A Model of Neuronal Responses in Visual Area MT

Eero P. Simoncelli [a] and David J. Heeger [b]

[a] *Center for Neural Science*
4 Washington Place, Room 809
New York, NY 10003

[b] *Department of Psychology*
Stanford University
Stanford, CA 94305

Received 24 December 1996; revised 26 May 1997

Abstract

Electrophysiological studies indicate that neurons in the Middle Temporal (MT) area of the primate brain are selective for the velocity of visual stimuli. This paper describes a computational model of MT physiology, in which local image velocities are represented via the distribution of MT neuronal responses. The computation is performed in two stages, corresponding to neurons in cortical areas V1 and MT. Each stage computes a weighted linear sum of inputs, followed by rectification and divisive normalization. V1 receptive field weights are designed for orientation and direction selectivity. MT receptive field weights are designed for velocity (both speed and direction) selectivity. The paper includes computational simulations accounting for a wide range of physiological data, and describes experiments that could be used to further test and refine the model.

Keywords: MT, model, velocity, opponency, normalization.

• Much of this research was performed while EPS was in the Computer and Information Science Department at the University of Pennsylvania, and was partially supported by an NSF Science and Technology Center Grant (STC SBR-89-20230). EPS is currently supported by the Sloan Foundation through the NYU Theoretical Neurobiology Program, and by NSF CAREER grant 9624855.
• DJH is supported by a grant from the National Institute of Health (MH50228), and an Alfred P. Sloan Research Fellowship.
• We thank J. A. Movshon for helpful discussions, and the anonymous reviewers for their suggestions.
• Please send all correspondence to the first author at the address given above [Fax: +1 212 995-4011; E-mail: eero@cns.nyu.edu].

Published in Vision Research, 38(5): 743–761, 1998

1 Introduction

Visual motion perception has been the subject of extensive research in the fields of perceptual psychology, visual neurophysiology, and computational vision. It is widely believed that the brain contains mechanisms specifically devoted to the processing of motion (Maunsell & Newsome, 1987; Albright, 1993), and that this processing occurs in a "motion pathway" consisting of at least two stages. The primary visual cortex (area V1) constitutes the first stage. Information passes from there to the middle temporal (MT or V5) visual area (Dubner & Zeki, 1971).

Many neurons in area MT are tuned for retinal image velocity; they respond vigorously to a visual stimulus moving with a particular speed and direction, and are somewhat indifferent to the stimulus' spatial pattern (Movshon *et al.*, 1986). An empirical link has been established between neural activity in area MT and the perception of motion (Siegel & Andersen, 1986; Newsome & Pare, 1988; Logothetis & Schall, 1989; Newsome *et al.*, 1989; Salzman *et al.*, 1990; Salzman *et al.*, 1992; Britten *et al.*, 1992). Area MT is also known to be involved in pursuit eye movements (Newsome *et al.*, 1985; Dursteler *et al.*, 1986; Movshon *et al.*, 1990).

This paper describes a computational model for neural extraction and representation of visual motion. The model is derived from fundamental properties of motion information encoded in visual signals, and consists of two stages, corresponding to cortical areas V1 and MT. The input to the model is a time-varying visual stimulus, and the output corresponds to the steady-state firing rates of a population of neurons. These values form a distributed representation (population encoding) of image velocity for each local spatial region of the visual stimulus.

The purpose of this paper is to define a precisely parameterized version of this model, to investigate its ability to account for the known physiology of MT neurons, and to propose physiological experiments that might be used to falsify or further extend it. Some of the material in this article has been reported previously (Simoncelli, 1993; Simoncelli and Heeger, 1994; Heeger, Simoncelli and Movshon, 1995). A software implementation of the model is available on the Internet at `http://www.cns.nyu.edu/~eero/MT-model.html`.

2 The Model

The model consists of two primary stages corresponding to cortical areas V1 and MT. The basic form of computation is identical in each of these two stages: A weighted sum of input values followed by rectification, squaring and response normalization. In the following sections, we describe these computations in detail, and link them to their intended physiological correlates.

2.1 V1 Simple Cells

V1 neurons have been classified into two primary categories: simple cells and complex cells (Hubel & Wiesel, 1962). There is a long tradition in which simple cell responses have been characterized using linear receptive fields (Hubel & Wiesel, 1962; Campbell *et al.*, 1968; Campbell *et al.*, 1969; Movshon *et al.*, 1978b). In such a model, the neuronal response is a weighted sum (over local space and recently past time) of the local stimulus contrast. Linear models are commonly used to explain simple cell selectivity for stimulus orientation and spatial frequency. More recently, researchers have also used linear models to explain direction selectivity by incorporating suitable timing differences (delays) in the responses evoked from different parts of the receptive field (Fahle & Poggio, 1981; Watson & Ahumada, 1983; Watson & Ahumada, 1985; Adelson & Bergen, 1985; van Santen & Sperling, 1985; Burr *et al.*, 1986; McLean & Palmer, 1989; DeAngelis *et al.*, 1993; McLean *et al.*, 1994).

Typical linear receptive fields can have negative responses, since they combine inputs using both positive and negative weights, but extracellular response measurements (firing rates) are by definition positive. This deficiency is usually addressed by imposing a form of rectification on the linear output. Our model uses a *half-squaring* operation (halfwave-rectification followed by squaring) for these purposes (Heeger, 1992b). Half-squaring does not drastically alter the tuning properties of the model neuron, which are primarily determined by the underlying linear receptive field. This particular form of rectification is chosen for mathematical convenience, as described in the next section and in the *Appendix*.

Rectified linear receptive field models do not account for several important simple cell nonlinearities, such as response saturation and cross-orientation inhibition. Many of these behaviors can be accounted for by incorporating *response normalization* (Robson, 1988; Bonds, 1989; Albrecht and Geisler, 1991; Heeger 1991, 1992a,b, 1993; DeAngelis *et al.*, 1992; Carandini and Heeger, 1994; Carandini *et al.*, 1997; Tolhurst and Heeger, 1997a,b; Nestares and Heeger, 1997). In our model, this is achieved by dividing the half-squared linear response of each neuron by a quantity proportional to the summed activity of a pool of neurons within a cortical "neighborhood". The neighborhood includes cells tuned for the full range of orientation, direction, and spatio-temporal frequency. The result is that the response of each neuron is normalized with respect to stimulus contrast, thereby limiting the dynamic range. Nevertheless, the normalization operation does not alter the relative responses of neurons in the pool, since they are each normalized by the same factor. In addition, the tuning characteristics of most of the neurons are unaffected by the normalization, since the normalization factor is constant over the full range of orientation, direction and spatio-temporal frequency. The complete simple cell model (at a single spatial location) is illustrated in Fig. 1.

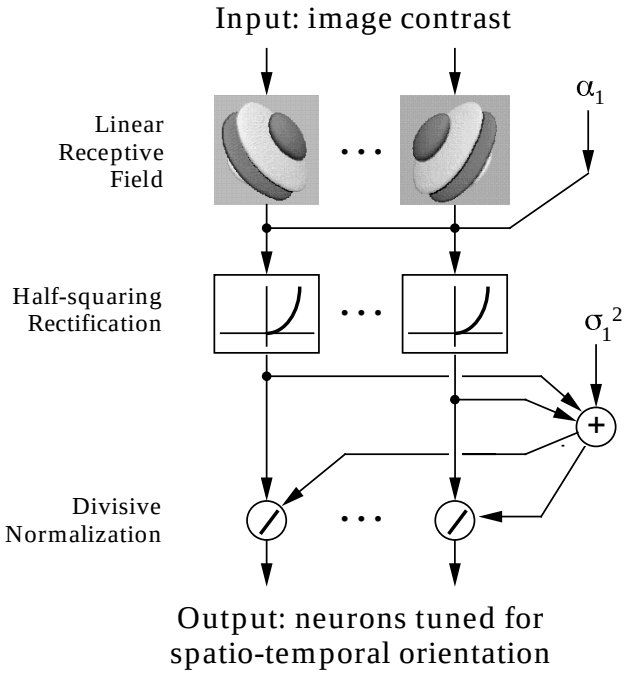

FIGURE 1: Model of V1 simple cells. Each neuron computes a weighted sum of its inputs followed by halfwave rectification, squaring, and response normalization. The underlying linear receptive fields (depicted as monochrome images) are localized in space and time, and are tuned for spatio-temporal orientation. An additive constant, α_1, is included in the summation, allowing for a spontaneous firing rate. The divisive normalization factor is computed as a sum of half-squared responses and a squared semi-saturation constant, σ_1.

2.2 V1 Complex Cells

Complex cells are similar to simple cells, in that they are selective for spatio-temporal orientation. But their responses are relatively independent of the precise stimulus position within the receptive field. It is widely believed that complex cells combine the responses of a set of underlying linear receptive field subunits (e.g., Movshon *et al.*, 1978a). In particular, the "motion energy" model accounts for a number of properties of complex cell responses (Pollen & Ronner, 1983; Adelson & Bergen, 1985; Emerson *et al.*, 1992; Heeger, 1992a). In its simplest form, a motion energy neuron sums the responses of four half-squared, linear receptive field subunits with phases in steps of $90°$, but with otherwise identical tuning properties. In addition to combination over phase, Emerson *et al.* (1992) found that spatial pooling (i.e., combination of subunits distributed over a localized spatial region) is needed to explain some aspects of complex cell responses. In our model, complex cell responses are computed as a weighted sum of simple cell afferents distributed over a local spatial region, but each having the same space-time orientation and phase. There is evidence that complex cells do not receive input from simple cell afferents, but directly from the LGN (see Heeger, 1992b for review). Likewise, our model

complex cell responses could be computed directly, although this essentially duplicates the processing that is performed by the simple cells.

2.3 MT Pattern Cells

The V1 neurons described thus far are not selective for stimulus velocity. Rather, they are selective only for the component of velocity orthogonal to their preferred spatial orientation (Adelson & Movshon, 1982). A velocity-selective neuron may be constructed by combining the outputs of a set of direction-selective V1 complex cells[1] whose preferred space-time orientations are consistent with the desired velocity, as illustrated in Fig. 2. This mechanism for velocity selectivity is a neural implementation of the "intersection-of-constraints" construction (Fennema & Thompson, 1979; Adelson & Movshon, 1982), and is conceptually the same as that proposed by Albright (1984). We emphasize, however, that the velocity selectivity in our model is not an explicit intersection-of-constraints calculation. Although each of our velocity-selective (MT) neurons has a preferred velocity (speed and direction), it will also respond (to a somewhat lesser degree) to non-optimal stimuli.

The details of this mechanism for velocity selectivity may also be described in the spatio-temporal frequency domain. The power spectrum of a translating two-dimensional pattern lies on a plane (Watson and Ahumada, 1983, 1985), and the tilt and spatial orientation of this plane depend only on the translational velocity. A model complex cell is selective for a localized band of spatio-temporal frequencies. A velocity-selective (MT) neuron is constructed by summing the responses of a particular set of V1 neurons, whose bands are bisected by a plane. The summation is over both orientation and spatial frequency. This construction is illustrated in Fig. 3, and is the same as that described (spatio-temporally) in Fig. 2. Because of the summation over spatial frequency, the resulting MT neurons have broader spatial frequency bandwidths than the V1 neurons, consistent with physiological data (Newsome *et al.*, 1983). We also subtract the responses of V1 neurons with bands that are far from the the plane (see *Appendix*).

In addition to the summation over spatio-temporal frequency illustrated in Figs. 2 and 3, each MT neuron sums the responses of V1 neurons with receptive field positions in a local spatial neighborhood. This is consistent with physiological evidence that MT receptive field diameters are roughly ten times those reported at corresponding eccentricities in V1 (Maunsell & van Essen, 1987; Albright & Desimone, 1987). The exact range of spatial pooling in the model is not critical for this paper, because our experimental simulations (see *Results*) are based on spatially homogeneous stimuli (moving and flickering gratings, plaids, and random

[1] Not all V1 neurons are direction-selective, but there is physiological evidence that those V1 neurons that project to MT have this property (Movshon & Newsome, 1996).

