PERGAMON

Vision Research 42 (2002) 1035–1051

Vision Research

www.elsevier.com/locate/visres

A model of speed tuning in MT neurons

John A. Perrone [a,*], Alexander Thiele [b]

[a] Department of Psychology, University of Waikato, Private Bag 3105, Hamilton, New Zealand
[b] Department of Psychology, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

Received 31 May 2001; received in revised form 10 October 2001

Abstract

We have shown previously that neurons in the middle temporal (MT) area of primate cortex have inseparable spatiotemporal receptive fields—their response profiles exhibit a ridge that is oriented in the spatiotemporal frequency domain, and this orientation predicts the neurons' preferred speed. When measured in spatiotemporal frequency space, such MT spectral receptive field (SRF) properties are closely matched to the spectrum generated by a moving edge. In contrast, V1 neurons have SRF properties that are poorly matched to moving edge spectra, indicating that V1 neurons are not tuned to a particular image speed but rather to specific spatial and temporal frequencies. Here we describe a neural mechanism based directly on the properties of V1 neurons that is able to explain the SRF change that occurs between V1 and MT. We outline the theory behind this transformation and posit an explanation for how the visual system extracts true speed (independent of spatial frequency) from retinal image motion. We tested this speed model against our MT neuron data and found that it provides an excellent account of speed tuning in MT.

Keywords: Motion; Speed; Area MT; V1; Spatiotemporal; Temporal frequency; Spatial frequency

1. Introduction

As mobile animals, we are constantly exposed to a changing pattern of light falling onto the retinas of our eyes. We can learn a lot about the environment around us from these moving, fluctuating light patterns and they often provide critical information for our survival (Gibson, 1950; Koenderink & van Doorn, 1975; Longuet-Higgins & Prazdny, 1980; Nakayama, 1984). It is therefore not surprising that much research has been directed at understanding the properties and function of neurons located along the motion pathways of the brain. Many of these properties have been revealed over the past three decades but an understanding of how and where in the brain the speed of moving objects is extracted has been elusive until very recently (Perrone & Thiele, 2001). In a plot of spatial frequency versus temporal frequency, an object or edge moving at speed v has a spectrum that lies on a line of slope $-v$ (Fahle & Poggio, 1981; Watson & Ahumada, 1983) (see shaded line in Fig. 1a). A change in the speed of the edge alters the slope of the edge spectrum with a faster speed producing a steeper slope. The degree of orientation of the spectrum is therefore the critical spectral feature for any mechanism sensitive to the speed of the moving edge.

Moving edges are a common feature in our visual environment and the receptive fields of neurons in primary visual cortex (V1) and the middle temporal area (MT) are continuously exposed to them. By measuring a neuron's response to moving sine-wave gratings of different spatial and temporal frequencies, it is possible to map out the 'spectral receptive field' (SRF) of the neuron. The SRF gives an indication as to which spatial and temporal frequencies will cause the neuron to respond and it corresponds to the Fourier transform of the neuron's spatiotemporal receptive field. We have recently shown that many MT neurons have SRFs with ridges of peak sensitivity that are elongated and oriented relative to the spatial and temporal frequency axes, i.e., they are inseparable (Perrone & Thiele, 2001). Fig. 1b shows an example of the SRF of one of the MT neurons in our sample, plotted in contour plot form. This SRF map corresponds to the upper right quadrant of the spatiotemporal frequency space plotted in Fig. 1a. The SRF of this neuron is nicely matched to the spectrum of an edge moving at a particular speed (~3°/s); a faster or

* Corresponding author. Tel.: +64-7-838-4466x8292; fax: +64-7-856-2158.
E-mail address: jpnz@waikato.ac.nz (J.A. Perrone).

0042-6989/02/$ - see front matter
PII: S0042-6989(02)00029-9

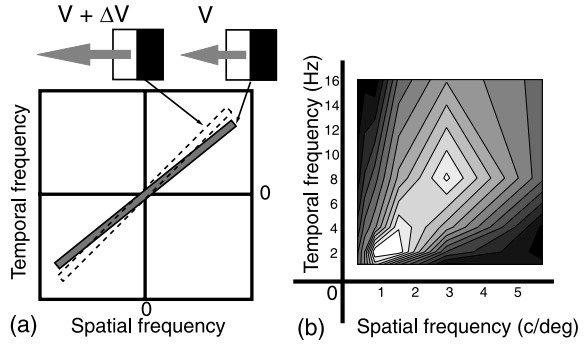

Fig. 1. Spatiotemporal frequency plots. (a) Representation of moving edge spectra for two edges, one moving right to left at speed v (shaded outline) and the other at $v + \Delta v$ (dashed outline). Discrimination of different edge speeds requires a mechanism that can respond selectively to a particular slope of the edge spectrum. (b) SRF of an MT neuron tested with 30 different combinations of spatial and temporal frequency (Perrone & Thiele, 2001). The neuron is tuned for a particular edge spectrum slope and a speed of approximately 3°/s. Only the upper right quadrant of frequency space is shown in this contour plot.

slower speed will result in poor alignment between the edge spectrum and the SRF and consequently a weaker response from the neuron.

