal and spatial frequency. Each histogram shows the average firing rate in 50 msec bins for multiple repetitions of a single stimulus. The grating was visible throughout each histogram, and the bold horizontal lines below the bottom row of histograms indicate the time when the grating was moving.

would respond selectively to the motion of an object from the right to the left. However, the envelope of the filter in Figure 1*A* is not oriented in space–time, whereas that in Figure 1*D* is oriented. When these filters are viewed in Fourier space, they are accordingly nonoriented as in Figure 1*B* and oriented as in Figure 1*E*. We will refer to the response viewed in frequency domain as oriented and nonoriented motion filters. In Fourier space, for a sinusoidal grating stimulus, speed follows the relationship:

$$\text{Speed} = \frac{\text{Temporal Frequency}}{\text{Spatial Frequency}} \qquad (1)$$

Because Figure 1, *B* and *E*, uses logarithmic axes, the contour lines of equal speed are parallel to one another (dashed lines).

The Fourier transform of a nonoriented motion filter (Fig. 1*B*, contour lines) reveals independence along the spatial and temporal frequency axes: the temporal frequency that gives the best response is the same at each spatial frequency, and vice versa. As a consequence, speed tuning for sine-wave grating stimuli changes as a function of spatial frequency. As shown in Figure 1*C*, the best response occurs at a speed of 8°/sec when the stimulus has a low spatial frequency of 1/8 cycles per degree. As spatial frequency increases, the preferred speed decreases: when the spatial frequency is 4 cycles per degree, the best response occurs at 0.25°/sec. Thus, for each subunit of the motion–energy model, the preferred speed varies over a 32-fold range as a function of spatial frequency. In contrast, the Fourier transform of the oriented motion filter is tilted (Fig. 1*E*) so that its spatial and temporal frequency tunings are not independent: as the spatial fre-
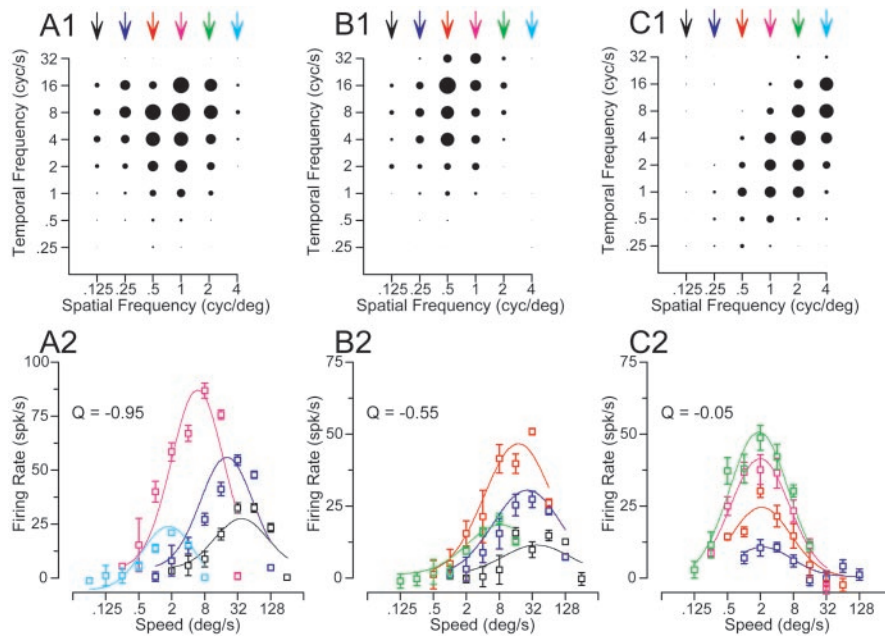
Figure 3. Effect of spatial frequency on the preferred speed of three MT neurons chosen to indicate the diversity of effects. *A1–C1*, Response fields in the coordinate system of temporal and spatial frequency. Each symbol indicates the response to one combination of spatial and temporal frequency, and the diameter of the symbol gives the size of the response. *A2–C2*, Symbols plot the response of each neuron as a function of speed; the different colors indicate gratings of different spatial frequencies. Curves show the result of fitting the data with Equations 2 and 3. As before, the colors are coordinated for the symbols and curves in the bottom row of graphs and the arrows above the top row of graphs. Error bars indicate SEM of the firing rate.

quency increases, the preferred temporal frequency also increases. In Fourier space, the long axis of the response field now lies on a contour of constant speed instead of crossing different speed contours for each spatial frequency. When the speed tuning is measured for each spatial frequency, shown in Figure 1*F*, the speed tuning curves overlap one another; thus, the preferred speed does not vary as the spatial frequency is changed.

In this paper, we will first ask whether the speed tuning of MT neurons resembles better the predictions of the model in the top or bottom row of Figure 1 by quantifying the relationship between preferred speed and spatial frequency of MT neurons: stimuli will consist of sine-wave gratings chosen to tile the parameter space of spatial and temporal frequency. We will next demonstrate that the responses to single gratings underestimate the true speed tuning of MT neurons: stimuli will consist of spatially overlapping pairs of sine-wave gratings, square-wave gratings, and textures.

Speed tuning of MT neurons for single sine-wave gratings
We recorded the responses of 118 MT neurons to gratings moving in the preferred direction; 104 neurons provided adequate data to allow a proper estimation of the response field in Fourier space. Each combination of spatial and temporal frequency yielded a single histogram, like those shown in Figure 2, which summarizes the responses of one MT neuron to all combinations of six spatial frequencies and eight temporal frequencies. Inspection of Figure 2 reveals that the largest responses occur for spatial frequencies of 0.5–1 cycles per degree and temporal frequencies of 8–32 cycles per second. However, inspection of the family of histograms does not reveal whether the response field is separable (Fig. 1*B*), oriented (Fig. 1*E*), or in between.

We created response fields from the mean response to each combination of spatial and temporal frequency for each neuron

like those shown for three representative neurons in Figure 3*A1–C1*. In these graphs, the amplitude of each steady-state response of the neurons to the combinations of spatial and temporal frequency is indicated by the diameter of the symbol. For the neuron in Figure 3*A1*, the response field is similar to the model filter shown in the top row of Figure 1: the spatial and temporal frequency profiles appear to be approximately independent. The same data have been plotted differently in Figure 3*A2*, now showing response as a function of stimulus speed separately for each spatial frequency. As expected for a neuron with a nonoriented response field, the preferred speed of the neuron changed as a function of spatial frequency: the preferred speed varied from 40 to 1.5°/sec, as the spatial frequency of the moving stimulus was changed from 0.125 to 4 cycles per degree. The data in Figure 3*B*, from the same neuron shown in Figure 2, show a somewhat tilted spatiotemporal response field (Fig. 3*B1*), and the plots of firing rate as a function of speed (Fig. 3*B2*) reveal that the preferred speed of the neuron decreased from 36 to 8°/sec as the spatial frequency of the moving stimulus increased from 0.125 to 1 cycles per degree. Finally, the neuron in Figure 3*C* shows an oriented spatiotemporal response field (Fig. 3*C1*) and no relationship between preferred speed and spatial frequency (Fig. 3*C2*). The speed tuning curves in Figure 3*C2* show different amplitudes at different spatial frequencies, but peak at the same speed for each spatial frequency.