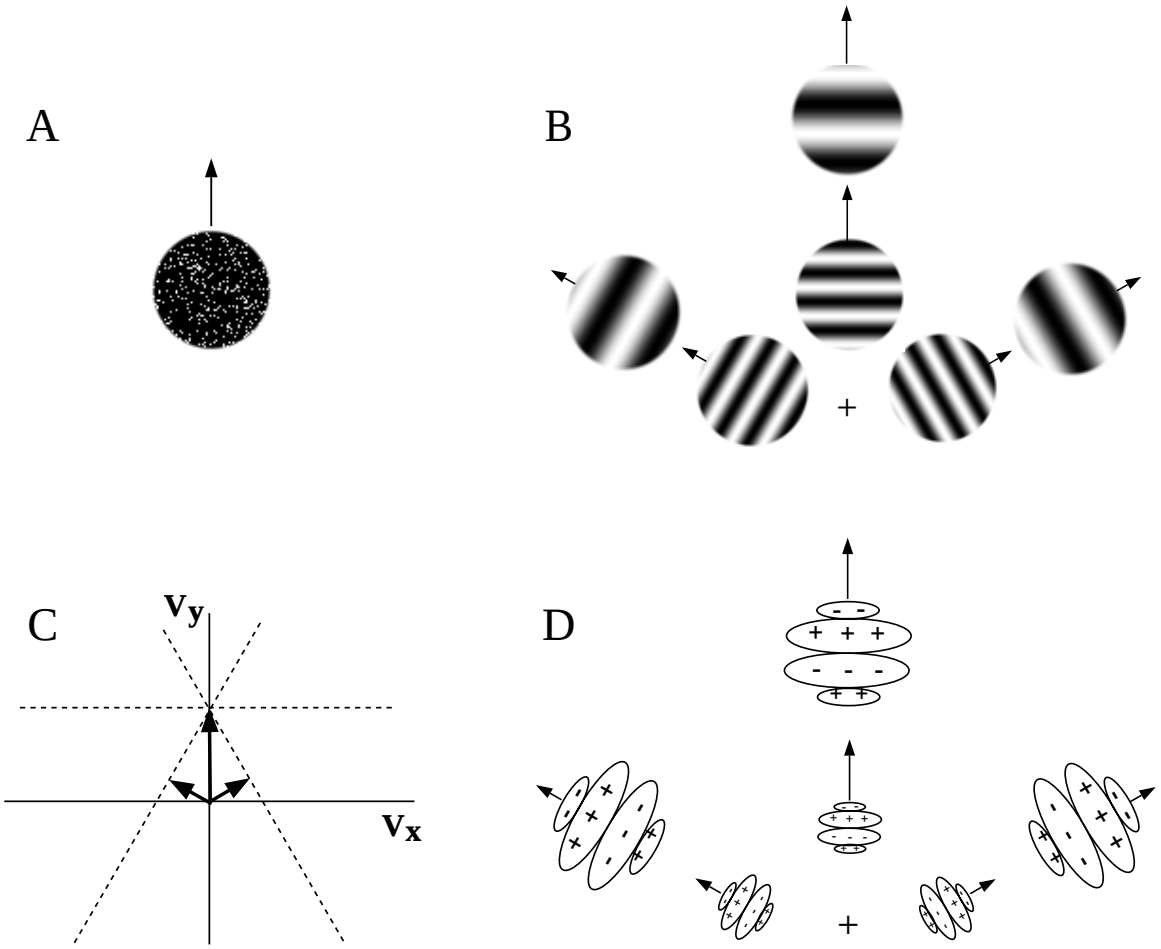

FIGURE 2: Construction of MT pattern cell velocity selectivity via combination of V1 complex cell afferents. **A:** Random dot field stimulus, drifting upward. **B:** Fourier decomposition of the dot stimulus. The stimulus is written as a sum of drifting sinusoidal components with appropriate normal velocities. A small subset of these are shown. **C:** Intersection of constraints (IOC) construction. The motion of a grating is ambiguous, since the component of velocity along the grating stripes produces no change in image intensity. Each arrow corresponds to the normal component of velocity for two of the gratings shown in **B**, and the dashed lines indicates the set of velocities consistent with the motions of those gratings. The intersection point of these constraint lines is the only velocity consistent with the motion of all of the components, and corresponds to the velocity of the dot stimulus. **D:** Set of V1 complex receptive fields selective for each of the components shown in **B**. The summed responses of such V1 neurons yields a pattern MT response that is selective for this stimulus velocity.

dot patterns).

The complete MT pattern cell construction is illustrated in Fig. 4. In addition to the weighted sum of complex cell afferents, a small additive constant, denoted α_2 is included to provide a spontaneous firing rate. Finally, the MT responses are half-

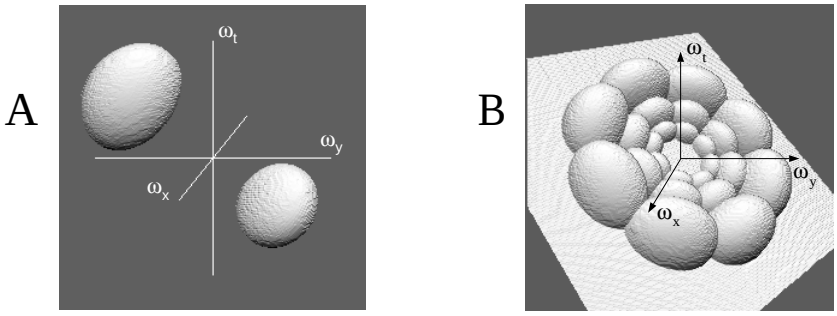

FIGURE 3: Construction of MT pattern cell velocity selectivity via combination of V1 complex cell afferents, shown in the Fourier domain. **A**: Selectivity of a V1 neuron corresponds to a pair of localized spatio-temporal frequency bands, symmetrically arranged about the origin. **B**: Selectivities of V1 neurons tuned for four orientations and three spatial scales, each consistent with a common velocity. The illustrated plane corresponds to the power spectrum of a stimulus moving at this common velocity. Responses of these V1 neurons are summed using positive (excitatory) weights to yield an MT response selective for this velocity. Not shown are a set of V1 neurons whose tuning bands lie off of the plane: these are combined using negative (inhibitory) weights. Also not indicated is the fact that the summation is performed over V1 neurons with receptive fields distributed over a local spatial region.

squared and normalized, as in the V1 stage of the model. The additional squaring nonlinearity leads to an MT contrast-response function that is steeper than that of the V1 cells, consistent with data reported by Sclar *et al.* (1990).

2.4 Model Implementation

This section provides the equations used to simulate neuronal responses. Additional mathematical details are given in the *Appendix*. A visual stimulus projected on the retina can be described by its light intensity distribution, $I(x, y, t)$, a function of two spatial dimensions (x,y) and time t. This representation ignores the color of the stimulus and assumes monocular viewing, but is in all other respects complete. The stimulus can also be characterized by its local contrast,

$$A(x, y, t) = [I(x, y, t) - \bar{I}]/\bar{I}, \tag{1}$$

where $\bar{I}$ is the average stimulus intensity (over space and time). This characterization is particularly relevant because (to first approximation) the retina produces a "neural image" of local contrast (Shapley & Enroth-Cugell, 1984).

The underlying linear response $L_n(t)$ of the nth simple cell is a weighted sum of the local contrast and a constant value, α_1:

$$L_n(t) = \int\int\int s_n(x, y, T)\, A(x, y, t - T)\, dx\, dy\, dT + \alpha_1. \tag{2}$$

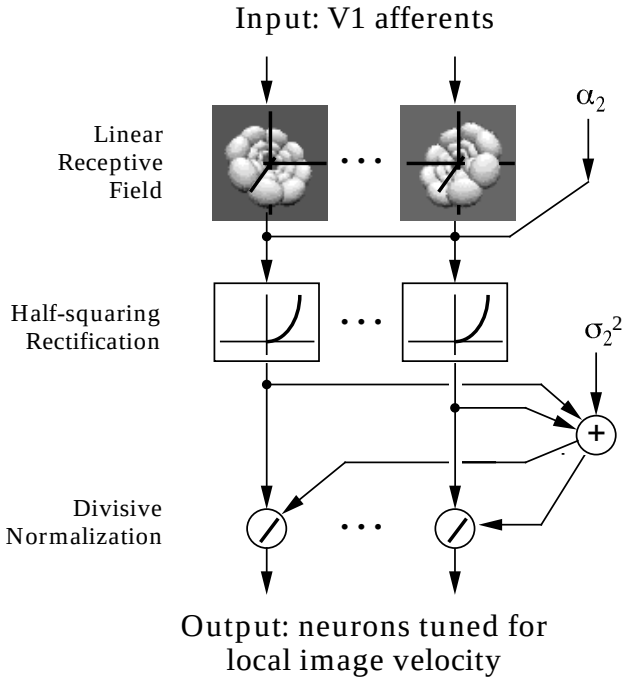

FIGURE 4: Model of MT pattern cells. Each neuron computes a weighted sum of V1 complex cell afferents, followed by half-squaring, and normalization. The V1 afferents are weighted to give a velocity-selective response, as described in Figs. 2 and 3. A constant, α_2, is added to provide a spontaneous firing rate. The divisive normalization factor is computed as a sum of half-squared responses and a squared semi-saturation constant, σ_2.

The weighting functions, $s_n(x, y, t)$, are a set of directional third derivatives of a Gaussian, with 28 different space-time orientations, replicated at all spatial locations. These functions contain both positive and negative values corresponding to the excitatory and inhibitory subregions of the receptive field, and are similar to linear operators used in previous receptive field models (Gabor functions, for example). The motivations for this particular choice of receptive field function are primarily computational, and are detailed in the *Appendix*.

The response of the nth simple cell, $S_n(t)$, is expressed as:

$$S_n(t) = \frac{K_1 \lfloor L_n(t) \rfloor^2}{\sum_m \lfloor L_m(t) \rfloor^2 + \sigma_1^2}, \tag{3}$$

where σ_1 is the semi-saturation constant of the normalization, K_1 determines the maximum attainable response, and $\lfloor \cdot \rfloor$ denotes the half-squaring operation:

$$\lfloor L(t) \rfloor^2 \equiv \max[0, L(t)]^2. \tag{4}$$

We interpret the response, $S_n(t)$, as the model equivalent of a post-stimulus time histogram (PSTH), a measure of the neuron's firing rate. Note that each neuron

suppresses itself (i.e., the summation in the denominator includes the numerator term). Assuming σ_1 is nonzero, the normalized response will always be a value between 0 and K_1, saturating for high contrasts. In our simulations, the responses $S_n(t)$ are computed directly, using Eqn. (3). Physiologically, these responses could be computed via inhibitory feedback mechanisms (for example, see Heeger, 1992b, 1993; Carandini and Heeger, 1994; Carandini *et al.*, 1997).

V1 complex cell responses are computed as local averages of simple cell responses:

$$C_n(t) = \sum_m c_{nm}\, S_m(t),$$

where the simple cell subunits are distributed over a local spatial region, but have the same space-time orientation and phase. The weights c_{nm} are all positive.

The underlying linear velocity-selective response $Q_n(t)$ of an MT pattern cell is expressed as:

$$Q_n(t) = \sum_m p_{nm} C_m(t) + \alpha_2, \tag{5}$$

where the p_{nm} are a set of weights (as illustrated in Figs. 2 and 3, and described in the *Appendix*), and α_2 determines the maintained (spontaneous) response level. Note that the summation includes pooling over spatial position, and that the weights p_{nm} assume both positive and negative values.[2] The index, n, parameterizes both the neuron's spatial position and velocity selectivity.

Finally, the nth pattern MT neuron's response, $P_n(t)$, is expressed as:

$$P_n(t) = \frac{K_2 \lfloor Q_n(t) \rfloor^2}{\sum_m \lfloor Q_m(t) \rfloor^2 + \sigma_2^2}, \tag{6}$$

where σ_2 and K_2 are constants that determine, respectively, the semi-saturation level and the maximum attainable response of the MT neuron. Again, we interpret the response, $P_n(t)$, as the model equivalent of a PSTH.