A large proportion of the MT neurons in our sample had SRFs of this form and a wide range of preferred edge speeds were found across the sample (Perrone & Thiele, 2001). The oriented MT SRFs can account for the speed tuning data found from tests using moving edges and bars (Felleman & Kaas, 1984; Lagae, Raiguel, & Orban, 1993; Maunsell & Van Essen, 1983; Rodman & Albright, 1987) and confirm that these neurons are truly 'speed tuned' in that they tend to respond equally well over a range of spatiotemporal frequency combinations, provided the image speed is kept constant. Because moving edges (with oriented spectra) are a very common feature of our visual environment, it is not really too surprising to discover neurons with oriented (inseparable) SRFs at some stage of the visual motion pathway. There is also psychophysical evidence for velocity-tuned (inseparable) mechanisms in humans (Reisbeck & Gegenfurtner, 1999). However there remains a lot of uncertainty as to how MT neuron speed tuning properties are generated from the V1 neurons preceding them in the chain of neural motion processing. This paper addresses the issue of how neural information at the level of V1 might be modified before it reaches MT, an area specialized for motion processing and whose neurons receive a direct projection from V1 (Movshon & Newsome, 1996). We start by describing the properties of V1 neurons in the same frequency (Fourier) domain used above for the edge spectra and MT neurons (see Fig. 1). From these V1 properties, we construct a model of speed tuning within MT and then compare the model and MT neuron responses. Some of the material in this article has been reported previously (Perrone, 1998; Perrone & Thiele, 2000).

2. The model

2.1. Spectral receptive fields of V1 neurons

In the following section we will give a brief overview of V1 neuron SRFs and their construction, followed by a description of how these can be used to generate 'speed-tuned mechanisms'. The V1 neuron SRFs are made up from separate temporal and spatial frequency contrast sensitivity tuning functions as described in the following section.

2.2. V1 temporal frequency contrast sensitivity tuning

The temporal frequency response of V1 neurons can be measured by moving sine-wave grating patterns across the neuron's receptive field. The grating spatial frequency and direction are initially optimized for a given cell and then the temporal frequency is varied by changing the grating speed (Foster, Gaska, Nagler, & Pollen, 1985; Hawken, Shapley, & Grosof, 1996; Movshon, Thompson, & Tolhurst, 1978; Tolhurst & Movshon, 1975). Typical data from such an experiment are shown in Fig. 2a and they give an indication of the amplitude response of the two types of neuron when exposed to a range of temporal frequencies. The fitted curves are based on the temporal frequency tuning functions used by Watson (1986) and Watson and Ahumada (1985) in their motion model (see Appendix A).

In primates, the picture that has emerged is that some V1 neurons (both simple and complex) have temporal frequency response profiles that are low-pass while others tend to be band-pass (Foster et al., 1985; Hawken et al., 1996) (see Fig. 2a). The former type respond best to static patterns (temporal frequency = 0 Hz) whereas the latter prefer moving features. Human psychophysical studies have also revealed low-pass and band-pass temporal frequency tuning (Robson, 1966) and the terms 'sustained' and 'transient' have been used to refer to the two different putative temporal channels (Cleland, Dubin, & Levick, 1971; Kulikowski & Tolhurst, 1973). When applied to neurons, 'sustained' indicates a response that extends for the duration of the stimulus, whereas 'transient' indicates a response primarily at stimulus onset and offset.