To quantify the dependence of speed tuning on spatial frequency, we used a variant of a two-dimensional Gaussian on logarithmic axes to fit spatiotemporal response fields like that illustrated in Figure 3, *A1* to *C1*:

$$R(sf, tf) = A*e^{\frac{-(\log_2(sf)-\log_2(sf_0))^2}{\sigma_{sf}^2}}*e^{\frac{-(\log_2(tf)-\log_2(tf_p(sf)))^2}{\sigma_{tf}^2}}$$

(2)

where tf_p depends on spatial frequency and is defined as:

$$tf_p(sf) = 2^{(Q+1)*(\log_2(sf)-\log_2(sf_0))+\log_2(tf_0)}$$

(3)

Fits and statistical significance were derived from the responses to the full set of individual presentations of each stimulus, whereas the variance accounted for by the fits was based on how well they predicted the mean responses for each stimulus. Equations 2 and 3 have the advantage that they use a single equation to fit all the data, quantifying the slope of the relationship between preferred speed and spatial frequency using a single parameter, the exponent Q. When Q is 0, there is no relationship between spatial frequency and the preference of a neuron for speed, indicating that the neuron has speed tuning that is independent of spatial frequency (Fig. 1*F*, inset). When Q is −1, there is a strong dependence of the preferred speed on the spatial frequency: as the spatial frequency is increased by a log unit, the preferred speed of the neuron is decreased by a log unit (Fig. 1*C*, inset). A Q value of −1 indicates that the spatial and temporal frequency tunings of the
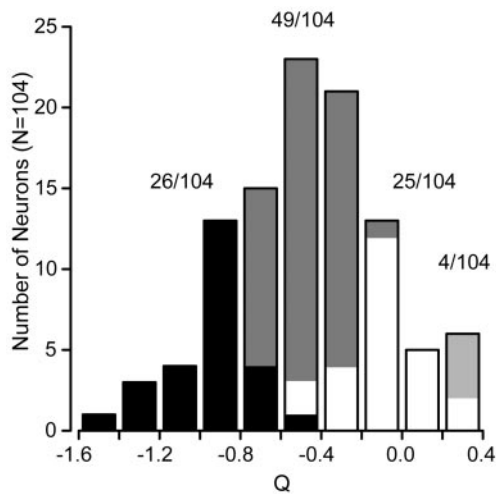
Figure 4. Summary of the effect of spatial frequency on preferred speed across the population of MT neurons. The histogram plots the distribution of the value of Q (Eq. 2) for all 104 neurons in our sample. A Q value of -1 indicates spatial and temporal frequency independence. A Q value of 0 indicates no relationship between spatial frequency and preference for speed. The dark bars indicate neurons whose 95% confidence intervals for Q overlapped with -1. The white bars indicate neurons whose 95% confidence intervals for Q overlapped with 0. Gray bars indicate the neurons whose confidence intervals lie between -1 and 0, whereas the light gray bars indicate neurons whose Q values and confidence intervals were >0. The values above the corresponding portions of the histogram indicate the number of cells falling into each classification.

neuron are independent. The Q value assumes that the interaction between spatial frequency and preferred speed is linear in logarithmic space, following a power law in linear frequency space. For the example neurons shown in Figure 3A–C, Q was -0.95, -0.55, and -0.05, indicating a strong, medium, and weak dependence of preferred speed on spatial frequency.

The distribution of the parameter Q calculated for our population of 104 MT neurons is unimodal and peaks near the mean Q value of -0.52 (Fig. 4). To compare with other studies, we classified the neurons according to whether the 95% confidence intervals of Q overlapped 0 or -1: if they overlapped -1, then we classified the neuron as "spatiotemporally independent" (26 of 104) (Fig. 4, black bars); if the confidence intervals overlapped 0, we classified the neuron as speed tuned (25 of 104) (Fig. 4, white bars); if Q was between -1 and 0 but the confidence intervals overlapped neither, we called the neuron unclassed (49 of 104) (Fig. 4, gray bars), although it had features of both speed tuning and spatiotemporal independence. A few neurons (4 of 104) had Q values >0 and confidence intervals that did not overlap 0, indicating that their speed tuning shifted with spatial frequency, but in the opposite direction predicted by a spatiotemporal-frequency-independent model. For the remainder of the paper, we have considered these neurons as part of the speed-tuned group. The model defined by Equations 2 and 3 provided excellent fits to the spatial and temporal frequency tuning of MT neurons, accounting for the majority of the variance in their mean responses (94.8 ± 3.6%; mean ± SD).

As additional independent tests of speed tuning we used two alternative analysis methods. First, we refitted the relationships between firing rate and speed (Fig. 3A2–C2) separately for each spatial frequency and then performed regression analysis for the log of the preferred speed as a function of the log of the spatial frequency. This analysis confirmed the validity of the assumption of a linear relationship between the logarithms of the spatial frequency and the preferred speed. It yielded values of Q that were

nearly the same in individual neurons ($r = 0.91$) but were slightly closer to -1 (less speed-tuned) than for the fits based on Equations 2 and 3. Because fitting the data for each spatial frequency separately used more parameters, we did obtain the expected slight improvement in how well the equations fitted the data.

Second, we used a variant of a method devised by (Levitt et al., 1994) for classifying neurons according to where they fall along the axis of spatiotemporal separability. In this method, we fitted the spatial and the temporal frequency tuning of the neuron independently with Gaussian functions on logarithmic axes (Fig. 5A). We then used the independent fits of spatial and temporal frequency to make two predictions of the response of the neuron: (1) a spatiotemporal-frequency-independent prediction computed by taking the outer product of the two tuning curves (Fig. 5B); and (2) a speed-tuned prediction computed by shifting the temporal frequency as a function of spatial frequency so that preferred speed was independent of the speed tuning of the neuron[a] (Fig. 5C). We then assessed whether the actual tuning of the neuron was closer to the speed prediction or the independent prediction by computing the partial correlation of the actual response with each of the simulated responses using the following equations:

$$R_{\text{indep}} = \frac{(r_i - r_s * r_{is})}{\sqrt{(1 - r_s^2)(1 - r_{is}^2)}} \quad (4)$$

$$R_{\text{speed}} = \frac{(r_s - r_i * r_{is})}{\sqrt{(1 - r_i^2)(1 - r_{is}^2)}} \quad (5)$$

where R_{indep} and R_{speed} are the partial correlations of the response field with the independent and speed-tuned predictions, r_i is the correlation of the data with the independent prediction, r_s is the correlation of the data with the speed-tuned prediction, and r_{is} is the correlation of the two predictions. We then plotted R_{speed} as a function of R_{indep} (Fig. 5D) and divided the population of neurons according to whether they were speed-tuned, spatiotemporally independent, or unclassed. The results in Figure 5D are similar to those shown by the distribution of values of Q in Figure 4B: the population included 28 speed-tuned neurons, 25 spatiotemporally independent neurons, and 51 unclassed neurons. Finally, there was a strong correlation between the classifications based on using the correlation analysis and the Q analysis ($r = 0.91$).

Our conclusions about the speed tuning of MT neurons disagree with those in a recent paper of Perrone and Thiele (2001), in which they did similar experiments but claimed that the majority of MT neurons are tuned for speed. We do not have access to their data, but it seems to us that the difference between their conclusion and ours lies solely in the criteria used to assign neurons to the speed-tuned class. We called neurons speed-tuned only if they had a value of Q that was not statistically different from 0. They used a more liberal criterion to classify neurons as speed-tuned, including any neurons that showed any tilt in their spatiotemporal response field. This would have caused most of our unclassed neurons to be classified as speed-tuned, despite the effect of spatial frequency on speed preference for these neurons.