2.5 Model Parameters and Simulation

We have thus far described the basic computational structure of the model. The remaining details of the computation are determined by the following adjustable

[2] The spatial pooling in this summation is computationally redundant with that of the V1 Complex cells. As such, our implementation performs both summations simultaneously.

parameters:

(i) The spatio-temporal frequency coverage of the full set of V1 simple cells. The *Appendix* explains this in detail.
(ii) The velocity coverage of the full set of MT pattern cells. Again, details are provided in the *Appendix*.
(iii) The spatial pooling regions used in the V1 normalization, in the V1 complex cell summation, in the linear summation of the MT pattern cells, and in the MT normalization. For our simulations, we pooled over the entire stimulus region, since the stimuli are all spatially homogeneous.
(iv) The constants, σ_1 and σ_2, that determine the semi-saturation contrast levels.
(v) The constants, α_1 and α_2, that determine the spontaneous firing rates.
(vi) The scale factors, K_1 and K_2, that determine the maximum attainable firing rates. These were set to give maximal firing rates of one.

The scalar parameters were hand-adjusted to qualitatively match the physiological data:

$$\sigma_1 = 0.2, \quad \alpha_1 = 0.07, \quad K_1 = 4,$$
$$\sigma_2 = 1, \quad \alpha_2 = 0.8, \quad K_2 = 1.8.$$

This set of values was retained for all of the simulation results, except where specifically noted.

V1 responses to some stimuli (transparent dots, square gratings, stochastic dot patterns, etc.) were computed by creating a movie of the stimulus and convolving with spatio-temporal filters. Time-averaged V1 responses to grating and plaid stimuli, on the other hand, were computed analytically in the Fourier domain, using expressions for the frequency responses of the V1 receptive fields. Responses to drifting dot fields are also computed analytically in the Fourier domain. These simulations were checked for consistency by performing simulations on movies of the same stimuli. Simulations are quite efficient due to the choice of linear operators (see *Appendix*): simulation times are typically on the order of seconds or minutes, even for full population responses such as those shown in Figs. 5–8.

2.6 Distribution of Responses

A single MT neuron cannot encode stimulus velocity. First, in addition to velocity, a neuron's response depends on stimulus contrast and spatial pattern. Second, even for a fixed contrast and spatial pattern, there are families of velocities (arranged in concentric contours around the preferred velocity) that evoke the same response. Thus, the representation of velocity in our model is implicitly encoded in the simultaneous responses of a population of neurons. The responses of this population may be interpreted as discrete samples of a continuous two-dimensional

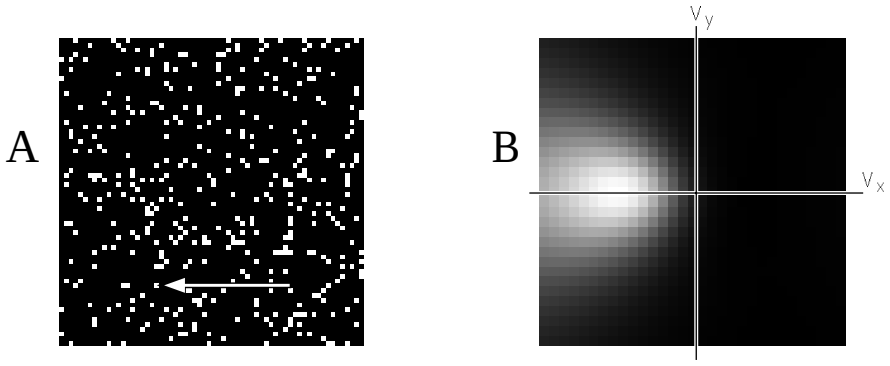

FIGURE 5: A: Drifting random dot field stimulus. Arrow indicates direction of motion. **B:** Distribution of model MT neuronal responses. Each point in the monochrome image corresponds to an MT neuron with a different preferred velocity. Intensity is proportional to the neuron's response. For example, the intensity at the center of the figure corresponds to the response of an MT neuron tuned for zero motion and the intensity at the top-right corner of the figure corresponds to the response of an MT neuron tuned for motion upward and to the right.

response distribution.

Figure 5 shows such a response distribution for a drifting dot field stimulus. Each point in Fig. 5B corresponds to an MT neuron with a different preferred velocity; intensity in the figure is proportional to the neuron's response. The location of the response distribution peak corresponds to the stimulus velocity. Note that the breadth of this response distribution does not imply a poor ability to distinguish different pattern velocities based on the population response, just as the broad spectral sensitivity of human photoreceptors does not imply poor ability to distinguish monochromatic lights.

The collection of MT responses provides a population code for local image velocity, similar to those that have been studied in motor areas. For example, population coding of motor commands has been investigated by Sparks *et al.* (1976) for saccadic eye movements, by Groh *et al.* (1995) for pursuit eye movements, and by Georgopoulos *et al.* (1986) for arm movements.

Figure 6 shows the distribution of MT responses for a drifting sine grating stimulus. Note that the response distribution is elongated: Such an elongation will occur for any one-dimensional stimulus pattern (e.g., an extended edge, grating, or striped pattern) because the two-dimensional velocity of any such pattern is inherently ambiguous. The ambiguity of one-dimensional patterns has been termed the "aperture problem", since the motion of a moving one-dimensional pattern viewed through a small circular aperture is ambiguous. Of course, the problem is not really due to the aperture, but to the one-dimensional structure of the stimulus.

Figure 7 shows that the motion is disambiguated when there is more spatial structure. A sine grating plaid stimulus is composed of two moving gratings. Each grat-

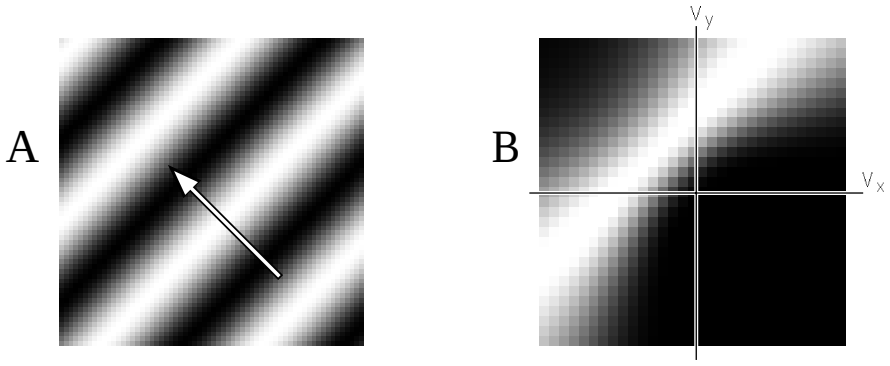

FIGURE 6: A: Drifting sinusoidal grating stimulus. **B:** Distribution of model MT neuron responses (same format as Fig. 5). The distribution is elongated because the stimulus velocity is inherently ambiguous.

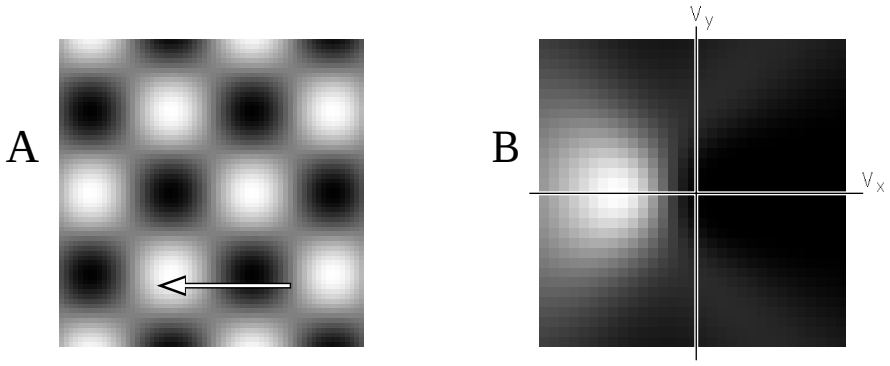

FIGURE 7: A: Plaid stimulus composed of two drifting sinusoidal gratings. **B:** Distribution of model MT neuron responses (same format as Fig. 5).

ing, displayed by itself, would evoke an elongated (ridge shaped) response distribution. When the two gratings are superimposed to produce a plaid, there is a peak in the response distribution at the intersection of the two ridges. We note, however, that the individual ridges associated with the two gratings are still visible.

Figure 8 shows the response distribution for a transparent motion stimulus composed of two texture fields moving in different directions. Large responses are evoked in two different subsets of velocity-selective MT neurons, each tuned for one of the texture velocities. We do not address the interpretation of population responses in this paper, but bimodal responses such as this suggest that interpretation of sensory population codes might be different than that of motor populations. In particular, there is experimental evidence that the mean of the response distribution (also known as the “vector average”) of a motor population ultimately determines the motor activity (see references above). Since stimuli such as that of Fig. 8 are known to produce multiple motion percepts (i.e., motion transparency) in human observers, condensing the population response to a single peak or mean value may be inappropriate. It seems plausible that the full response distribution might be propagated forward to later stages of processing and ultimately used to

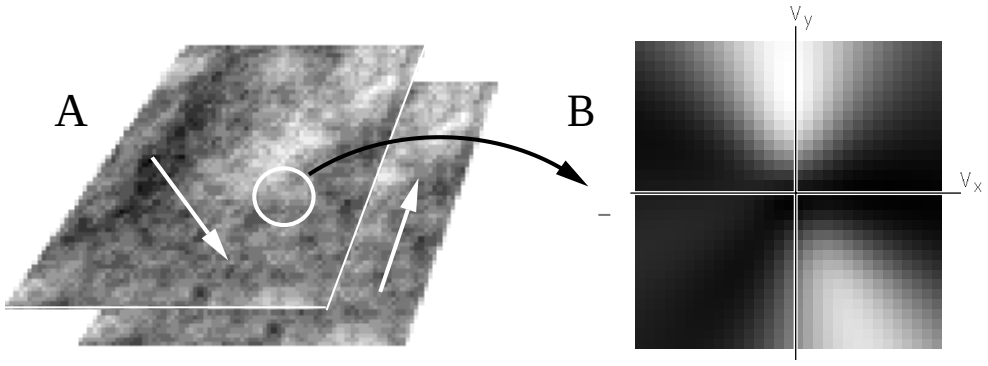

FIGURE 8: A: Transparent motion stimulus composed of a sum of two random texture fields moving in different directions. **B:** Distribution of model MT neuron responses (same format as Fig. 5). The bimodal population response distribution represents both motions.

drive behavior (i.e., psychophysical judgments).

3 Results

As described in the previous section, the model is designed to represent velocity via the relative responses of neurons within a population. Thus the most direct method of examining the model's behavior is to compute responses of all neurons in the population to a single stimulus. In order to compare with single-cell electrophysiological data, however, we must examine the responses of a *single* model neuron to a parameterized set of stimuli. This section reviews the single-cell physiology of MT neurons and compares the behavior of model cells with that of real cells.

3.1 Direction Tuning

A V1 neuron will respond to complex stimuli containing multiple oriented components when any one of the components is near the neuron's preferred orientation and direction of motion (Movshon *et al.*, 1986; Gizzi *et al.*, 1990). Figures 9A and 9B show polar plots of direction tuning for a particular V1 neuron using two stimuli: (1) drifting sinusoidal gratings, and (2) drifting plaid patterns composed of two gratings. The direction tuning curve for gratings is unimodal, but the direction tuning curve for plaids shows two distinct lobes. Each lobe corresponds to one of the plaid's component gratings. Movshon *et al.* (1986) used the phrase "component-motion selective" to characterize this behavior, which was common to all of the V1 neurons in their sample. Model V1 neurons behave similarly, as shown in Figs. 9C and 9D.