2.3. Spatial frequency contrast sensitivity tuning

Hawken and Parker (1987) carried out one of the most detailed investigations of the spatial frequency responses of V1 neurons in primates and specifically examined the shape of the spatial contrast sensitivity function by testing the neurons with a wide range of spatial frequencies. The data from one of their neurons are shown in Fig. 2b. This is representative of the cells in their sample but they also discovered many unusual curves as

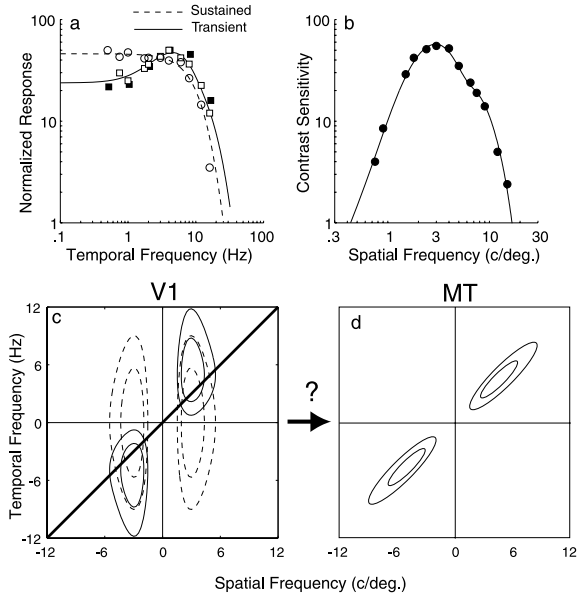

Fig. 2. Comparing V1 and MT neurons' SRFs. (a) Temporal frequency response data from V1 neurons. Vertical axis indicates normalized responses for data and contrast sensitivity for curves. The open circles and squares are from Foster et al. (1985, Fig. 8b,c). The solid squares are data supplied by Hawken (personal communication). Both sets of data have been normalized relative to the peak response and scaled by a factor of 50. One neuron type (sustained) is low-pass in shape (open circles and dashed curve), the other (transient) is band-pass. The fitted curves are amplitude response functions used to model the temporal frequency contrast sensitivity tuning of the sustained and transient neurons (see Appendix A). (b) Re-plotted spatial frequency contrast sensitivity data from Hawken and Parker (1987, Fig. 6a). The fitted curve is from the amplitude response function they used to model the spatial frequency tuning of their data and which we use in our model (see Appendix A). (c) Plan view of the SRFs of the two types of V1 neurons. Note that this plot uses linear axes. The contour lines connect values at 50% and 75% of the peak. The solid diagonal line represents the spectra of a moving edge. (d) Stylized version of typical MT neuron SRF (see Fig. 1b). Our model is designed to explain how the V1 non-oriented (one quadrant separable) SRFs are converted into the oriented (inseparable) form evident in MT.

well (see their Fig. 13). This data set happens to be from a simple cell (their Fig. 6a) but we will use it as a generic V1 complex neuron spatial contrast sensitivity function. This approach can be justified by the fact that Hawken and Parker found no differences between the shapes of the spatial contrast sensitivity functions from the simple and complex cells in their sample. The fitted curve is based on the difference of difference of Gaussians with separation (d-DOG-s) function used by Hawken and Parker to model their data and which we also use in our MT speed tuning model (see Appendix A).

2.4. Combination of temporal and spatial contrast sensitivity tuning functions

When the spatial function (Fig. 2b) is combined with the transient and sustained temporal contrast sensitivity functions (Fig. 2a), the three-dimensional (3-D) non-oriented SRFs of typical V1 neurons are obtained. Fig. 2c includes two plan views (i.e., looking down the contrast sensitivity axis) of two 3-D spectra (transient and sustained) formed by combining the functions in Fig. 2a and b. Within a single quadrant of frequency space, the SRF is generated by multiplying the spatial and temporal functions. The sustained SRF is generated by combining the spatial function in Fig. 2b with the dashed temporal function in Fig. 2a; the transient SRF is constructed from the Fig. 2b spatial function and the solid curve in Fig. 2a.

The fact that the temporal functions are combined with the spatial functions using a multiplication operation assumes that the temporal function does not change shape as the spatial frequency changes and vice versa (i.e., it assumes separability within a particular quadrant of frequency space). There is physiological evidence to support this assumption (Foster et al., 1985; Tolhurst & Movshon, 1975). The above combination of spatial and temporal functions also involves a 'mixing' of data based on neuron responses and data from contrast sensitivity (threshold) experiments. However there is evidence that, for cells in cat area 17 at least, the two different measures (firing rate and thresholds) are compatible and closely related (Movshon et al., 1978). We will work on the basis that the fitted curves represent the amplitude response functions of the neurons and reflect the relative output of the neurons in response to different spatial and temporal frequencies.

The spatial functions for positive and negative frequencies are assumed to be equivalent but the temporal functions for the transient neurons are asymmetric with the peak sensitivity falling in two diagonally opposite quadrants, dependent upon the directional tuning of the neuron (Watson & Ahumada, 1983). The sustained neuron's SRFs are depicted in Fig. 2c as being symmetrical along both dimensions (four quadrant separable) and so they would respond about equally to opposite directions of motion (non-directional response). However, this need not be the case and it is not critical to the functioning of the model. In this paper we are concerned mainly with the response properties of the neurons in just two diagonally opposite quadrants of frequency space, corresponding to the most responsive or preferred direction of the two neurons.