[a]In Levitt et al. (1994), the speed tuning of V2 neurons was evaluated by comparing the correlations of the actual spatial and temporal frequency profile to predictions based on independent versus speed-tuned spatial temporal frequency tuning, as done here. However, to create the speed-tuned prediction they altered the preferred spatial frequency as a function of temporal frequency. In this paper, we made the speed-tuned prediction by the converse procedure of changing the preferred temporal frequency as a function of spatial frequency. This subtle change is important, because changing the peak spatial frequency as a function of temporal frequency does not produce spatial-frequency-independent speed tuning, whereas the converse procedure does.
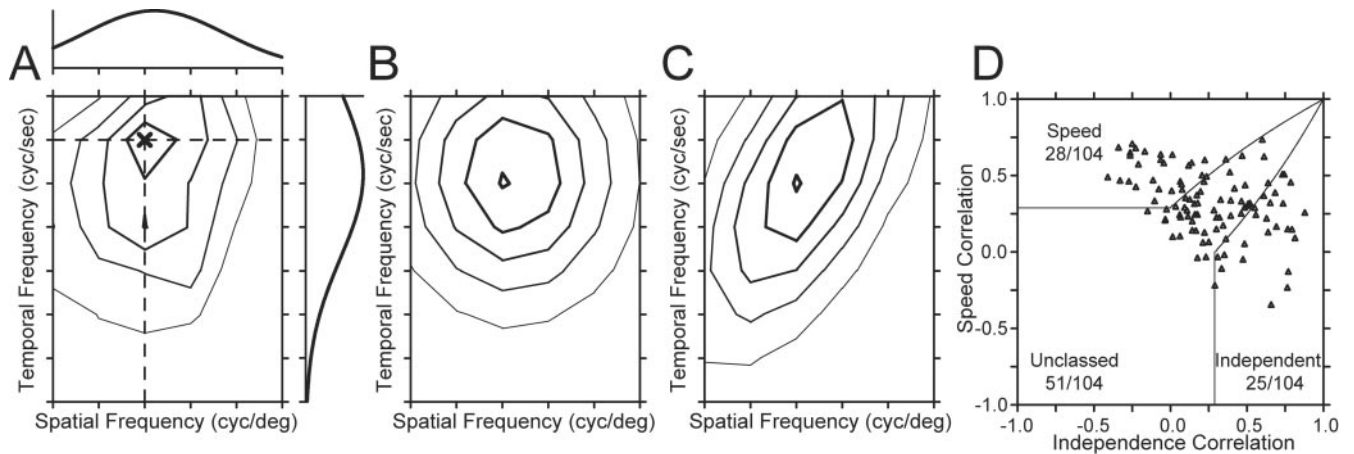
Figure 5. A correlation-based analysis to classify the speed tuning of MT neurons (after Levitt et al., 1994). *A*, Contour plot of the response field of the example neuron shown in Figure 2 and Figure 3*B1,B2*. The spatial and temporal frequency that evoked the peak response is indicated by the dark X. The curve above the contour plot shows the spatial frequency tuning of the neuron at the temporal frequency that elicited the peak response, indicated by the horizontal dashed line. The curve to the right of the contour plot shows the temporal frequency tuning at the spatial frequency that elicited the peak response, shown by the vertical dashed line. *B, C*, Contour plots predicted by using the spatial and temporal tuning in *A* were used to create either a spatiotemporal-frequency-independent model of the neuron (*B*) or a speed-tuned model of the neuron (*C*). *D*, Summary of the correlation of the response field of each neuron with the predictions of models based on spatiotemporal independence (*x*-axis) and speed tuning (*y*-axis). Each symbol summarizes the classification for a different MT neuron in our sample. Solid lines indicate the dividing lines used to characterize neurons as speed-tuned, unclassed, or independent.

A nonlinearity that improves speed tuning of MT neurons for pairs of sine-wave gratings

Cognizant that objects comprise multiple spatial frequencies, we next asked whether speed tuning became form-invariant for stimuli that contained multiple spatial frequencies. We tested MT neurons with dual-grating stimuli consisting of two superimposed gratings of the same orientation. We paired gratings of the same speed from either the higher or lower pairs of four spatial frequencies (Fig. 6*A*, connected by black or gray lines). Predicting the responses to the two gratings in each pair by applying the linear model (adding the responses to each grating presented alone) implied that we should find distinctly different preferred speeds for the different ranges of spatial frequency, as shown by the open symbols and dashed curves in Figure 6*B*. However, recording the responses to dual gratings yielded speed-tuning curves with similar peaks for the different ranges of spatial frequency, as shown by the filled symbols and solid curves in Figure 6*B* (top).

To evaluate this effect in the population of MT neurons, we fitted the predictions and data with Gaussian functions and used the midpoint and σ^2 from the fits to estimate the preferred speed and the tuning width for each function. We then computed the absolute difference in preferred speed between the higher and lower pairs of spatial frequencies for both the linear predictions and the data, and plotted the data from each neuron as a point in Figure 6*D*. All but two of the points fell below the line of slope 1, indicating that the difference in preferred speed for dual-grating stimuli over the two ranges of spatial frequencies was consistently smaller in the data than predicted by the linear model (means of 0.73 and 1.35 octaves, respectively). Figure 6*E* shows that the tuning width for the dual-grating stimuli was narrower in the data than predicted, for almost every neuron. Population averages were: linear prediction, 1.69 ± 0.60; data, 1.28 ± 0.53 (mean ± SD). This experiment indicates the presence of nonlinear mechanisms based on multiple spatial frequencies and suggests that the nonlinearities could create form-invariant speed tuning for real object motion.

To characterize the nonlinear mechanism, we calculated "gain," defined as the actual response to each dual-grating stim-

ulus divided by the prediction from summing the responses to the two gratings singly. Figure 6*B*, bottom, plots the gain separately for the high and low ranges of spatial frequencies: gain was highest near the preferred speed for this neuron (upward black arrow). To summarize the nonlinearity across our sample of MT neurons, we normalized the speed axis in the gain plots for each neuron so that preferred speed had a value of 1. We then averaged the curves that plotted gain as a function of normalized speed for all 30 neurons tested with dual-grating stimuli. The average gain (Fig. 6*C*) was largest near the preferred speed of the neurons and declined as speed moved away from preferred. Furthermore, the average gain curves were similar for the higher and lower ranges of spatial frequencies (Fig. 6*C*, gray vs black curves).

To further investigate the source of the nonlinear interaction that yielded more form-invariant speed-tuning curves, we presented stimuli composed of two overlapping gratings whose spatial and temporal frequencies were adjusted independently according to the design illustrated in Figure 7, *A* and *B*. For each MT neuron presented with the dual-grating stimulus, we first determined its preferred spatial and temporal frequencies. We then chose pairs of spatial and temporal frequencies surrounding its preference, yielding four sine-wave gratings. The way these gratings were selected and combined is indicated by the position of the four histograms for single gratings in Figure 7*A*. Each histogram is placed as if it were in a plot of spatial versus temporal frequency, so that the two histograms outlined in gray had the same speed and, when combined into a dual-grating stimulus produced the "same-speed" response that is also outlined in gray. For this neuron, the actual response to the same-speed dual grating was twice as large as that predicted by adding the mean responses to the two gratings presented singly (horizontal dashed line). The two histograms outlined in black represent responses to two gratings that had the same spatial and frequency components as the same-speed gratings, but now in different combinations so that the two gratings moved at different speeds. For this neuron, the response to the two different-speed gratings (outlined in black) was approximately the same amplitude as that predicted by adding the mean responses to the two gratings presented singly (horizontal dashed line). The example neuron illus-

trated in Figure 7B is more typical of the responses we found. The response to the same-speed dual grating is slightly smaller than predicted by linear summation, whereas the response to the different-speed dual-grating stimulus is half as large as predicted by the linear summation of the mean responses to the two gratings presented singly.

We again quantified the response to the dual gratings by computing the gain, defined as the actual response divided by that predicted by summing the responses to the two component gratings singly. For the population, the gain for the same-speed dual grating was almost always greater than the gain for the different-speed dual-grating stimulus. When the gain along the same-speed axis was plotted as a function of that along the different-speed axis (Fig. 7C), 75% of the recorded neurons (36 of 48) showed higher gains along the same-speed axis, indicating the presence of a nonlinearity that favored the same-speed gratings. The rest of the neurons were grouped around the line of slope 1, indicating that the nonlinearity did not favor same-speed gratings, although the gains were almost always <1.0, indicating the presence of a nonlinearity. The gains averaged 0.79 and 0.56 along the same-speed and different-speed axes. For same-speed dual gratings there was a slight difference in the change in the gain of response that depended on the degree that speed tuning was affected by spatial frequency. Using the broad classifications described above, neurons described as spatiotemporal-frequency-independent (Fig. 7C, black circles), unclassed (gray circles), or speed-tuned (open circles) had average gains of 0.86, 0.82, and 0.65, respectively. For different-speed dual gratings, gain averaged 0.53, 0.58, and 0.53 for the three classes of MT neurons.