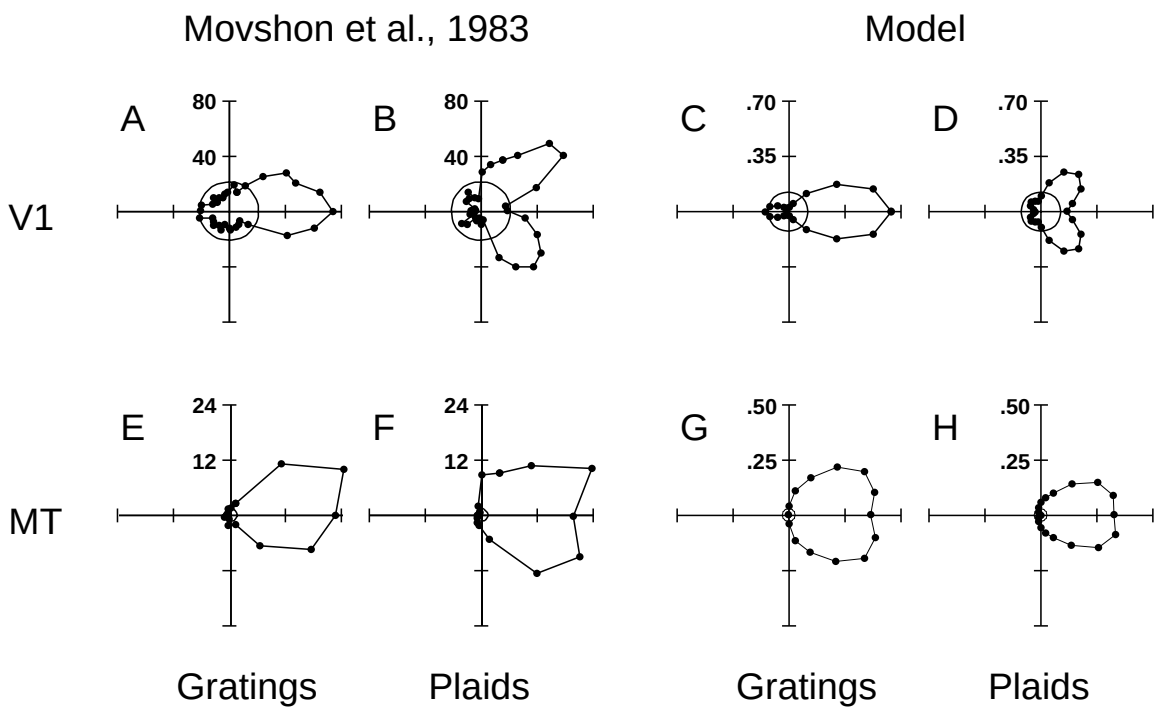

FIGURE 9: Component-motion selectivity and pattern-motion selectivity. **A, B, C, D:** Direction tuning curves of a real V1 neuron (re-plotted from Movshon *et al.*, 1986) and a model V1 neuron. Stimuli are drifting gratings, and plaid patterns composed of two gratings. Response is plotted radially and the direction of stimulus motion is indicated by the angular coordinate. Circles indicate the spontaneous firing rates. The direction tuning for plaids is bimodal, indicating that these neurons respond separately to the motions of the two component gratings. **E, F, G, H:** Direction tuning curves for a real MT neuron (re-plotted from Movshon *et al.*, 1986) and a model MT neuron. The direction tuning curves for plaids are unimodal indicating that these neurons respond to the combined motion of the entire plaid pattern, not to the motions of the component gratings. Pattern-motion selectivity arises in the model because each model MT neuron combines inputs from several V1 afferents, each selective for a different component motion (see text).

In area MT, Movshon *et al.* (1986) found that roughly one third of the neurons were component-motion selective. But another third exhibited a different behavior, in which the direction tuning curves for grating and plaid stimuli were similar. They used the phrase "pattern-motion selective" to describe this behavior. An example is shown in Figs. 9E and 9F. This MT neuron responded to the motion of the entire plaid pattern rather than to the motions of the individual component gratings. Figures 9G and 9H show that an example model MT pattern neuron behaves similarly. In order to match the spontaneous rate of the model neuron to this particular real neuron, we had to use a modified value for the additive constant: $\alpha_2 = 0.15$.

Albright (1984) classified MT neurons into two types using an alternative criterion. Both types are selective for the direction of motion of a dot field and for the orientation of a (flashed, stationary) bar. For Type I neurons, the preferred dot direction

is perpendicular to the preferred bar orientation. For Type II neurons, however, the preferred dot direction is parallel to the preferred bar orientation. Rodman and Albright (1989) demonstrated that the Type II classification is highly correlated with Movshon *et al.*'s pattern-motion classification. The behavior of Type II neurons is also consistent with the velocity selectivity of our model MT neurons; the neurons "interpret" a stationary bar as if it is moving parallel to its orientation.

3.2 Speed Tuning

The speed tuning properties of MT neurons have been measured by a number of researchers (e.g., Maunsell & van Essen, 1983a; Lagae *et al.*, 1993). Rodman and Albright (1987) divided MT neurons into three distinct classes based on their speed tuning properties. Figures 10A, 10C, and 10E show speed tuning curves of representative neurons. The first neuron (Fig. 10A) was broadly tuned for speed when the stimulus moved in the preferred direction; the tuning width is approximately three octaves at half-height. This is comparable to tuning widths measured in other labs (e.g., Maunsell and van Essen, 1983), although for both real and model neurons the tuning width depends on the particular neuron being examined. In addition, this neuron was suppressed by motion in the anti-preferred (opposite) direction; the suppression was strongest when the stimulus moved in the opposite direction at roughly the preferred speed.

The other two neurons (Figs. 10C and 10E) responded to motion in both preferred *and* anti-preferred directions. One of them (Fig. 10C) exhibited a preference for low speeds in both directions. The other (Fig. 10E) responded to high speed stimuli in both directions. In addition to the behaviors depicted here, Mikami *et al.* (1986a) reported that some MT neurons are suppressed by motion in the preferred direction, but at non-preferred speeds.

The model is capable of producing quite similar speed tuning curves, as illustrated in Figs. 10B, 10D, and 10F. The shapes of the simulated speed tuning curves are most easily explained in the Fourier domain. Each model MT neuron receives excitatory input from a particular set of V1 neurons with preferred spatio-temporal frequency lying on a plane in the Fourier domain. Each MT neuron also receives inhibitory input from V1 neurons with preferred spatio-temporal frequency lying off the plane. This subtractive inhibition is responsible for the suppressive effects observed in Fig. 10. The two stages of response normalization in the model have little effect on speed tuning, since changing the speed of the stimulus does not significantly alter the total activity of the neurons in either of the two normalization pools.

The Fourier planes corresponding to low speed motions in opposite directions are both close to the $\omega_t = 0$ plane, and thus close to each other. This means that a model MT neuron with a low preferred speed in one direction (e.g., Fig. 10D) will also re-

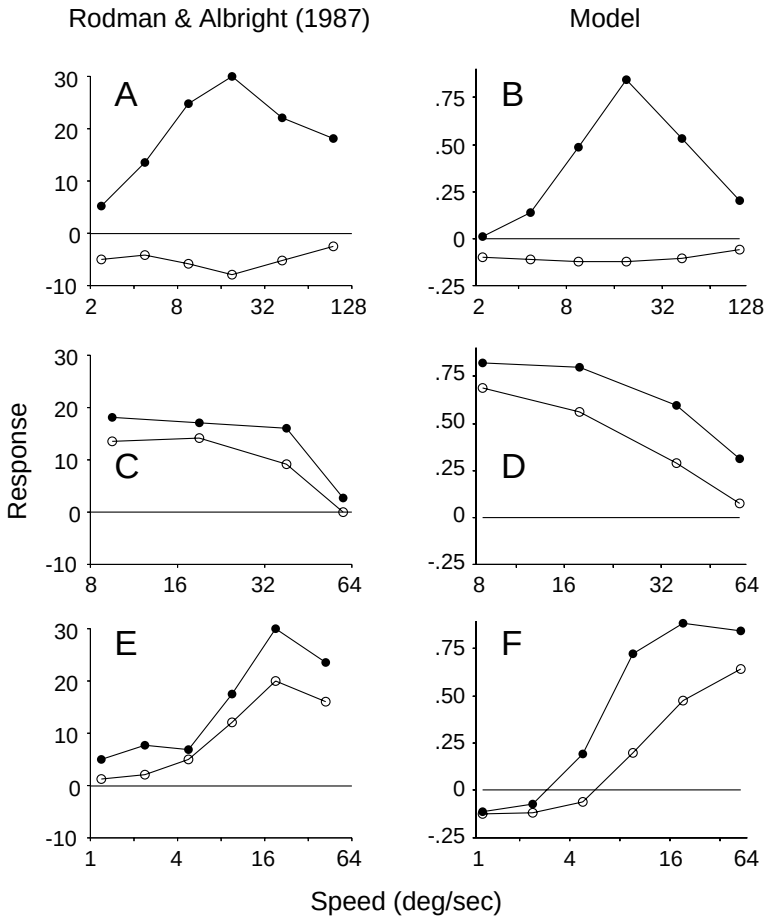

FIGURE 10: A, C, E: Speed tuning curves of MT neurons (re-plotted from Rodman and Albright, 1987) for bar stimuli moving in the preferred (closed symbols) and anti-preferred (open symbols) directions, after subtracting the spontaneous firing rate. The three neurons exhibited different speed tuning characteristics: "speed-tuned", "low-pass", and "high-pass", respectively. **B, D, F:** Speed tuning curves of model MT neurons, tuned for three different speeds. V1 afferents in **D/F** are tuned for spatial frequencies four times lower / higher than those in **B** (see *Appendix*).

spond to low speed stimuli moving in the opposite direction. Such a neuron will be suppressed by fast stimuli (in either direction). Similarly, a neuron with a high preferred speed in one direction (e.g., Fig. 10F) will be excited by fast stimuli moving in the opposite direction and also by flickering stimuli, but will be suppressed by slow stimuli in either direction.[3]

[3] The definitions of "fast" and "slow" depend on the relative spatial and temporal frequency tuning of the underlying population of V1 neurons.

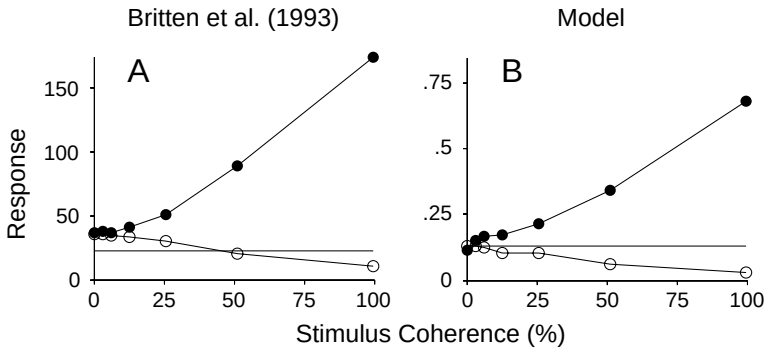

FIGURE 11: MT response as a function of motion signal strength. **A:** Response of an MT neuron as a function of the percentage of coherently moving dots (re-plotted from Britten *et al.*, 1993). Horizontal line is the spontaneous firing rate. Filled symbols correspond to motion in the preferred direction. Open symbols correspond to motion in the anti-preferred (opposite) direction. **B:** Simulated responses of a model MT neuron.

3.3 Response versus Motion Signal Strength

Newsome and colleagues (Newsome *et al.*, 1989; Salzman *et al.*, 1990, 1992; Britten *et al.*, 1992) recorded activity of MT neurons in response to stimuli consisting of a field of coherently moving dots superimposed on a field of randomly moving dots. The strength of the motion signal was controlled by varying the ratio of coherent to random dots.

Figure 11 shows the response of an MT neuron as a function of motion signal strength (percentage of coherently moving dots). In both the real and simulated data sets, the response function of most cells rises nearly linearly with stimulus coherence for motion in the preferred direction, and falls nearly linearly with stimulus coherence for motion in the opposite direction.

Britten *et al.* (1993) explain that an opponent motion energy neuron would behave similarly provided that its spatio-temporal frequency bandwidth is broad enough. The linear increase / decrease in response arises because the distribution of spectral power depends on stimulus coherence. For 100% coherence, the spectral power is concentrated on a plane in the spatio-temporal frequency domain. For 0% coherence, the spectral power is distributed uniformly throughout the frequency domain. Between these two extremes, the percentage of spectral power in the plane is proportional to the coherence level. Thus, the response of an energy neuron will be proportional to stimulus coherence.