A stylized version of a moving edge spectrum has been included in the Fig. 2c plot as well. Notice that—unlike the moving edge spectrum—the SRFs of the sustained and transient V1 neurons have principle axes that are parallel to the spatial and temporal frequency axes. Individually, therefore, they lack the orientation required to respond selectively to the oriented spectra generated by moving edges. It follows therefore that they are going to be poor at differentiating between edge spectra arising from slightly different edge speeds (orientations). Yet,

somehow by the time we reach MT, the SRFs have acquired the orientation required to selectively respond to particular edge speeds (Fig. 2d). Indeed a long-standing problem in the field of visual perception is how the visual system determines local image speed, given the limited properties of V1 neural SRFs.

One of the earliest suggestions for how humans and primates could determine the speed of moving features was a mechanism based on the ratio of the outputs from a 'fast' mechanism similar to the transient type of V1 neuron (T) and from a 'slow' mechanism similar to the sustained V1 type (S). A number of theorists argued that the T/S ratio is linearly related to the temporal frequency of the moving stimulus and hence its velocity can be determined from the output of a unit that divides the T mechanism output by the S mechanism output (Adelson & Bergen, 1986; Harris, 1986; Thompson, 1982; Tolhurst, Sharpe, & Hart, 1973). The principle behind these ratio models can be illustrated using the V1 temporal frequency data shown in Fig. 2a. The temporal frequency curves have been re-plotted in Fig. 3a. Since the transient (T) and sustained (S) V1 functions are plotted on a log vertical axis, the difference between them ($\log T - \log S$) is equivalent to $\log(T/S)$. This is plotted in Fig. 3b. One can see that an increase in the value of the T/S ratio is associated with an increase in the temporal frequency (i.e., the speed) of the stimulus. The problem with this scheme is that it does not produce a *speed-tuned* unit; the output goes on increasing as long as the stimulus speed increases. It therefore cannot be a good account of what is happening in MT neurons.

2.5. Temporal frequency tuning

A mechanism that does produce temporal frequency tuning from the two V1 temporal functions (T and S) is depicted in the remaining panels of Fig. 3 (c, d, solid lines). Consider the introduction of an absolute operation after the $\log T - \log S$ stage—perhaps through the use of two divergent inputs with one responding when $\log T > \log S$ and another responding when $\log S > \log T$ (Fig. 3c). This results in an inverted function with the minimum at a temporal frequency corresponding to the point at which the two temporal frequency functions intersect (see solid arrow in Fig. 3a). An inversion operation applied to Fig. 3c then produces a peaked function tuned to a particular temporal frequency (Fig. 3d). As long as one temporal frequency tuning curve is low-pass and the other band-pass, the two functions will intersect at some point. The use of this 'cross-over' point has been previously identified as a potential method for coding image speed (Perrone, 1994; Thompson, 1983). The additional absolute and inversion operations enable the two broadly tuned transient and sustained temporal filters to be converted into a narrowly tuned unit. We make use of this feature in our MT speed tuning model.