The movement of the two overlapping gratings contained both a first-order component of motion, defined by luminance changes in time, and a second-order component, defined by contrast changes in time (Cavanagh and Mather, 1989). When the gratings moved at the same speed, both the first- and second-order motion components were in the same direction; when the two gratings moved at different speeds, the second-order motion component opposed the first-order component. Therefore, we conducted a control experiment to determine whether the reduced gain for the different-speed stimulus could be explained by the presence of the opposing second-order motion. We tested 16 MT neurons using dual-grating stimuli in which the two gratings were positioned side by side rather than overlapping. The side-by-side gratings were positioned within the receptive field of MT neurons and together spanned the area taken up by the stimulus in the experiments that used spatially superimposed dual gratings. We then performed the same exper-
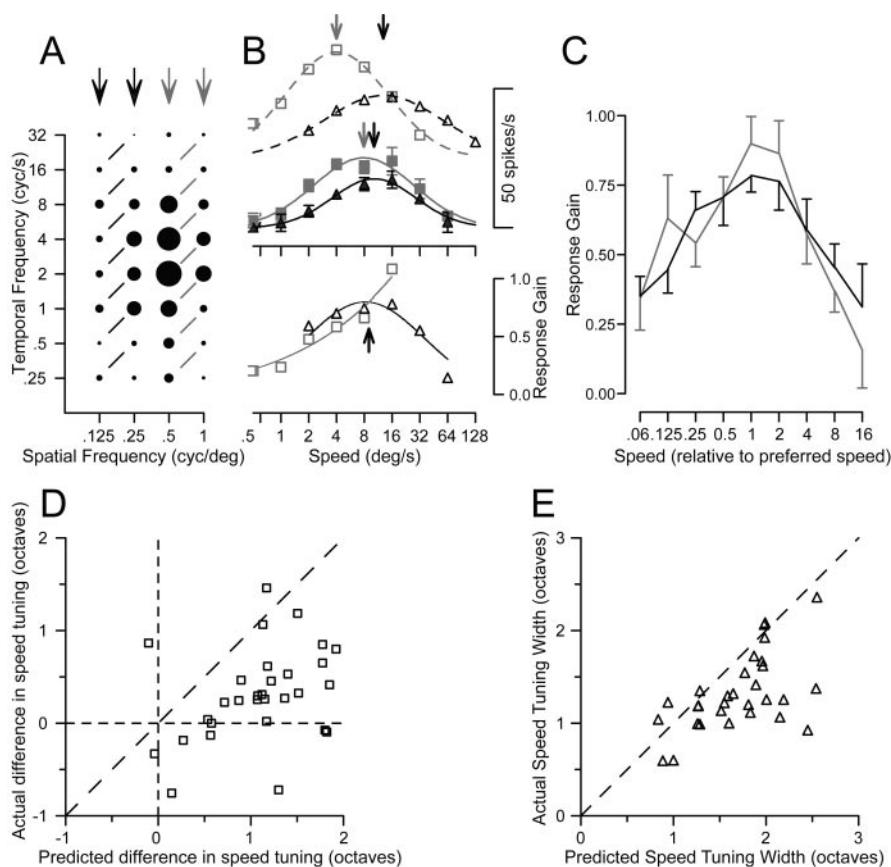

Figure 6. Demonstration of a nonlinear mechanism that creates better speed tuning for dual gratings than for single gratings. *A*, Response field of an example MT neuron to single gratings of varying spatial and temporal frequency, as in Figure 3. The black and gray arrows indicate the spatial frequencies used for the two groups of dual grating stimuli. Responses connected by black or gray lines indicate dual-grating stimuli in which each grating provided motion at the same speed. *B*, Top, The open symbols and dashed curves plot the speed tuning predicted by the linear sum of the responses to individual gratings, for the spatial frequency pairs of stimuli in *A* comprising the lower (black) or higher (gray) frequencies. The filled symbols and continuous curves plot the actual speed tuning for dual gratings. Arrows plotted along the curves indicate the preferred speeds for predicted responses and the actual responses. *B*, Bottom, black and gray symbols plot the normalized gain of the response to dual gratings for the lower and higher ranges of spatial frequency pairs. The black arrow shows the mean of the preferred speeds for the actual responses over the two ranges of spatial frequencies. Gain was not computed for a given dual-grating stimulus unless the predicted response, in the denominator, was at least 10% of the maximum response. *C*, The population average of normalized response gain is plotted as a function of speed, in which the data of each neuron has been normalized so that preferred speed was 1. Black and gray curves show responses for low and high ranges of spatial frequencies. Error bars indicate SEM. *D*, Comparison of the actual and predicted difference in preferred speed between the dual-grating stimuli comprised of different pairs of spatial frequencies. Each symbol shows data for a single MT neuron. *E*, Comparison of the actual and predicted speed tuning for the dual grating stimuli. Each symbol indicates the tuning width for a single MT neuron. As detailed in Results, predicted responses were obtained by summing the responses to each component grating presented singly.

iment and analysis detailed above and in Figure 7. When the gratings were spatially separate, eliminating second-order motion from the different-speed pair, MT neurons still showed a nonlinear response, in which the gain of the response to the same-speed gratings was, on average, 1.23 times greater than the response to the different-speed gratings. For comparison, the gain was 1.41 times greater when the gratings were superimposed.