The suppression of response below the spontaneous rate is due primarily to the inhibitory region of the MT receptive field. The slight bend in the model response curves is due to the squaring and normalization of the MT stage of the model. The V1 normalization factor depends very little on coherence level: Although the *distribution* of spectral power varies with coherence level, the *total* power does not.

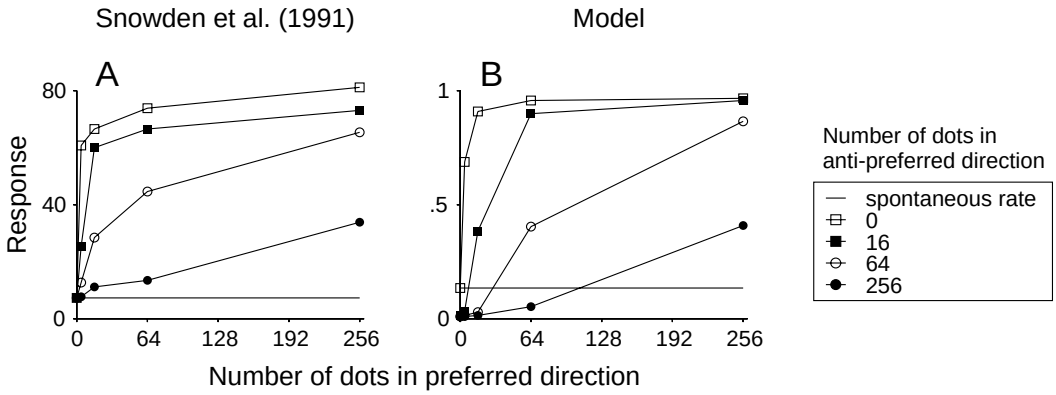

FIGURE 12: Suppression of MT responses. **A:** Responses of an MT neuron to superimposed pairs of drifting dot fields (re-plotted from Snowden *et al.*, 1991). One field moved in the preferred direction and the other moved in the anti-preferred (opposite) direction. The number of dots drifting in each directions was varied. Response was suppressed in the presence of anti-preferred dots. **B:** Simulated responses of a model MT neuron.

There is a minor discrepancy between the the real and simulated data in Fig. 11. Most MT neurons respond well above their spontaneous rate for the 0% coherence stimulus. For model MT neurons, the spontaneous rate and the 0% coherence rate are equal. In our current implementation, the excitatory and inhibitory weights in the MT receptive fields are exactly balanced. Since the 0% coherence stimulus has a flat power spectral density, it will excite all of the V1 neurons equally, and thus will provide equal amounts of excitation and inhibition to each MT cell. One could introduce a model parameter to control the balance between excitatory and inhibitory weights, in order to try to fit these data quantitatively.

3.4 Suppression by Non-Preferred Motions

A number of physiological experiments conclude that opponency (mutual inhibition and / or suppression between cells tuned for opposite directions of motion) is an essential aspect of the behavior of MT neurons (Mikami *et al.*, 1986; Rodman & Albright, 1987; Snowden *et al.*, 1991; Qian & Andersen, 1994). Snowden *et al.*, for example, investigated this suppression by recording MT neuronal activity in response to superimposed pairs of drifting dot fields. One dot field moved in the preferred direction and the other dot field moved at the same speed in the anti-preferred (opposite) direction. The curves in Fig. 12 show the response of an MT neuron as a function of the density, or number, of dots drifting in the preferred direction. Each curve corresponds to a different density of anti-preferred dots. The responses of both real and model MT neurons show suppression in the presence of a dot field drifting in the anti-preferred direction.

The computation embodied in our model includes two types of motion opponency: subtractive (due to the underlying weighting of V1 afferents in the MT

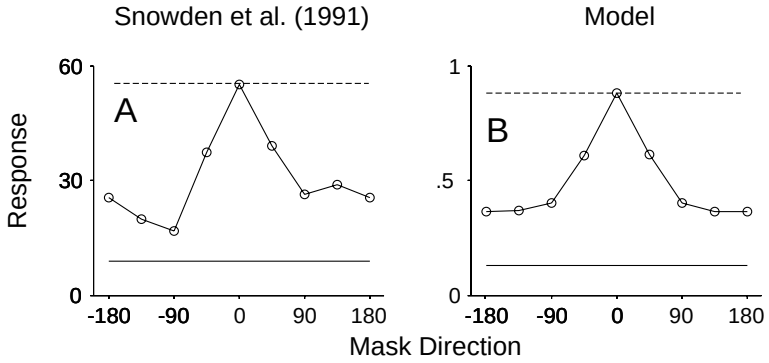

FIGURE 13: A: Responses of an MT neuron to superimposed pairs of drifting dot fields (re-plotted from Snowden *et al.*, 1991). One field moved in the preferred direction and the other (mask) field moved in variable directions. Dashed horizontal line indicates the response to the preferred dot field alone. Solid horizontal line indicates the spontaneous firing rate. Response was suppressed when the second (mask) dot field moved in non-preferred directions. **B:** Simulated responses of a model MT neuron.

stage of the model) and divisive (due to the normalization in both stages of the model). Both of these types of motion opponency contribute to the suppressive behavior observed in Fig. 12B. The change in gain (slope) of the curves is due to divisive opponency, as was suggested by Snowden *et al.*(1991). The suppression below the spontaneous firing rate for stimuli containing mostly anti-preferred motions is primarily due to the inhibitory portion of the MT receptive field weights. This behavior is not seen in the real neuron. But, as in the previous simulations, the model behavior could be adjusted by introducing a parameter to vary the balance of excitatory and inhibitory weighting.

Figure 13 demonstrates motion opponency in another manner. The dashed horizontal line indicates the response to a dot field moving in the preferred direction. The solid curve shows the effect of superimposing a second (mask) dot field. The response of the neuron was suppressed by the presence of a second field moving in a non-preferred direction.

Figure 14 demonstrates this suppression in yet a third manner. Snowden *et al.* (1991) measured the direction tuning for drifting dot fields. They repeated the direction tuning measurements in the presence of a second (masking) dot field moving in the anti-preferred direction. Responses in the presence of this masking stimulus (open symbols) are suppressed below those of the original direction tuning curve (closed symbols).

Finally, there is some evidence that suppression contributes to the speed and direction tuning of MT neurons (Mikami *et al.*, 1986; Rodman & Albright, 1987). As discussed above (see *Speed Tuning*), suppression does indeed contribute to simulated tuning curves. For some model MT neurons, the suppression is strongest at the preferred speed, but in the opposite direction (Fig. 10B). For other model

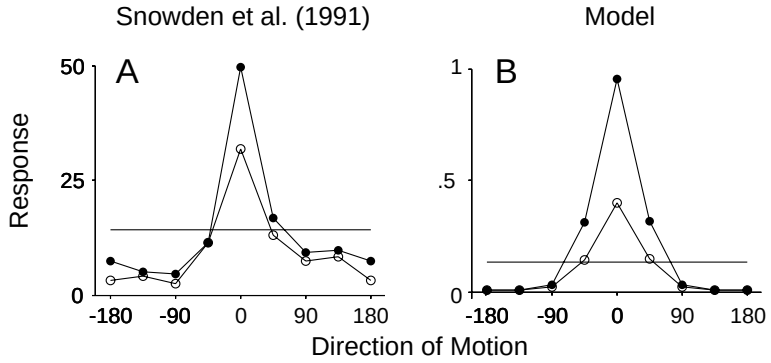

FIGURE 14: A: Responses of an MT neuron to superimposed pairs of drifting dot fields (re-plotted from Snowden *et al.*, 1991). Closed symbols indicate the direction tuning curve for a single drifting dot field stimulus. Open symbols indicate the direction tuning curve when a second (mask) dot field was superimposed, drifting in the anti-preferred direction of motion. Horizontal line is the spontaneous firing rate. **B:** Simulated responses of a model MT neuron.

MT neurons, the suppression is strongest for motion in the preferred direction at non-preferred speeds, either too fast (Fig. 10D) or too slow (Fig. 10F).

4 Discussion

The simulations of the previous section demonstrate that the proposed model is consistent with a variety of physiological data. This section reviews several aspects of the physiology that have not been addressed in the simulations, examines some failures of the current model along with potential solutions, compares the model to models proposed by other researchers, and describes a set of physiological experiments that are motivated by the model.

4.1 Unmodeled Results

Early Nonlinearities. According to the model, V1 complex cells combine the responses of a set of underlying linear receptive field subunits. Hence, we have implicitly assumed that LGN neurons have linear receptive fields. This assumption does not hold for the relevant LGN neurons. In particular, although parvocellular neurons are quite linear, magnocellular neurons exhibit significant nonlinearities (Derrington & Lennie, 1984; Shapley, 1990; Benardete *et al.*, 1992). Parvocellular neurons constitute approximately 90% of the LGN in monkeys, but the majority of MT afferents appear to be magnocellular in origin (Tootell *et al.*, 1988; Maunsell *et al.*, 1990).

This contradicts the precise structure (anatomy) of the present model, but may not have a substantial effect on the simulated physiological responses. The nonlinear-

ity in magnocellular neurons appears to be a gain-control mechanism analogous to the contrast normalization in the V1 stage of the model (Carandini, 1996). Thus, some of the effects of normalization that we have attributed to cortical interactions could be occurring in earlier stages.

Spatio-temporal Frequency Tuning. For V1 neurons, spatial-frequency tuning (measured with drifting grating stimuli) is largely independent of stimulus temporal frequency (e.g., Hamilton *et al.*, 1989), a property that has been referred to as "separability" of spatial- and temporal-frequency tuning. Since speed is the ratio of temporal frequency divided by spatial frequency, this means that a V1 neuron's preferred speed depends on the spatial structure of the stimulus.

For our model V1 neurons, however, the preferred speed remains constant under changes in spatial frequency (at least over the relatively narrow range of spatial frequencies to which the neuron responds). This is due to the particular linear receptive field weights underlying the V1 responses. These receptive field weights were chosen for mathematical and computational convenience, but alternate receptive fields with separable spatio-temporal frequency tuning curves could be used instead (see *Appendix*).

Our model MT neurons also have inseparable spatio-temporal frequency tuning, since they are designed to have a constant preferred speed regardless of the underlying spatial structure of the stimulus. This is consistent with physiological data indicating that some MT neurons maintain a constant preferred speed over a broad range of spatial and temporal frequencies (Newsome *et al.*, 1983).

MT Component Motion Cells. MT neurons have been categorized into two distinct classes: component-motion selective and pattern-motion selective neurons (Movshon *et al.*, 1986). Although we have discussed only the latter category in this paper, MT component cells could also be included in our model, either as an intermediate stage before the pattern cell computation, or as an independent parallel computation. Such a model for component cells is illustrated in Fig. 15. These model neurons combine V1 afferents over spatial position and spatial frequency, but *not* over orientation. Thus, they are orientation- and speed-tuned, but with a broad spatial frequency bandwidth. Model MT pattern cell responses could potentially be computed as a weighted sum of these component cell responses.

Second-Order Motion. "First-order" motion perception models determine velocity primarily from the distribution of power in the Fourier domain. Our model MT neurons, for example, are designed to respond when spectral power is concentrated on a plane through the origin in the spatio-temporal frequency domain.