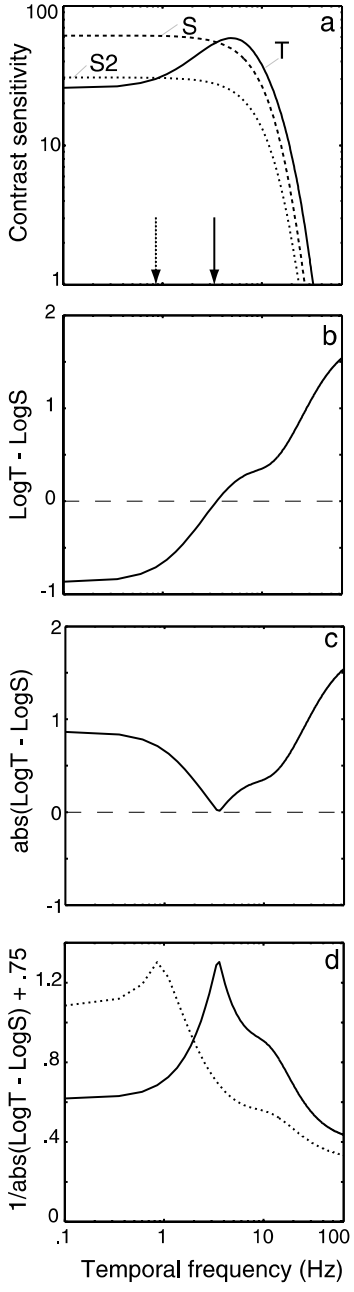

Fig. 3. Generating temporal frequency tuning. (a) Temporal frequency contrast sensitivity functions for sustained (S) and transient (T) V1 neurons. Because one is low-pass and the other is band-pass, the two curves intersect at a particular temporal frequency (solid arrow). The dotted line ($S2$) is for another sustained V1 neuron whose weighted output is reduced relative to S and has lower overall sensitivity. It intersects the T curve at a lower temporal frequency (dotted arrow). (b) Plot showing the relationship between the temporal frequency of the stimulus and $\log T - \log S$ or equivalently, $\log(T/S)$. (c) Temporal frequency as a function of $\text{abs}(\log T - \log S)$. The function has a minimum at the point at which the S and T functions intersect. (d) Temporal frequency as a function of $1/(\text{abs}(\log T - \log S) + 0.75)$. The function now peaks at the intersection of the S and T curves. This mechanism (based on the two V1 neurons' temporal properties) is tuned to a particular temporal frequency (~3 Hz). The 0.75 term is a constant that controls the width of the tuning curve. The dotted curve is for the case where the weighted version of S ($S2$) is used in the mechanism instead of S. This mechanism is tuned to a different temporal frequency (~1 Hz).

A second key idea behind the model is illustrated by the second sustained function shown in Fig. 3a (dotted line marked as $S2$). The relative sensitivity of this V1 neuron is reduced relative to that of the transient neuron. Therefore the two functions (T and $S2$) now intersect at a different temporal frequency (see dashed arrow). Application of the same absolute and inversion operations now results in a tuning curve that peaks at a lower temporal frequency than the case where S was used instead of $S2$ (see dotted curve in Fig. 3d). Therefore, by weighting the output of the sustained V1 neuron relative to that of the transient neuron (or vice versa), it is possible to control where the peak temporal frequency response of the new mechanism occurs. We also incorporate this feature into our MT speed tuning model.

Unfortunately the mechanisms described above are limited to the temporal frequency dimension only (i.e., they are not speed tuned) and they do not directly address the important question as to how the oriented SRFs in MT are generated. To do this we need to look at both the spatial and temporal frequency tuning of the T and S V1 neurons.

2.6. Combining sustained and transient V1 neuron properties

In Fig. 3d above we showed that it is possible to produce a mechanism that is maximally sensitive whenever the T and S V1 outputs are equal. We now extend this concept to two dimensions (u, ω). Fig. 4a shows a perspective plot of the spatiotemporal frequency contrast sensitivity function for a sustained V1 neuron. It is based on the functions fitted to the spatial and temporal data shown in Fig. 2a and b and represents a 3-D view of the upper right quadrant of Fig. 2c (but note that it uses log axes instead of the linear axes of Fig. 2c). Fig. 4b shows the same type of plot for the transient type of V1 neuron. We are attempting to construct a mechanism tuned to a particular speed v, i.e., to all spatial–temporal frequency combinations (u, ω) such that $v = \omega/u$.

The required mechanism turns out to be very simple. The two V1 contrast sensitivity functions (T and S) can be made to overlap along the $v = \omega/u$ line by modifying the spatial frequency tuning of the transient neuron relative to the sustained neuron (see Appendix A and Fig. 4c). In the earlier Fig. 2c, both the sustained and transient spatial frequency tuning functions were assumed to be identical. However a key component of our model is that they actually *differ slightly* (see back wall of Fig. 4c). The amount they differ from each other can be calculated exactly from the T and S temporal frequency tuning functions (see Eq. (A.1) in Appendix A).

In three dimensions, this manipulation means that the locus of intersection of the two SRF surfaces can be made to trace out a curve that is oriented with respect to the spatial and temporal frequency axes (see Fig. 4c for the case where $v = 1°/s$). The result is that the two V1 neuron contrast sensitivity functions, $S(u, \omega)$ and $T(u, \omega)$, overlap on the diagonal (as in Fig. 4c) and will have equal outputs whenever the temporal and spatial frequency of the stimulus are related by the equation $\omega = vu$, i.e., they will respond equally to all speeds $v = \omega/u$ (see black line at bottom of Fig. 4c).