Additional evidence for a nonlinearity that improves speed tuning of MT neurons

If the presence of multiple spatial frequencies in the stimulus acts through the kind of nonlinearity described in the previous section to improve speed tuning, then speed tuning should vary in a number of consistent ways depending on the exact form of the stimulus. The presence of multiple spatial frequencies should render speed tuning both less dependent on form and more nar-
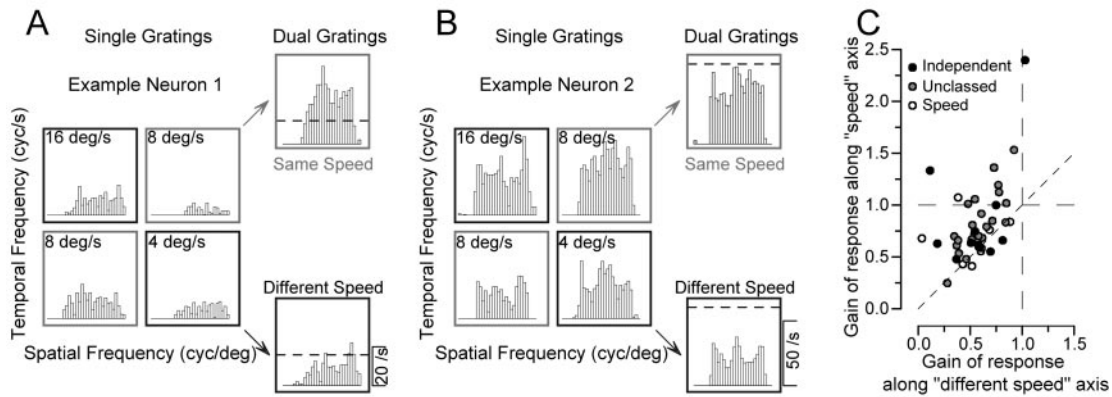
Figure 7. Additional analysis of the nonlinear response to dual grating stimuli. *A, B*, Poststimulus time histograms showing the responses of two example neurons to both single gratings and dual gratings. In each panel, the four responses grouped in a square were obtained from single gratings with all combinations of two spatial and temporal frequencies. The numbers in each box indicate the speed of the stimulus. The two histograms at the top and bottom right of each panel show the responses to dual gratings composed of gratings moving at the same speed (gray boxes) or different speeds (black boxes). The dashed line shows the response level predicted by adding the mean firing to the two component grates presented alone. *C*, Group data for all 48 neurons whose responses were measured to dual gratings. Each symbol summarizes data from one neuron and shows the average gain of the response to same-speed dual gratings as a function of that for different-speed dual gratings. Gain is defined as the actual response divided by the linear prediction of the response. The filling of each circle indicates the classification of the neuron according to the effect of spatial frequency on preferred speed for single grating: black, independent; gray, unclassed; open, speed-tuned.
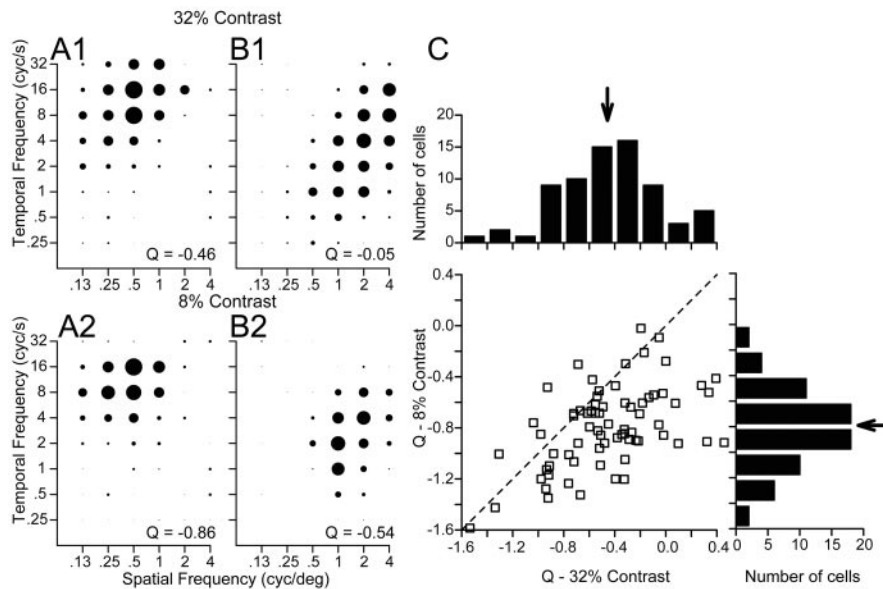
Figure 8. Effect of spatial frequency on preferred speed of MT neurons for high- versus low-contrast sine-wave gratings. *A1, B1*, Spatiotemporal frequency response fields of two example MT neurons for high-contrast gratings (32%), in the same format as Figure 3. *A2, B2*, Spatiotemporal frequency responses fields of the same two neurons for low-contrast gratings (8%). For each contrast, the diameter of the symbols was normalized relative to the peak response at that contrast. The numbers in the bottom right of each graph give the value of Q obtained by fitting Equations 2 and 3 to the data in that graph. *C*, Comparison of the values of Q for single sine-wave gratings of high versus low contrast. In the graph, each symbol summarizes the responses of an individual neuron and the dashed line indicates a line of slope 1 and *y*-intercept of 0. The histograms above and to the right of the scatter plot show the distributions of Q for single sine-wave gratings of 32% and 8%, respectively. Neurons were included in the histograms only if they were tested with gratings of both contrasts. The arrows in both histograms indicate the mean values across our sample of MT neurons.

wave gratings. In the present section, we test each of these predictions.

Effect of contrast on spatiotemporal frequency response fields

Lowering the contrast of the moving gratings moved the responses in MT substantially toward spatiotemporal independence and away from speed tuning, without altering the preferred spatial or temporal frequency of individual neurons. The effect of lowering contrast was measured in 61 MT neurons by using interleaved trials to present moving gratings composed of the high contrast used in the previous section (32%) (Fig. 8*A1,B1*) and a lower contrast (8%) (Fig. 8*A2,B2*). For the two neurons summarized in Figure 8, *A* and *B*, the spatiotemporal response field was visibly less oriented when the grating contrast was lower. Quantitative analysis showed that both neurons became less speed-tuned: reducing the contrast of the grating changed the value of Q from -0.46 to -0.86 in one neuron (Fig. 8*A1,B1*) and from -0.05 to -0.54 in the other (Fig. 8*A2,B2*). The population summary in Figure 8*C* shows that lowering the contrast of the stimulus caused the Q values of many MT neurons to shift toward spatiotemporal independence (i.e., toward -1), leading to a shift in the mean Q for this subset of our population of MT neurons from -0.45 to -0.79.

Responses to square-wave gratings

When the stimulus consisted of a square-wave grating instead of a sine-wave grating, the responses of MT neurons were essentially speed-tuned. Comparison of the spatiotemporal response fields for two example neurons illustrates responses that are much more nicely oriented for square-wave gratings (Fig. 9*A2,B2*) than for sine-wave gratings (Fig. 9*A1,B1*). For the two sample neurons,

rowly tuned than predicted by the summation of the responses to the component gratings singly. Neurons might be closer to speed-tuned for high-contrast than for low-contrast gratings if the former cause some saturation earlier in the visual pathways. Neurons should show better speed tuning for square-wave than sine-wave gratings, because the former comprise multiple spatial frequencies. The speed tuning for moving random-dot textures should be narrower than predicted from the responses to sine-

the values of Q obtained by applying Equations 2 and 3 were -0.88 and -0.58 for sine-wave gratings and 0.02 and -0.20 for square-wave gratings. The same effect appeared in almost all 20 MT neurons we tested with sine-wave and square-wave gratings. As summarized in Figure 9*C*, the values of Q for square-wave gratings grouped around 0 and were less than the values of Q for sine-wave gratings in all but two neurons.

The use of the spatiotemporal response fields for square-wave gratings is formally incorrect because square-wave gratings are composed of multiple sine-wave gratings including the fundamental spatial frequency (F) and higher harmonics ($3F$, $5F$, $7F$, etc.). The spatial and temporal frequency content of a moving square-wave grating is a series of points aligned along an isospeed contour in the Fourier space. Because the stimulus itself is oriented in plots of temporal versus spatial frequency, the responses to square-wave gratings also should appear to be more oriented, when plotted according to the fundamental frequency, as shown in Figure 9. Thus, the critical question is not whether MT neurons appear more speed-tuned for square-wave gratings, as they do, but rather whether the shift toward speed tuning can be accounted for using a linear summation of the responses to the fundamental and harmonic components of the stimulus, after adjusting for the contrast change attributable to the amplitude reduction. For each neuron tested, we predicted the value of Q for square-wave gratings responses using the parameters of Equations 2 and 3 fitted to the responses to sine-wave gratings at 32% contrast. On average, the linear prediction did not predict the full extent of the shift toward speed tuning for square-wave gratings: across the 20 neurons the average value of Q changed from -0.53 to -0.08 when the stimulus was changed from sine- to square-wave gratings, whereas the linear model predicts a more modest change to -0.30 (Fig. 9*D*).