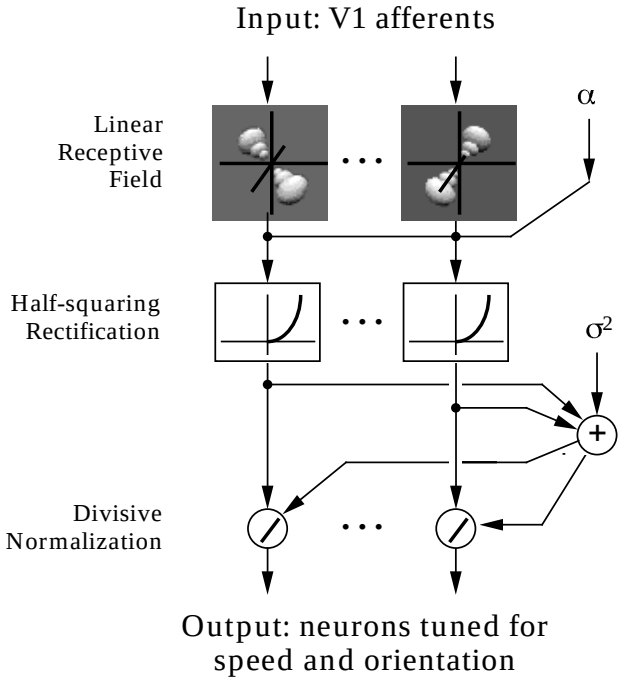

FIGURE 15: Model of MT component cells. Each neuron computes a weighted sum of V1 complex cell afferents, followed by half-squaring, and normalization. The complex cells are all tuned for the same spatial orientation and speed, but a range of spatial frequencies and receptive field locations. A constant, α, is added to provide a spontaneous firing rate. The divisive normalization factor is computed as a sum of all of the half-squared responses and a squared semi-saturation constant, σ.

There are however a number of stimuli, known as "non-Fourier" or "second-order" motion stimuli, whose appearance is inconsistent with a first-order motion model. (Chubb and Sperling, 1988, 1989; Cavanaugh and Mather, 1989; Fleet and Langley, 1994). In the most standard examples, human subjects see movement despite the fact that the spectral power of the stimulus is directionally balanced. This observation has been used to refute simple energy models of motion perception.

The role of MT neurons in the perception of second-order motion is unclear. Albright (1992) found that many MT neurons respond to second-order motion stimuli, and that their selectivity for second-order motion is similar to their selectivity for first-order motion. O'Keefe *et al.* (1993), using a different suite of second-order motion stimuli, found that very few MT neurons respond in a direction-selective manner to second-order motion.

Current models of second-order motion perception are based on the first-order motion models. They include an additional early nonlinearity (typically a form of rectification) which transforms the second-order motion signal into a first-order motion signal. Many of the models also include another linear filter prior to the nonlinearity. A motion response is subsequently computed via an appropriate first-order motion computation (Chubb and Sperling, 1988, 1989; Cavanaugh and Mather,

1989; Wilson *et al.*, 1992, 1994; Fleet and Langley, 1994). In a similar fashion, our model may be extended to properly compute motion of second-order stimuli by inserting an additional (nonlinear) stage of computation.

Motion Coherence. The motion of plaid stimuli, formed via superposition of two drifting periodic gratings, can be perceived as coherent (i.e., as a single moving pattern) or non-coherent (as two separate component gratings), depending on a variety of stimulus parameters (e.g., Adelson & Movshon, 1982; Movshon *et al.*, 1986; Stoner *et al.*, 1990). As shown in Fig. 8, our model can produce a bimodal response distribution for some types of transparently combined stimuli. But the current model cannot directly explain the numerous psychophysical observations concerning coherent versus non-coherent motion.

Stoner and Albright (1992) performed the only published physiological study comparing MT responses for coherent and non-coherent motion stimuli. They used three square-wave plaid patterns, differing only in the intensity of the diamond-shaped regions at the intersections of the dark bars of the two gratings. One of these intensities corresponded to an additive combination of the gratings, and another to a multiplicative combination. The multiplicative pattern appeared non-coherent, and the other two patterns appeared coherent (to human observers). Consistent with this percept, MT direction tuning curves were more component-like for the multiplicative plaids (i.e., the responses in the component directions were larger, and the response in the pattern direction was smaller).

Stoner *et al.* (1990) explain that a model based on stimulus Fourier components predicts that the additive plaid should appear most transparent. Our model also makes this prediction. The response distribution for an additive plaid stimulus is shown in Fig. 7. Note the faint ridges emanating from the peak in the population. Increasing or decreasing the luminance of the grating intersections is equivalent to superimposing a stimulus composed of a moving pattern of diamonds. These moving diamonds are a strong stimulus for an MT neuron. Superimposing this pattern on the plaid stimulus sharpens the response peak (and suppresses the faint ridges), resulting in a sharper direction tuning curve. The point of maximal transparency could be shifted toward the multiplicative plaid by including an early compressive nonlinearity in the model (e.g., a logarithm) that would convert the multiplicative signal combination into an additive one.

Disparity-Dependent Suppression. MT neurons are also selective for binocular disparity (Maunsell & van Essen, 1983b), but we have not included binocular tuning in the current model. In some MT neurons, suppression occurs mainly between motion signals with similar disparities (Bradley *et al.*, 1995). This type of behavior could be replicated by incorporating disparity tuning, and by using separate normalization pools for each range of preferred disparities.

MT Surround Antagonism. The behavior of some MT neurons can be greatly influenced by stimuli outside the classically defined receptive field. Stimuli in these surrounding regions do not by themselves evoke a response, but can strongly modulate the response to a stimulus within the receptive field (Allman *et al.*, 1985a, 1985b; Tanaka *et al.*, 1986; Born and Tootell, 1992). The most common effect is a suppression of the response to optimal stimulation of the classical receptive field when the surround is stimulated by motion in the same direction. For some of these neurons, surround motion in the opposite direction enhances the response to stimulation in the classical receptive field.

Our model currently does not address center-surround antagonism. Suppression in the MT stage of our model is both subtractive (due to the linear combination of V1 afferents) and divisive (due to normalization). Both forms of suppression are spatially restricted to the classical receptive field. The simplest means of incorporating surround antagonism would be to enlarge the scope of the normalization pool to include neurons with surrounding receptive fields and with direction preferences opposite to the preferred direction in the center. This would produce a divisive interaction between the center (classical receptive field) and the surround, and yet stimulation of the surround in isolation would not evoke a response.

4.2 Relationship to Other Models

A number of models have been proposed to describe visual motion computations performed in area MT. The present model incorporates many concepts from these previous models. Below, we describe some of the similarities and differences.

Adelson and Bergen (1985) developed a two-dimensional (x-t) spatio-temporal energy model based on space-time oriented quadrature pairs of linear filters. They subtracted rightward from leftward motion energy to produce an opponent energy response, and demonstrated that this response could be made identical to that of an elaborated Reichardt detector (van Santen and Sperling, 1985), given a suitable choice of filters for each model. Our model shares much of the conceptual framework of the Adelson/Bergen work. In particular, our complex cells compute a form of motion energy, and the linear weighting function in the MT stage of our model has both positive and negative lobes, and therefore may be viewed as computing a form of opponent motion energy. An essential difference is that our model includes an additional spatial (y) dimension, thus allowing orientation tuning in V1 neurons, and an intersection-of-constraints calculation (i.e., summation over orientation) in the MT neurons. In addition, we use a set of third-derivative spatio-temporal filters, and include response normalization.

A number of authors (e.g., Albright, 1984; Heeger, 1987; Grzywacz & Yuille, 1990; Smith & Grzywacz, 1993) have used neural implementations of the intersection-of-constraints construction to describe the behavior of MT neurons. Heeger (1987)

and Grzywacz and Yuille (1990) developed and implemented models for pattern velocity computation based on spatio-temporal energy computed via quadrature pairs of Gabor filters. The present model is conceptually similar, although quite different in implementation. In addition to the use of both subtractive and divisive opponency, we have used directional derivatives for our V1 receptive fields so that responses may be properly interpolated from a small set of neurons. This provides an unbiased velocity representation (Simoncelli, 1993), and allows efficient simulation (see *Appendix*).

Sereno (1993) and Nowlan and Sejnowski (1995) adopted two-stage architectures similar to our own, in which V1 neurons compute motion energy, and MT responses are computed from a linear combination of V1 afferents. Unlike our model, the underlying linear receptive fields of these MT model neurons were determined via an artificial neural network learning algorithm. In addition, the dynamic range of neurons in these models is limited by static sigmoidal nonlinearities, that produce behaviors inconsistent with physiological data (see *Model* for citations). An important contribution of the Nowlan and Sejnowski model is that it performs simultaneous motion estimation and motion-based segmentation.

A number of authors have devised biological models based on the "gradient constraint" that is commonly used in computer vision for motion estimation (e.g., Wang *et al.*, 1989; Johnston *et al.*, 1992; Young & Lesperance, 1993). This constraint provides a relationship between local image velocity and the spatial and temporal derivatives of image intensity. Gradient-based velocity estimation may be written in terms of normalized opponent motion energy computations (Adelson & Bergen, 1986; Simoncelli, 1993), similar to those used in our model. But whereas our model represents both speed and direction implicitly (via a population code), the models referenced above represent speed (and in some cases, direction) explicitly. The resulting MT responses increase monotonically with stimulus speed, inconsistent with speed tuning curves such as those depicted in Figs. 10A or 10C.

Wilson *et al.* (1992, 1994) developed a model with two parallel pathways that compute Fourier motion (via a Reichardt detector) and non-Fourier motion (via a Reichardt detector preceded by rectified oriented filtering). These pathways include a type of divisive normalization. The MT stage of the model combines the responses of these two pathways. Competitive inhibition allows the most active neurons to encode the motion direction. This model differs markedly from our own. The Reichardt motion detectors are spatially one-dimensional, and there are two parallel computations which are fused in the output. Most importantly, these model neurons encode direction, but not speed.

Qian *et al.* (1994a) observed that MT responses are strongly suppressed by superimposing drifting dot fields, particularly when the dots moving in opposite directions were "paired". The authors explained these results using a two-dimensional (x-t) opponent motion energy model, in which they incorporated either subtractive

or divisive opponency. Our model differs by incorporating a second spatial (y) dimension, by using a different set of spatio-temporal filters, and by incorporating both subtractive and divisive opponency.

4.3 Proposed Experiments

An important role for a computational model of this type is to provide a framework for experimentally characterizing the physiological behavior of an MT neuron. For example, responses of an MT neuron to drifting sinusoidal gratings of varying spatio-temporal frequency could be used to estimate MT receptive field weights (including the subtractive inhibitory regions). This could be checked for consistency against the velocity tuning of the cell, as measured with texture stimuli drifting at a range of speeds and directions. In addition to providing this type of characterization framework, a number of predictions are made by the model which could be examined experimentally:

- The model predicts that speeds evoking the maximal suppression should vary inversely with the preferred speed for MT neurons. That is, neurons with high preferred speeds should be suppressed most by slow stimuli moving in the anti-preferred direction, and vice versa. In addition, neurons with high preferred speeds should give an excitatory response to fast stimuli in the "anti-preferred" direction, and show some suppression for slow stimuli in the preferred direction. Similarly, neurons with slow preferred speeds should respond to slow stimuli in the opposite direction, and should be suppressed by high speeds in the preferred direction. This is due to the construction of the MT linear weights. As shown in Fig. 10, this behavior is consistent with some data (Rodman and Albright, 1987; Lagae *et al.*, 1993), but it has not been studied systematically.
- The model predicts that broadband stimuli (such as drifting dots) should elicit the strongest responses from pattern cells, but that sinusoidal stimuli (of equivalent spectral power) should produce more effective suppression. This is due to the linear weighting of our model MT neurons, which subtracts spectral power lying off of the plane. A broadband stimulus moving in the anti-preferred direction will simultaneously contribute some suppression and a small amount of excitation. A sinusoidal stimulus moving in the anti-preferred direction will only contribute suppressively.
- The model predicts that MT pattern cells should have bimodal direction-tuning curves for gratings moving significantly slower than the preferred speed. Albright (1984) made a similar prediction, but subsequent investigations (Rodman & Albright, 1987) did not find the predicted behavior in a sample of 13 "Type II" neurons. Rubin and Hochstein (1992) concluded from these findings that IOC-motivated models of MT neurons were probably incorrect, and proposed an alternative strategy. But the stimuli in the Rodman and Albright (1987) experiments were moving bars, rather than gratings. Preliminary data using drifting

sine gratings (Simoncelli *et al.*, 1996) indicate that some MT pattern cells clearly exhibit the predicted bimodality.

- The model also predicts that component motion cells (both V1 complex cells and MT component cells) should have unimodal direction tuning curves for gratings of all speeds, and bimodal direction tuning curves for rapidly moving texture patterns. This behavior is typical of complex cells in cat area 17 (Hammond & Reck, 1980; Hammond & Smith, 1983). Preliminary data (Simoncelli *et al.*, 1996) indicate that some MT component cells exhibit this behavior.