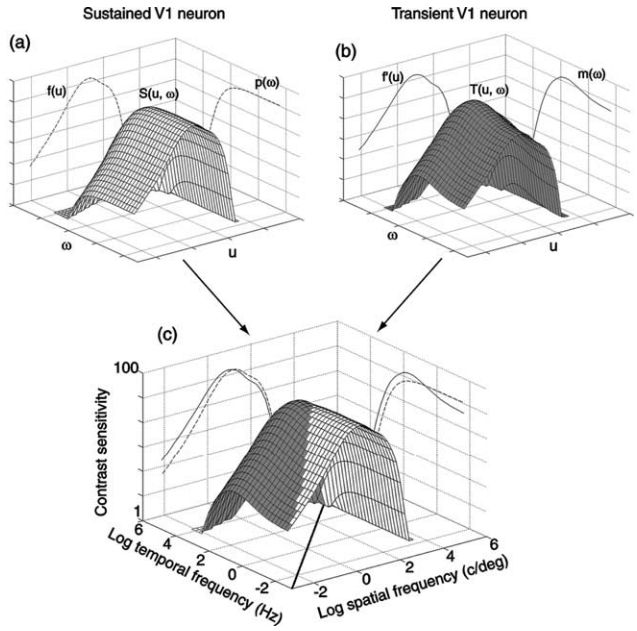

Fig. 4. Perspective plots of V1 sustained and transient SRFs. For clarity, only the upper right quadrant of frequency space is shown. (a) Combined contrast sensitivity of a sustained type V1 neuron with low-pass temporal tuning $p(\omega)$ and spatial frequency tuning $f(u)$. (b) Contrast sensitivity function for transient type V1 neuron with band-pass temporal frequency tuning $m(\omega)$ and spatial frequency tuning $f'(u)$. (c) Both sustained (white) and transient (shaded) contrast sensitivity shown on the same plot. If the spatial frequency tuning of the transient neuron differs in a special way from that of the sustained neuron (see back wall of plot and Eq. (A.1) in Appendix A), the two curves intersect at spatial and temporal frequencies (u_i, ω_i) that fall along a straight line in (u, ω) space (see light–dark border and line on floor of plot). This combination of spatial and temporal frequencies correspond to a stimulus speed $v = \omega_i/u_i$.

2.7. Speed-tuned (oriented) SRFs

We next introduce a mechanism that responds maximally to the spatial and temporal frequencies that lie along the S–T intersection. There are a number of techniques that could be adopted to achieve the requirements of this new mechanism, but the one selected uses mainly subtraction and addition operations in order to keep it physiologically plausible. If T and S represent the transient and sustained neurons' contrast sensitivities, respectively (i.e., the combined spatial and temporal contrast sensitivities, as shown in Fig. 4c), the

contrast sensitivity of the new model speed-tuned mechanism is given by

$$M(u,\omega) = \frac{\log(T+S+\alpha)}{|\log T - \log S| + \delta}. \quad (1)$$

This differs slightly from the basic mechanism outlined in Fig. 3. The sum $(T+S)$ has been added to the numerator and is there to ensure that the response is maximal when both neuron types are responding significantly. It is log transformed in order to broaden the response of the speed-tuned mechanism so that it responds across a wider range of spatial frequencies. The α term is a constant and it also helps broaden the response profile of the mechanism along the direction of the SRF ridge. It is analogous to the background or spontaneous activity of the neurons (but note that the units and scale are arbitrary). The last term (δ) in the denominator is a constant to prevent division by zero and which makes the output less sensitive to noise. This parameter controls the width of the peak sensitivity ridge and it is used to set the bandwidth of the new speed-tuned mechanism.

2.8. Weighted intersection mechanism model (WIM)

The operation of the proposed model can be broken down into two simple principles:

Principle 1. The maximum output from the new speed-tuned mechanism occurs whenever the outputs of the two neuron types (transient and sustained) are equal.

Principle 2. The peak response of the speed-tuned mechanism is maximal only for specific edge speeds, i.e., for spatial and temporal frequency combinations that lie along an oriented line ($\omega = vu$) in frequency space.

The two principles which set up the oriented spectrum and which generate the maximum response when the transient and sustained outputs are equal will be referred to collectively as the WIM model. The decision to implement this particular algorithm (Eq. (1)) was based on the excellent fit it provides between the model and existing literature on MT neuron speed tuning (see below).

Fig. 5a and b show re-plotted speed tuning data from Lagae et al. (1993) and Maunsell and Van Essen (1983). Both sets of data are from a large number of MT neurons tested with a broad range of speeds. These studies were carried out using moving bars and edges to test the neurons and therefore the stimuli are very broad-band in terms of the range of spatial frequencies presented. Thus the speed tuning curves should mainly reflect the temporal frequency tuning of the neurons. Fig. 5c is the WIM model output when the spatial frequency tuning is fixed at some value, u, and only the temporal frequency tuning of the V1 neurons is considered. Using temporal frequency curves similar to those shown in Fig. 2a and combining the T and S neuron outputs using the algorithm described by Eq. (1) results in a 'speed tuning' curve with characteristics that closely match the MT data. In both sets of data (Fig. 5a and b), a sharp peak is clearly obvious and concave regions on either side of the peaks are also evident. The same characteristics are present in the WIM model output.