The relationship between spatial and temporal frequency tuning speed tuning to random dots
For a subset of our sample of neurons (71 of 104), we were able to compare the speed tuning for a random-dot stimulus to the spatiotemporal frequency response field. The random-dot stimulus contains a broad spatial frequency spectrum; our strategy, again, was to compare the speed tuning of MT neurons for random-dot stimuli with predictions based on the responses of each neuron to single sine-wave gratings of different spatial and temporal frequencies. To predict the response to random dots, we averaged the actual responses to sine-wave gratings along contours of equal speed, as indicated by the dashed lines in Figure 10*A*. In the example of Figure 10*B*, the preferred speed of the predicted speed tuning curve (black symbols and curve) is similar to that for the actual response to the random dots (gray symbols and curve). However, the width of the speed tuning predicted by the sine-wave gratings is greater than that for the actual responses to
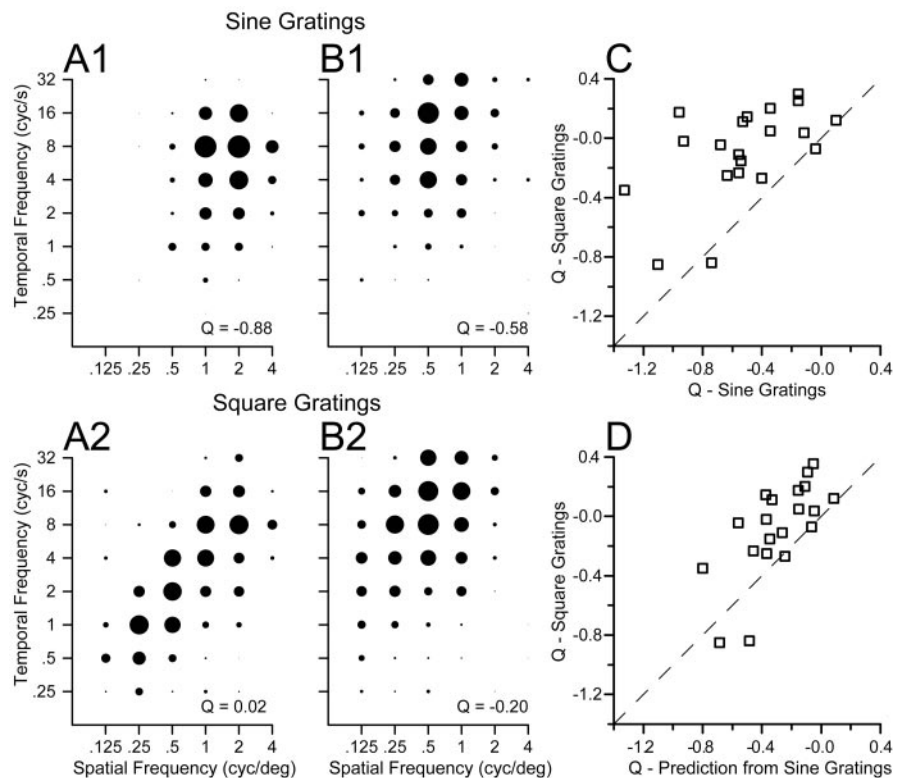

Figure 9. Comparison of the effect of spatial frequency on preferred speed for sine-wave versus square-wave gratings. *A1, B1*, Spatiotemporal frequency response fields of two example MT neurons for high-contrast sine-wave gratings (32%), in the same format as Figure 3. *A2, B2*, Spatiotemporal frequency responses fields of the same two neurons for square-wave gratings. For each class of stimulus, the diameter of the symbols was normalized relative to the peak response for that stimulus. The numbers in the bottom right of each graph give the value of Q obtained by fitting Equations 2 and 3 to the data in that graph. *C*, Comparison of the values of Q for sine-wave versus square-wave gratings. *D*, Comparison of the values of Q fitted to the responses to square-wave gratings with those predicted by a linear sum of the responses to sine-wave gratings representing the frequency components in the square wave: the fundamental frequency (F) plus $3F$, $5F$, $7F$, $9F$, and $11F$. In *C* and *D*, each symbol summarizes the responses of an individual MT neuron and the dashed line has a slope of 1 and an intercept of 0.

random-dot motion. These two features were representative of our sample population. The actual and predicted preferred speeds for moving random dots were highly correlated ($r = 0.86$) (Fig. 10*C*). However, in most cells the tuning was narrower for the dot stimuli than predicted by applying the linear model to the responses to sine-wave gratings (Fig. 10*D*). Across the population, σ^2 averaged 1.49 octaves for the dot stimuli and 1.98 octaves for the predictions from sine-wave gratings ($p < 0.01$; t test). The random-dot stimulus did contain higher spatial and temporal frequencies than were used in gratings to measure the spatial and temporal frequency tuning of MT neurons. Although most MT neurons did not respond to spatial frequencies higher than four cycles per degree, the discrepancy between the predicted and actual speed tuning width could be an artifact of responses to the higher spatial frequencies contained in the random-dot stimulus.

For each neuron in our sample population we also used the fits of Equations 2 and 3 to estimate the spatial and temporal frequency combination that would elicit the best response. For high-contrast gratings (32%), the peak spatial frequency varied from 0.125 to 4 cycles per degree (mean, 0.55 cycles per degree; SD, 1.1 octaves). The preferred temporal frequency of MT neurons ranged from 0.75 to 25 Hz (mean, 3.94; SD, 1.02 octaves). Reducing the contrast of the moving gratings from 32 to 8% caused the preferred spatial and temporal frequencies to decrease by an average of 0.05 log unit. Comparison of the preferred parameters for the two contrasts in each individual yielded correla-
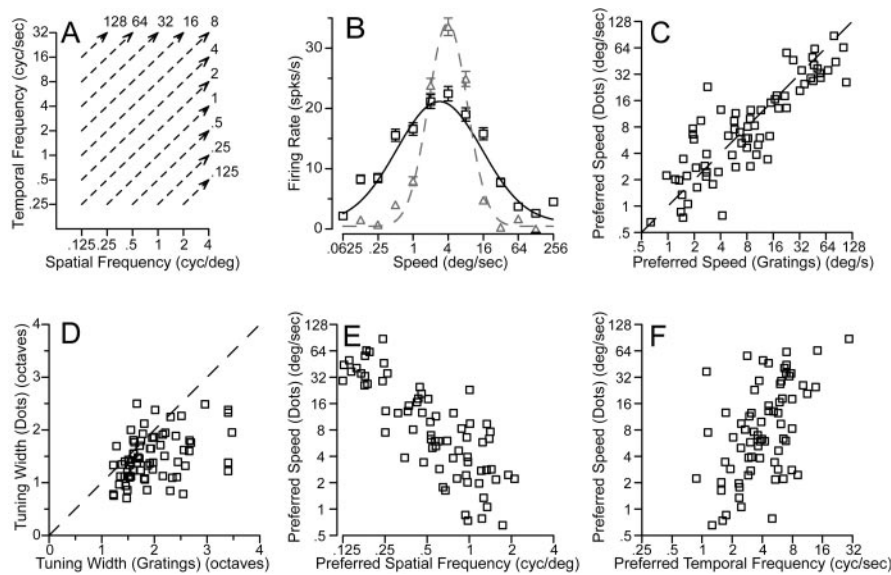
Figure 10. Comparison of MT neuron responses to sine-wave gratings versus random-dot textures. *A*, A diagram of the method used to predict speed tuning for textures from response to gratings. For each speed of texture motion, the response of the neuron was averaged for sine-wave gratings that fell along the isospeed contours indicated by the dashed lines. *B*, Black versus gray symbols and curves show predicted and actual responses and turning curves from an example MT neuron as a function of texture speed. In *C–F*, each symbol shows measurements made from an individual neuron. *C, D*, Comparison of predicted and actual preferred speed (*C*) and tuning width (*D*) for the population of neurons. The dashed line shows the expected relationship if the predicted and actual tuning matched. The values plotted in these graphs were taken from the Gaussian curves fitted to the data for each neuron. *E, F*, Comparison of the preferred speed to random-dot textures with the preferred spatial frequency (*E*) or temporal frequency (*F*) for sine-wave gratings. Preferred spatial and temporal frequencies were obtained from the parameters that gave the best fits of Equations 2 and 3 to the data.
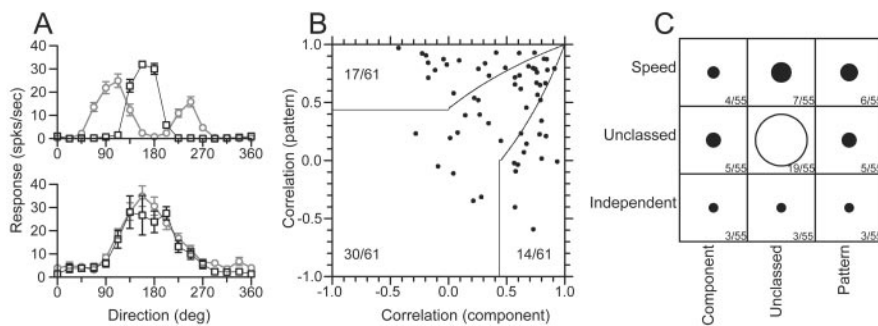
Figure 11. Comparison of the responses of MT neurons to moving plaids with the degree of form-invariant speed tuning. *A*, Direction tuning of two example MT neurons for moving plaids and single sine-wave gratings. *A*, Top, Component-selective neuron; bottom, pattern-selective neuron. Black squares show the response to single sine-wave gratings, whereas the gray circles indicate the response to the plaid stimulus. Error bars indicate the SEM response for each direction of plaid movement. The plaid components were separated by 135°. *B*, The distribution of pattern versus component selective neurons, based on the correlation assay defined by Movshon et al. (1986). Solid lines indicate the basis for classification. The numbers indicate how the sample was distributed in each classification area. *C*, The joint distribution of classification for pattern selectivity (*x*-axis) and form-invariant speed tuning (*y*-axis). The diameter of each symbol, as well as the accompanying numbers, indicate the number of cells belonging to each joint classification. For speed tuning, the classification was taken from the analysis of the values of *Q* (Figs. 3, 4).