4.4 Conclusions

We have presented a two-stage physiological model for local image velocity representation in visual area MT. The model is functionally motivated, and its behavior is determined by a small number of free parameters. The computations are simple enough that one can understand and make predictions about its behavior. An additional feature is the commonality of computation in the two sequential stages. It is often noted that neocortical areas throughout the cortex exhibit common structure: The types, arrangements, and connections of neurons are highly stereotyped. Despite its simplicity, the model is able to account for much of the physiology of MT neurons.

One drawback of this simplicity is that the population of model neurons is unrealistically homogeneous. For example, the underlying linear receptive field of each V1 cell is perfectly anti-symmetric, and there are identical collections of neurons at each spatial position (i.e., in each "hypercolumn" of V1 and MT). Receptive fields of real neurons are quite irregular in comparison. Heterogeneity could be incorporated into the model (at great computational expense) without altering the overall functionality. We could, for example, use a different collection of linear receptive fields for each patch of the visual field. Each new collection of linear receptive fields would be related to our current collection of linear receptive fields via an invertible linear transformation, and would thus represent the same information in a modified format.

In addition to the potential model enhancements described in *Unmodeled Results*, we note that the squaring nonlinearities could be replaced by fullwave-rectification (Pollen & Ronner, 1983), or an exponent greater than 2 (Albrecht *et al.*, 1984; Sclar *et al.*, 1990; Albrecht & Geisler, 1991), if warranted by the physiological data. Such modifications would make the simulations more expensive computationally, and would induce small quantitative changes. But the qualitative behavior of the model would be unaffected.

One notable deficiency of the model is the lack of realistic temporal dynamics. In the current implementation, outputs correspond to steady-state firing rates. The model should be extended to include more realistic temporal behavior. For ex-

ample, one might try to model adaptation by including a slow temporal average in the computation of the normalization factor. Experimental examination of neuronal response dynamics may provide evidence for both the functional form and the implementation of these computations (Carandini, 1996).

Finally, we believe this model should be able to account for a number of psychophysical findings. Previous incarnations of the model have been used to account for biases in the perception of speed and direction of plaid patterns (Simoncelli & Heeger, 1992; Simoncelli, 1993; Heeger & Simoncelli, 1993). In order to use the current model for such purposes, one must compute an estimated (or "perceived") velocity from the responses of the MT population.

References

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*, 284–299.

Adelson, E. H. & Bergen, J. R. (1986). The extraction of spatio-temporal energy in human and machine vision. In *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, pages 151–156, Charleston, S Carolina.

Adelson, E. H. & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*(5892), 523–525.

Albrecht, D. G., Farrar, S. B., & Hamilton, D. B. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *Journal of Physiology (London)*, *347*, 713–739.

Albrecht, D. G. & Geisler, W. S. (1991). Motion sensitivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, *7*, 531–546.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*, 1106–1130.

Albright, T. D. (1992). Form-cue invariant motion processing in primate visual cortex. *Science*, *255*, 1141–1143.

Albright, T. D. (1993). Cortical processing of visual motion. In Miles, F. A. & Wallman, J., editors, *Visual Motion and its Role in the Stabilization of Gaze*, pages 177–201. Elsevier Science Publishers.

Albright, T. D. & Desimone, R. (1987). Local precision of visutopic organization in the middle temporal area (MT) of the macaque. *Experimental Brain Research*, *65*, 582–592.

Allman, J., Miezin, F., & McGuinness, E. (1985a). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area. *Perception*, *14*, 105–126.

Allman, J., Miezin, F., & McGuinness, E. (1985b). Stimulus specific responses from beyond the classical receptive field. *Ann. Rev. Neurosci.*, *8*, 407–428.

Benardete, E. A., Kaplan, E., & Knight, B. W. (1992). Contrast gain control in the primate retina: P cells are not X-like, some M cells are. *Visual Neuroscience*, *8*, 483–486.

Bonds, A. B. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, *2*, 41–55.

Born, R. T. & Tootell, R. B. (1992). Segregation of global and local motion processing in middle temporal visual area. *Nature*, *357*, 497–499.

Bradley, D. C., Qian, N., & Andersen, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature*, *373*, 609–611.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, *12*, 4745–4765.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, *10*, 1157–1169.

Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London B, 227*, 249–265.

Campbell, F. W., Cleland, B. G., Cooper, G. F., & Enroth-Cugell, C. (1968). The angular selectivity of visual cortical cells to moving gratings. *Journal of Physiology (London), 198*, 237–250.

Campbell, F. W., Cooper, G. F., & Enroth-Cugell, C. (1969). The spatial selectivity of visual cells of the cat. *Journal of Physiology (London), 203*, 223–235.

Carandini, M. (1996). *Linearity, Gain Control and Spike Encoding in the Primary Visual Cortex*. PhD thesis, New York University, Center for Neural Science, New York, NY.

Carandini, M. & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science, 264*, 1333–1336.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience, 17*, 8621–8644.

Cavanaugh, P. & Mather, G. (1989). Motion: the long and short of it. *Spatial Vision, 4*, 103–129.

Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-fourier motion perception. *Journal of the Optical Society of America A, 5*, 1986–2006.

Chubb, C. & Sperling, G. (1989). Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proceedings of the National Academy of Science USA, 86*, 2985–2989.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). The spatiotemporal organization of simple cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology, 69*, 1118–1135.

DeAngelis, G. C., Robson, J. G., Ohzawa, I., & Freeman, R. D. (1992). The organization of suppression in receptive fields of neurons in the cat's visual cortex. *Journal of Neurophysiology, 68*, 144–163.

Derrington, A. M. & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *Journal of Physiology (London), 357*, 219–240.

Dubner, R. & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus of the monkey. *Brain Research, 35*, 528–532.

Dursteler, M. R., Newsome, W. T., & Wurtz, R. H. (1986). Directional and retinotopic pursuit deficits following lesions of the foveal representation within the superior temporal sulcus of the macaque monkey. *Journal of Neurophysiology, 57*, 1262–.

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32*, 203–218.

Fahle, M. & Poggio, T. (1981). Visual hyperacuity: spatiotemporal interpolation in human vision. *Proceedings of the Royal Society of London, B, 213*, 451–477.

Fennema, C. L. & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Vision, Graphics, and Image Processing, 9*, 301–315.

Fleet, D. J. & Langley, K. (1994). Computational analysis of non-fourier motion. *Vision Research, 22*, 3057–3079.

Fleet, D. J. & Langley, K. (1995). Recursive filters for optical flow. *IEEE Pattern Analysis and Machine Intelligence, 17*, 61–67.

Freeman, W. T. & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Pattern Analysis and Machine Intelligence, 13*, 891–906.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Nature, 233*, 1416–1419.

Gizzi, M. S., Katz, E., Schumer, R. A., & Movshon, J. A. (1990). Selectivity for orientation and direction of motion of single neurons in cat striate and extrastriate visual cortex. *Journal of Neurophysiology, 63*, 1529–1543.

Groh, J. M., Born, R. T., & Newsome, W. T. (1995). Microstimulation of area MT affects both saccades and smooth pursuit eye movements. *Soc. Neurosci. Abstr., 21*, 281.

Grzywacz, N. M. & Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London A, 239*, 129–161.

Hamilton, D. B., Albrecht, D. G., & Geisler, W. S. (1989). Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function. *Vision Research, 29*, 1285–1308.

Hammond, P. & Reck, J. (1980). Influence of velocity on directional tuning of complex cells in cat striate cortex for texture motion. *Neuroscience Letters, 19*, 309–314.

Hammond, P. & Smith, A. T. (1983). Directional tuning interactions between moving oriented and textured stimuli in complex cells of feline striate cortex. *Journal of Physiology (London), 342*, 35–49.

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A, 4*, 1455–1471.

Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In Landy, M. & Movshon, J. A., editors, *Computational Models of Visual Processing*, pages 119–133. MIT Press, Cambridge, MA.

Heeger, D. J. (1992a). Half-squaring in responses of cat simple cells. *Visual Neuroscience, 9*, 427–443.

Heeger, D. J. (1992b). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–198.

Heeger, D. J. (1993). Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology, 70*, 1885–1898.

Heeger, D. J. & Simoncelli, E. P. (1993). Model of visual motion sensing. In Harris, L. & Jenkin, M., editors, *Spatial Vision in Humans and Robots*, chapter 19, pages 367–392. Cambridge University Press, New York.

Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (1996). Computational models of

cortical visual processing. *Proceedings of the National Academy of Science USA, 93*, 623–627.

Hubel, D. & Wiesel, T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology (London), 160*, 106–154.

Johnston, A., McOwan, P. W., & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc. R. Soc. Lond. B, 250*, 297–306.

Koenderink, J. J. & van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biol. Cybern., 55*, 367–375.

Lagae, L., Raiguel, S., & Orban, G. A. (1993). Speed and direction selectivity of macaque middle temporal neurons. *Journal of Neurophysiology, 69*, 19–39.

Logothetis, N. K. & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science, 245*, 761–763.

Maunsell, J. H. R., Nealy, T. A., & DePriest, D. D. (1990). Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *Journal of Neuroscience, 10*, 3323–3334.

Maunsell, J. H. R. & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience., 10*, 363–401.

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

Maunsell, J. H. R. & van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49*, 1148–1167.

Maunsell, J. H. R. & van Essen, D. C. (1987). Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries. *J. Comp. Neurol., 266*, 535–555.

McLean, J. & Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Research, 29*, 675–679.

McLean, J., Raab, S., & Palmer, L. A. (1994). Contribution of linear mechanisms to the specification of local motion by simple cells in areas 17 and 18 of the cat. *Visual Neuroscience, 11*, 271–294.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. i. Mechanisms of direction and speed selectivity in extrastriate area MT. *Journal of Neurophysiology, 55*, 1308–1327.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1986). The analysis of moving visual patterns. In Chagas, C., Gattass, R., & Gross, C., editors, *Experimental Brain Research Supplementum II: Pattern Recognition Mechanisms*, pages 117–151. Springer-Verlag, New York.

Movshon, J. A., Lisberger, S. G., & Krauzlis, R. J. (1990). Visual cortical signals supporting smooth pursuit eye movements. *Cold Spring Harbor Symposia on*

Quantitative Biology, *15*, 707–716.

Movshon, J. A. & Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *Visual Neuroscience*, *16*(23), 7733–7741.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978a). Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology (London)*, *283*, 79–99.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology (London)*, *283*, 53–77.

Nestares, O. & Heeger, D. J. (1997). Modelling the apparent frequency-specific suppression in simple cells responses. *Vision Research*, *37*, 1535–1543.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*, 52–54.

Newsome, W. T., Gizzi, M. S., & Movshon, J. A. (1983). Spatial and temporal properties of neurons in macaque MT. *Investigative Opthalmology and Visual Science Supplement*, *24*, 106.

Newsome, W. T. & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, *8*, 2201–2211.

Newsome, W. T., Wurtz, R. H., Dursteler, M. R., & Mikami, A. (1985). Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *Journal of Neuroscience*, *5*, 825–840.

Nowlan, S. J. & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. *Journal of Neuroscience*, *15*, 1195–1214.

O'Keefe, L. P., Carandini, M., Beusmans, J. M. H., & Movshon, J. A. (1993). MT neuronal responses to 1st- and 2nd-order motion. *Society for Neuroscience Abstracts*, *19*, 1283.

Pollen, D. & Ronner, S. (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*, 907–916.

Qian, N. & Andersen, R. A. (1994). Transparent motion perception as detection of unbalanced motion signals: II. Physiology. *Journal of Neuroscience*, *14*, 7367–7380.

Robson, J. G. (1988). Linear and nonlinear operations in the visual system. *Investigative Opthalmology and Visual Science Supplement*, *29*, 117.

Rodman, H. R. & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Vision Research*, *27*, 2035–2048.

Rodman, H. R. & Albright, T. D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, *75*, 53–64.

Rubin, N. & Hochstein, S. (1992). Global motion computation and the behavior of single cells. In *Neurosci Abstracts*, volume 464.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, *346*, 174–177.