We tried a number of alternative forms of Eq. (1) (with and without the log terms, without the numerator, etc.) but none were able to provide as good a fit to the speed tuning data as that shown in Fig. 5. We recognize that some of the proposed operations lack a direct physiological counterpart and that these computations will need to be specified in more detail in the future. The current implementation has the advantage that the transient and sustained neurons' outputs can be simulated using spatial and temporal contrast sensitivity functions based on actual V1 neuron data. However this form of the model does not address factors such as the actual response in time of the transient and sustained V1 neurons. This aspect is currently being examined (Perrone & Krauzlis, 2002). A full version of the model using image-based spatiotemporal filters (e.g., Watson & Ahumada, 1985) will also enable us to examine the effect of factors such as

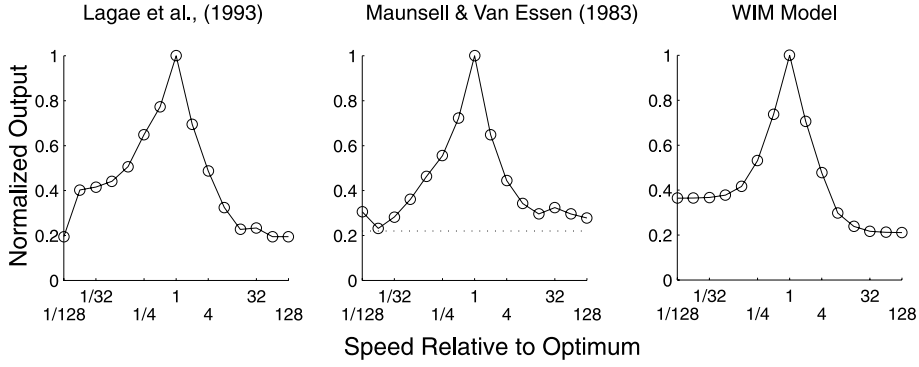

Fig. 5. Comparing the WIM model output to existing speed tuning data obtained with broad-band stimuli (moving edges and bars). Because of the broad-band nature of the input, the model output could be simulated using just the temporal frequency tuning of the sustained and transient V1 input neurons (see Fig. 2a).

stimulus contrast on the speed extraction process (Stone & Thompson, 1992; Thompson, 1982).

2.9. WIM model implementation

The main hypothesis underlying the WIM model is that the temporal and spatial frequency contrast sensitivity tuning functions of the sustained and transient V1 neurons have evolved in order to enable a special interaction that produces speed tuning in directionally selective units. In order to demonstrate how this process could work, we present the results of a simulation (see Fig. 6) in which the two neurons (sustained and transient) start out as being equivalent in their spatial and temporal frequency contrast sensitivity tuning and then the transient neuron gradually alters its properties. This is not intended to imply a particular evolutionary or developmental order but to illustrate the steps in the WIM model algorithm.

In Fig. 6a the temporal frequency contrast sensitivity tuning of the sustained neuron is simulated using the function and parameters obtained from the V1 data shown in Fig. 1a (see Appendix A). The transient neuron's temporal frequency tuning is initially set to be identical to that of the sustained neuron (hence the functions are superimposed). In Fig. 6b the sustained neuron's spatial frequency contrast sensitivity tuning has been simulated using the equation and parameters taken from the Hawken and Parker (1987) paper (see their Fig. 6a neuron, re-plotted in our Fig. 1b). The desired speed tuning of the model mechanism was arbitrarily set at 1°/s. The transient neuron's spatial frequency tuning was then calculated using Eq. (A.1) (Appendix A). Since the sustained and transient temporal frequency functions are identical, the value of R in Eq. (A.1) is equal to 1.0 for all spatial frequencies and so the calculated transient spatial function is identical to that of the sustained function (the functions are superimposed in Fig. 6b). The respective temporal and spatial functions in Fig. 6a and b were then multiplied together to produce the simulated sustained and transient neurons' contrast sensitivities (S and T). The model speed-tuned mechanism SRF was then calculated from the sustained and transient contrast sensitivities using Eq. (1). The α parameter in the model was set at 0.0 for all of the Fig. 6 simulations. The result is shown as a contour plot in Fig. 6c (note that only the upper right quadrant of frequency space is depicted). The SRF of the model mechanism exhibits no orientation under these conditions.