tion coefficients of 0.82 and 0.92 for spatial and temporal frequency. There was no statistically significant effect of neuron categorization (spatiotemporally independent, unclassed, or speed tuned) on the preferred spatial or temporal frequency. Although it did not change the tuning for spatial or temporal frequency, reducing the contrast of the gratings from 32 to 8% did reduce the peak firing of MT neurons by an average of 42%.

Figure 10, *E* and *F*, compares the preferred spatial and temporal frequencies of MT neurons with their preferred speed to dot motion. The correlation between preferred spatial frequency and preferred speed was very high ($r_{sf} = -0.81$), whereas the relation-

ship was less strong between preferred temporal frequency and preferred speed ($r_{tf} = 0.53$). Correcting for the small correlation between preferred spatial and temporal frequency ($r = -0.25$) changed the estimated correlations between preferred spatiotemporal frequency tuning and preferred speed only slightly ($r_{sf} = -0.82$; $r_{tf} = 0.54$). Thus, it appears that spatial frequency tuning dominates temporal frequency tuning in determining the speed tuning of MT neurons. The dominance of spatial frequency tuning in determining preference for speed is consistent with the psychophysical finding that there are more channels for spatial frequency than for temporal frequency (Watson and Robson, 1981).

The response of MT neurons to moving plaids
Previous reports (Movshon et al., 1986) have used stimuli called plaids to subdivide the neurons according to their response to complex motions. Plaids are composed of two overlapping gratings at different orientations, each undergoing motion orthogonal to its orientation. To determine whether the responses to plaids was correlated with the degree of speed tuning, we tested 61 MT neurons with interleaved single sine-wave gratings and plaids whose grating components were separated by 135° of rotation. Sine-wave gratings and plaid components had the same spatial and temporal frequency, which was chosen to be close to the preferred values for each of these parameters. When tested with single gratings that moved in 16 different directions, each neuron showed traditional direction tuning curves with a single peak (Fig. 11*A*, open squares). When tested with plaids that moved in the same 16 directions (Fig. 11*A*, open circles), "component neurons" (top) showed direction tuning curves with two lobes: one lobe each for the directions of motion when the two component gratings were moving in the preferred direction of the neuron under study. Because the plaids were composed of gratings separated by 135° of rotation, the two lobes had peaks separated by 135°. "Pattern neurons" (Fig. 11*A*, bottom) respond to the direction of motion of the overall pattern; therefore, they have a direction tuning curve with only a single peak that is the same for plaids (filled circles) and single gratings (open squares). Most neurons showed direction tuning for plaids that was intermediate between the two extremes shown in Figure 11*A*.

We used the correlation analysis developed by Movshon et al. (1986) to classify MT neurons as pattern, component, or unclassed. This analysis is similar to that used in Figure 5 to quantify the degree of speed tuning of MT neurons. It uses the response to

single gratings to predict the expected response of a pattern and a component neuron and then performs partial correlation analysis to ask whether the actual response is better correlated with the prediction for the pattern or component models. The direction tuning for single sine-wave gratings was used as the prediction of the pattern model. The sum of this same direction tuning rotated 67.5° clockwise and counterclockwise was used as the prediction of the component model. Correlations were computed using the pairs of actual and expected responses for all directions of stimulus motion. The equations for partial correlation were:

$$R_p = \frac{(r_p - r_c * r_{cp})}{\sqrt{(1 - r_c^2)(1 - r_{cp}^2)}} \quad (6)$$

$$R_c = \frac{(r_c - r_p * r_{cp})}{\sqrt{(1 - r_p^2)(1 - r_{cp}^2)}} \quad (7)$$

where R_c and R_p are the partial correlations of the direction tuning curves for plaids with the component and pattern predictions, r_c is the correlation between the response to plaids and the model prediction for a component neuron, r_p is the correlation to the modeled pattern neuron response, and r_{cp} is the correlation of the two predictions. R_p was then plotted as a function of R_c (Fig. 11B) for each of the 61 neurons we studied with plaids, and the neurons were classified as component, unclassed, or pattern according to the criteria indicated by the solid curves. Consistent with previous reports, ~25% of our sample were classified as pattern neurons, ~25% as component neurons, and ~50% were unclassed (Movshon et al., 1986).

Figure 11C evaluates our sample of MT neurons to see whether there is any correlation between the classification of individual neurons along the separate axes of speed-tuned/spatiotemporally separable (y-axis) and pattern/component responses to plaids (x-axis). For each entry in the table, the diameter of the symbol indicates the number of neurons that fell in that class. The largest group was unclassed along both dimensions, but there was neither a visible nor a statistically significant correlation among neurons that were classed along one or both axes (Spearman rank correlation; $r = 0.067$; $p > 0.5$).

Discussion

Are MT neurons tuned for the speed of sine-wave gratings?
Current theories about the creation of motion-sensitive neurons revolve around two extremes that make different predictions about how motion is represented in the brain. At one extreme, motion-sensitive neurons are tuned independently for spatial and temporal frequency. Speed selectivity then becomes a consequence of the fact that they show their largest response for a specific pair of spatial and temporal frequencies, defining preferred speed as preferred temporal frequency divided by preferred spatial frequency. In this model, speed tuning would depend significantly on the spatial frequency of the stimulus. At the other extreme, neurons are tuned for speed and therefore show some covariance in their spatial and temporal frequency tuning, eliminating or minimizing the dependence of speed tuning on spatial frequency.

As we illustrated in Figure 1, one way to diagnose whether the response of a neuron shows spatiotemporal independence or speed tuning is to measure the responses to single sine-wave gratings and ask whether the response field is tilted in Fourier space. In a study similar to ours, Perrone and Thiele (2001), also demonstrated that most MT neurons show some degree of spatiotemporal tilt. They concluded that MT neurons are tuned for speed.

However, although tilt in the spatiotemporal response profile is a necessary condition for speed tuning, it is not sufficient. By performing the direct analysis of plotting response as a function of speed for each spatial frequency and by determining whether the amount of spatiotemporal tilt was consistent with speed tuning, we have now demonstrated that only 25% of MT neurons are tuned for the speed of sine-wave gratings in a way that is form-invariant. Indeed, the most striking aspect of the response to single gratings is the unimodal distribution of the value of Q, with most neurons falling between the two theoretical extremes. Although we have classified the neurons as independent, unclassed, and speed-tuned to ease analysis, the effect of spatial frequency on speed tuning is best described as a continuum.