Salzman, C. D., Murasugi, C. M., Britten, K. H., & Newsome, W. T. (1992). Microstimulation in visual area MT: Effects on direction discrimination performance. *Journal of Neurophysiology*, *12*, 2331–2355.

Sclar, G., Maunsell, J. H. R., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, *30*, 1–10.

Sereno, M. E. (1993). *Neural computation of pattern motion: modeling stages of motion anaysis in the primate visual cortex*. MIT Press, Cambridge, MA.

Shapley, R. (1990). Visual sensitivity and parallel retinocortical channels. *Annu. Rev. Psychol.*, *41*, 635–658.

Shapley, R. & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain control. *Progress in Retinal Research*, *3*, 263–346.

Siegel, R. M. & Andersen, R. A. (1986). Motion perceptual deficits following ibotenic acid lesions of the middle temporal area (MT) in the behaving rhesus monkey. *Soc. Neurosci. Abstr.*, *12*, 1183.

Simoncelli, E. P. (1993). *Distributed Analysis and Representation of Visual Motion*. PhD thesis, Massachusetts Institute of Technology, Dept of Electrical Engineering and Computer Science, Cambridge, MA.

Simoncelli, E. P., Bair, W. D., Cavanaugh, J. R., & Movshon, J. A. (1996). Testing and refining a computational model of neural responses in area MT. *Investigative Opthalmology and Visual Science Supplement*, *37*.

Simoncelli, E. P. & Heeger, D. J. (1992). A computational model for perception of two-dimensional pattern velocities. *Investigative Opthalmology and Visual Science Supplement*, *33*.

Simoncelli, E. P. & Heeger, D. J. (1994). A velocity representation model for MT cells. *Investigative Opthalmology and Visual Science Supplement*, *35*, 1827.

Smith, J. A. & Grzywacz, N. M. (1993). A local model for transparent motion based on spatio-temporal filtering. In Eeckman, F. H. & Bower, J. M., editors, *Computation and Neural Systems*, pages 177–181. Kluwer Academic Press, Hingham, MA.

Snowden, R. J., Treue, S., Erikson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience*, *11*, 2768–2785.

Sparks, D. L., Holland, R., & Guthrie, B. L. (1976). Size and distribution of movement fields in the monkey superior colliculus. *Brain Res.*, *113*, 21.

Stoner, G. R. & Albright, T. D. (1992). Neural correlates of perceptual motion coherence. *Nature*, *358*, 412–414.

Stoner, G. R., Albright, T. D., & Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature*, *344*, 153–155.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y., & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience*, *6*, 134–144.

Tolhurst, D. J. & Heeger, D. J. (1997a). Comparison of contrast normalization and threshold models of the responses of simple cells in cat striate cortex. *Visual Neuroscience*, *14*, 293–310.

Tolhurst, D. J. & Heeger, D. J. (1997b). Contrast normalization and a linear model for the directional selectivity of simple cells in cat striate cortex. *Visual Neuroscience*, *14*, 19–26.

Tootell, R. B. H., Hamilton, S. L., & Switkes, E. (1988). Functional anatomy of macaque striate cortex. IV. contrast and magno-parvo streams. *Journal of Neuroscience*, *8*, 1594–1609.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, *2*, 300–321.

Wang, H. T., Mathur, B., & Koch, C. (1989). Computing optical flow in the primate visual system. *Neural Computation*, *1*, 92–103.

Watson, A. B. & Ahumada, A. J. (1983). A look at motion in the frequency domain. In Tsotsos, J. K., editor, *Motion: Perception and representation*, pages 1–10. Association for Computing Machinery, New York.

Watson, A. B. & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, *2*, 322–342.

Wilson, H. R., Ferrara, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, *9*, 79–97.

Wilson, H. R. & Kim, J. (1994). A model for motion coherence and transparency. *Visual Neuroscience*, *11*, 1205–1220.

Young, R. A. & Lesperance, R. M. (1993). A physiological model of motion analysis for machine vision. In *Proc SPIE*, volume 1913, pages 48–123.

5 Appendix

This appendix describes additional mathematical and implementational details of the model.

5.1 V1 Linear Receptive Field Weights

The linear receptive field weights underlying the V1 responses are designed to satisfy two fundamental constraints:

– The sum of the V1 responses (as a function of stimulus spatio-temporal frequency) should be roughly constant over a desired frequency range. This function corresponds to the spatio-temporal frequency sensitivity of the overall system. We refer to this as the *tiling* property.
– The set of V1 neurons includes receptive fields tuned for a predetermined set of space-time orientations, but the response of a V1 neuron tuned to any intermediate space-time orientation may be precisely interpolated from this set. In other words, the information represented by the V1 neurons does not depend on the specific choice of space-time orientations. We refer to this as the *interpolation* property.

These two constraints are, in fact, independent. Both constraints are satisfied by linear receptive fields that are directional derivatives of a differentiable function, $g(x, y, t)$. For example, the first derivative taken in the direction of a unit vector $\hat{u}$ may be written as:

$$D_u\{g(x,y,t)\} = u_x \frac{\partial g(x,y,t)}{\partial x} + u_y \frac{\partial g(x,y,t)}{\partial y}, + u_t \frac{\partial g(x,y,t)}{\partial t},$$

where the $\{u_x, u_y, u_t\}$ are the components of the unit vector $\hat{u}$. This equation states that the partial derivative in an arbitrary direction $\hat{u}$ is a linear combination of the partial derivatives in the x, y, and t directions. In other words, it is possible to interpolate any directional first derivative as a linear combination of a fixed set of three derivatives.

The tiling property of the first derivatives is easily seen in the frequency domain. Computing a partial derivative (of space or time) corresponds to multiplication by a linear ramp function in the frequency domain. For example, the Fourier transform of the x derivative of $g(x, y, t)$ is:

$$\mathcal{F}\left\{\frac{\partial g(x,y,t)}{\partial x}\right\} = -i\omega_x G(\omega_x, \omega_y, \omega_t),$$

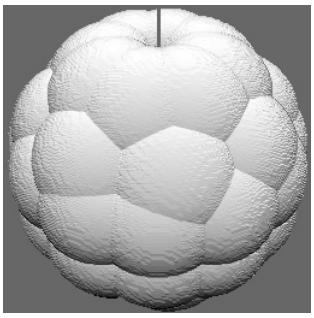

FIGURE 16: Distribution of the Fourier power spectra corresponding to our population of 28 V1 neurons. The spectra lie on the surface of a sphere in the spatio-temporal frequency domain, and correspond to space-time oriented linear receptive fields. The vertical line at the top indicates the temporal frequency axis.

where i is the imaginary number, and $G(\omega_x, \omega_y, \omega_t)$ is the Fourier transform of $g(x, y, t)$. Summing the power spectra of the derivatives in the (x, y, t) directions gives:

$$\begin{aligned} &|\omega_x G(\omega_x, \omega_y, \omega_t)|^2 + |\omega_y G(\omega_x, \omega_y, \omega_t)|^2 + |\omega_t G(\omega_x, \omega_y, \omega_t)|^2 \\ &\qquad = \ [\omega_x^2 + \omega_y^2 + \omega_t^2] \ |G(\omega_x, \omega_y, \omega_t)|^2 \\ &\qquad = \ \omega_r^2 \ |G(\omega_x, \omega_y, \omega_t)|^2 , \end{aligned}$$

where $\omega_r = \sqrt{\omega_x^2 + \omega_y^2 + \omega_t^2}$, the radial frequency. A suitable choice of $g(x, y, t)$ can give nearly any desired coverage of spatio-temporal frequency. For example, $G(\omega_x, \omega_y, \omega_t) = 1/\omega_r$ produces uniform coverage of the entire spatio-temporal frequency domain.

The tiling and interpolation properties extend to higher-order (either separable or directional) derivatives (Freeman & Adelson, 1991). Higher-order directional derivatives have narrower directional tuning curves than the first derivatives, and require a larger set to tile and interpolate all space-time orientations. In particular, in three dimensions, (x, y, t), a full set of Nth-order directional derivatives is of size $(N+1)(N+2)/2$. The interpolation property also extends to the squared directional derivatives (Simoncelli, 1993), but one needs to use an even larger set.

The V1 linear receptive fields in our current model are third derivatives of a spatio-temporal Gaussian.[4] The derivative order is chosen to match the typical orientation bandwidths of V1 neurons (e.g., see Fig. 9). Receptive fields at different scales are formed by adjusting the variance of the underlying Gaussian by factors of 2. The full set of V1 receptive fields includes 28 space-time orientations that are evenly distributed on the surface of a sphere in the spatio-temporal frequency domain, as illustrated in Fig. 16. The tiling region is approximately a spherical an-

[4] See Koenderink (1987) for a general characterization of receptive fields in terms of differential operators.

nulus, with bandwidth of roughly 1.5 octaves, and with reduced sensitivity near the temporal frequency axis. Frequency units are fixed such that the peak of the annulus crosses the temporal frequency axis at $\omega_t = 8$ cycles / sec and the spatial frequency axes at $\omega_x = 0.5$ cycles / deg. The spatial frequency intercept was increased / decreased by a factor of 4 to give the results shown in Figs. 10D / F, respectively.

Note that the choice of a Gaussian function for $g(x, y, t)$ is computationally convenient, as it allows separable convolution computations. But it has three clear disadvantages. First, Gaussian derivatives at different scales produce an uneven tiling of the Fourier domain. Second, the resulting spatial and temporal frequency tuning curves are not separable; in fact, the spatio-temporal frequency responses are polar-separable. As mentioned in the *Discussion*, this is inconsistent with V1 physiology. This problem may be partially ameliorated by substituting a function, $g(x, y, t)$, that is temporally low-pass but spatially band-pass. Such a change might affect other aspects of model behavior and we have not, as of yet, systematically examined this issue. Third, Gaussian derivative receptive fields are non-causal (i.e., the response depends both on past and future stimulus intensities). This problem can easily be alleviated by introducing a time delay, but a more appropriate solution is to use recursive (i.e., feedback-based) temporal derivative filters (e.g., Adelson & Bergen, 1985; Fleet & Langley, 1995).

5.2 MT Linear Receptive Field Weights

Each MT neuron linearly combines a set of V1 afferents such that the MT response is velocity-selective. In particular, an MT neuron with preferred velocity $\vec{v}$ sums the responses of four V1 neurons at the following space-time orientations:

$$\hat{u}_1 = \begin{pmatrix} -v_x \\ -v_y \\ s^2 \end{pmatrix} / \sqrt{s^4 + s^2}, \qquad \hat{u}_2 = \begin{pmatrix} -v_y \\ v_x \\ 0 \end{pmatrix} / s,$$

$$\hat{u}_3 = [\hat{u}_1 + \hat{u}_2] / \sqrt{2}, \qquad \hat{u}_4 = [\hat{u}_1 - \hat{u}_2] / \sqrt{2},$$

where $s = |\vec{v}| = \sqrt{v_x^2 + v_y^2}$. These four space-time orientations are equally distributed around a plane in the spatio-temporal frequency domain, as illustrated in Fig. 3, and the sum of the underlying V1 receptive field power spectra forms a smooth annular ring. Since the fixed population of V1 neurons may not include these particular space-time orientations, the responses corresponding to the desired space-time orientations must be interpolated as described earlier. The interpolated V1 responses are linear sums of the fixed set of V1 responses, which can be combined together such that the full MT linear response is expressed as a weighted linear sum of the fixed set of 28 V1 neurons.

Finally, the MT neuron subtracts the responses of V1 neurons that do not lie near its preferred velocity plane. This is accomplished by subtracting the mean of the 28 weights from each of the weights (thus producing an overall set of weights with zero mean). These 28 zero-mean weights correspond to the final linear receptive field weights of the MT neuron. The resulting spatio-temporal frequency response function is smooth and depends only on the Euclidean distance to the plane. The weighting could be adjusted if warranted by physiological data (see *Proposed Experiments*). For the MT normalization pool, we used a population of 19 MT neurons that approximately tile the velocity space. One neuron is tuned for zero velocity, six for moderate speeds (16 deg/sec), and twelve for high speeds (120 deg/sec).