How do MT neurons become tuned for the speed of real-world objects?
Sine-wave gratings provide a tool for the analysis of neural responses, but have the drawback that they do not occur frequently in natural visual scenes (Field, 1987; Dong and Atick, 1995). Thus, there is no reason to think that the visual system would be specialized for reporting accurately the speed of motion of sine-wave gratings. Indeed, several features of the responses of MT neurons suggest the existence of neural mechanisms that would solve this problem for real-world objects. First, dual-grating stimuli provided direct evidence for a nonlinear mechanism that would move MT neurons toward form-invariant speed tuning for real-world objects. Second, MT neurons show responses that are much closer to speed-tuned for square-wave gratings than for high-contrast sine-wave gratings. Square-wave gratings contain sharp edges often found in natural scenes. They can be described as the sum of multiple sine-wave gratings of different spatial frequencies, and would gain access to the speed-tuning nonlinearity. Third, MT neurons show responses that are closer to speed-tuned for high-contrast sine-wave gratings than for low-contrast gratings. High-contrast sine-wave gratings might create saturated responses at earlier stages of neural processing. As the neural response becomes distorted, it comprises multiple spatial frequencies and would trigger the effects of the speed-tuning nonlinearity.

We think that the speed-tuning nonlinearity is more than linear summation followed by the half-wave rectification created by a threshold, because the responses to a given set of stimuli are both more reliably and more narrowly speed-tuned than predicted by a linear summation based on the response of the neuron to single sine-wave gratings. Dual-grating stimuli not only altered the amplitude of the response relative to that predicted by linear summation, but it also caused the speed tuning to shift toward the single value that was revealed by testing with random-dot textures. These phenomena would result from the kind of excitatory interactions proposed in models by Simoncelli and Heeger (1998), in which MT neurons respond in a speed-tuned manner as a result of excitatory interactions between inputs from neurons that prefer the same speed but vary in their spatial and temporal frequency tuning (Heeger et al., 1996; Simoncelli and Heeger, 1998).

MT neurons seem to derive form-invariant speed tuning in a way that takes advantage of the fact that moving objects in natural scenes comprise multiple spatial frequencies. The nonlinearity revealed by our dual-grating experiments would make the response field appear to be oblique (Fig. 1E) when multiple spatial frequencies are present, even if it was cardinal for single sine-wave gratings (Fig. 1B). It does this by allowing strong responses when stimuli fall along the preferred speed line in spatiotemporal

frequency plots, while suppressing responses along speed lines above or below the preferred speed. It enables the desired result (a form-invariant assessment of speed) without creating receptive fields that are tuned for the speed of artificial stimuli such as sine-wave gratings.

Implications for models of speed tuning in MT

Our data do not support strongly any particular model of motion-selective neurons. Rather, they raise a number of important, related issues about future attempts to model motion-selective neurons, especially those in area MT. First, new models of MT neurons should have the goal of creating a population of MT responses that reflect the diversity of speed tuning rather than the two extremes. In contrast, previous models of motion-sensitive neurons have focused on creating responses at the two extremes. An important issue that will need to be addressed is whether direction-selective neurons in the primary visual cortex show the same diversity, or if they lie closer to spatiotemporal independence as suggested by Tolhurst and Movshon (1975). Our data, combined with the similarity of the population obtained for the same experiments in V2 (Levitt et al., 1994), suggest that the diversity we found in MT may be a general property of motion-sensitive neurons, rather than something specific about MT. Second, new models should include the nonlinear gain interaction described by our data so that the speed tuning for objects can be more form-invariant and narrower than would be predicted from the responses to single sine-wave gratings. Third, a better model for MT neurons should stress the correlation we found between preferred speed selectivity and the spatial frequency tuning of the neuron (Fig. 10*E*). Finally, our failure to find a relationship between where neurons fall on the axis of speed tuning versus spatiotemporal frequency independence and where they fall on the axis of pattern versus component responses to plaids implies that these two features of MT neuron responses can be modeled independently.

Relationship between the coding of speed in MT and motion psychophysics

We turn now to the problem of decoding the population response in area MT to reconstruct speed for use in generating perceptions and actions. Although neurons in area MT may not encode the speed of motion independently of spatial frequency, they may still contribute to our sensation of motion. In fact, psychophysical data indicate that the spatial frequency of a grating does affect the perception of speed: low spatial frequencies bias human observers to perceive faster speeds (Campbell and Maffei, 1981; Reisbeck and Gegenfurtner, 1999; Smith and Edgar, 1990). We have confirmed that spatial frequency biases the human perception of speed and have shown that spatial frequency also affects the initiation of smooth pursuit eye movements in monkeys: both of these effects are of a direction and magnitude predicted by the responses of the full population of neurons we have recorded in MT (our unpublished observations). Thus, it seems that neurons with speed tuning affected by spatial frequency are providing outputs from MT. These neurons that do not encode speed independently of spatial form cannot not be dismissed as interneurons that only perform computations within MT, but they also contribute to our perception of motion.

Finally, one might ask why it is important to have a speed-tuning nonlinearity when it is possible to obtain a reasonable estimate of speed by simply adding the responses to the component gratings of the visual stimulus without using the nonlinearity. We imagine two reasons. First, form-dependent speed tuning could cause serious misjudgments of object speed, especially if a small object is moving toward the observer, becoming larger and changing spatial frequency content as it looms. Second, as indicated by our data, one important function of the speed-tuning nonlinearity is to narrow the speed tuning of MT neurons. Narrower tuning of individual neurons means that a smaller population of neurons is activated by any given stimulus. Thus, the nonlinearity contributes to the creation of a sparse code, thereby increasing the efficiency of neural coding in a way that seems likely to benefit the organism (Olshausen and Field, 1996).

References

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

Campbell FW, Maffei L (1981) The influence of spatial frequency and contrast on the perception of moving patterns. Vision Res 21:713–721.

Cavanagh P, Mather G (1989) Motion: the long and short of it. Spat Vis 4:103–129.

Dong D, Atick J (1995) Statistics of natural time-varying images. Network Comput Neural Syst 6:345–358.

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A 4:2379–2394.

Friend SM, Baker Jr CL (1993) Spatio-temporal frequency separability in area 18 neurons of the cat. Vision Res 33:1765–1771.

Heeger DJ, Simoncelli EP, Movshon JA (1996) Computational models of cortical visual processing. Proc Natl Acad Sci USA 93:623–627.

Holub RA, Morton-Gibson M (1981) Response of visual cortical neurons of the cat to moving sinusoidal gratings: response-contrast functions and spatiotemporal interactions. J Neurophysiol 46:1244–1259.

Levitt JB, Kiper DC, Movshon JA (1994) Receptive fields and functional architecture of macaque V2. J Neurophysiol 71:2517–2542.

Lisberger SG, Movshon JA (1999) Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. J Neurosci 19:2224–2246.

Merrill EG, Ainsworth A (1972) Glass-coated platinum-plated tungsten microelectrode. Med Biol Eng 10:495–504.

Movshon JA (1975) The velocity tuning of single units in cat striate cortex. J Physiol (Lond) 249:445–468.

Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1986) The analysis of moving visual patterns. Exp Brain Res [Suppl 2]:117–151.

Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.

Perrone JA, Thiele A (2001) Speed skills: measuring the visual speed analyzing properties of primate MT neurons. Nat Neurosci 4:526–532.

Priebe NJ, Churchland MM, Lisberger SG (2002) Constraints on the source of short-term motion adaptation in macaque area MT. I. The role of input and intrinsic mechanisms. J Neurophysiol 88:354–369.

Reisbeck T, Gegenfurtner K (1999) Velocity tuned mechanisms in human motion processing. Vision Res 39:3267–3285.

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

Smith AT, Edgar GK (1990) The influence of spatial frequency on perceived temporal frequency and perceived speed. Vision Res 30:1467–1474.

Tolhurst D, Movshon J (1975) Spatial and temporal contrast sensitivity of striate cortical neurones. Nature 257:674–675.

Watson AB, Ahumada Jr AJ (1985) Model of human visual-motion sensing. J Opt Soc Am A 2:322–341.

Watson AB, Robson JG (1981) Discrimination at threshold: labeled detectors in human vision. Vision Res 21:1115–1122.