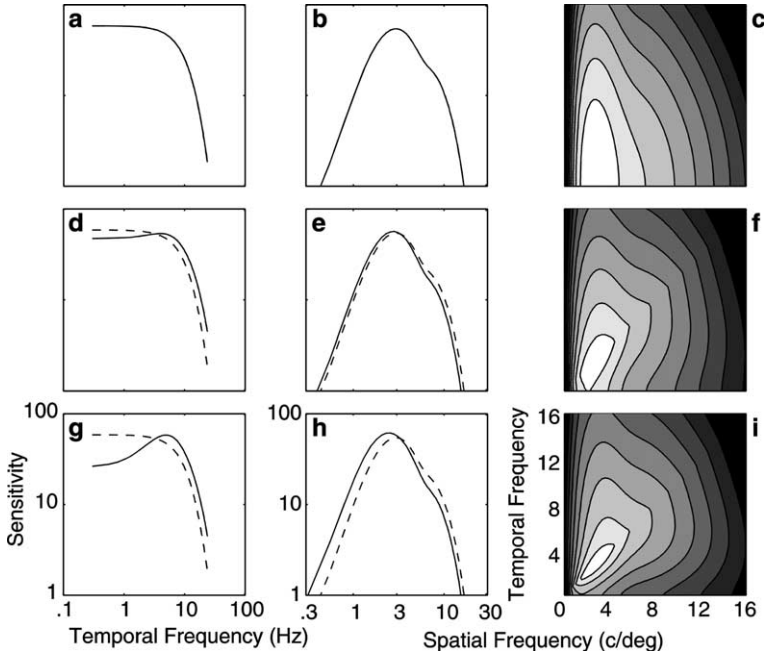

Fig. 6. Generating an oriented SRF using the WIM model. Spatial frequencies in the range 0.3–24 cycles/deg were sampled in steps of 0.25 cycles/deg. The speed tuning of the mechanism was set at 1°/s, and the temporal frequencies tested ranged from 0.3 to 24 Hz in 0.25 Hz steps. The SRF of the model mechanism (right hand panels) gradually acquires orientation relative to the spatial and temporal frequency axes as the transient neuron temporal frequency tuning (solid lines in left hand panels) changes from low-pass to band-pass. Left hand panels (a, d, g): The transience factor (ζ) controlling the band-pass extent of the temporal frequency function (see Appendix A) is set to 0.0, 0.2, and 0.6, respectively, in the three different panels. Middle panels (b, e, h): Spatial frequency contrast sensitivity functions for both the sustained (dashed lines) and transient (solid lines) V1 neurons. The sustained function is from Hawken and Parker (1987, Fig. 6a), and the transient theoretical function is calculated using Eq. (A.1) (Appendix A). Right hand panels (c, f, i): Upper right quadrant of spatiotemporal frequency space showing contour plots of the SRFs produced by the WIM model using the temporal and spatial functions shown to the left of the plots. The four innermost contours are drawn at 90%, 80%, 70% and 60% of the maximum sensitivity.

In Fig. 6d, the transient neuron's temporal frequency tuning (solid line) is no longer assumed to be identical to that of the sustained neuron but is now slightly band-pass. The temporal function described in Appendix A was used to model the transient neuron's temporal frequency tuning and the transience parameter (ζ) that controls the degree of band-pass tuning was set to 0.2. The other parameters in the function were matched to those of the transient curve in Fig. 1a. With the sustained and transient temporal contrast sensitivity tuning functions set as in Fig. 6d, the transient neuron's spatial frequency tuning was again calculated using Eq. (A.1) (Appendix A). This time, since the sustained and transient temporal curves are different, the transient spatial frequency function differs slightly from the sustained function (see Fig. 6e). This is a consequence of the Principle 2 of the model mechanism, and it forces the intersection of the two neurons' SRFs (transient and sustained) to lie on a straight line in frequency space. The SRF of the model mechanism was then generated using Eq. (1) from the functions shown in Fig. 6d and e and is plotted in Fig. 6f. The SRF has acquired a limited amount of orientation.

As the temporal frequency tuning of the transient neuron becomes even more band-pass (Fig. 6g) the transient spatial frequency tuning curve changes more relative to the sustained tuning (Fig. 6h) and the model mechanism SRF generated using Eq. (1) (Fig. 6i) has a region of peak sensitivity that is clearly oriented relative to the spatial and temporal frequency axis. A combined neural mechanism with an oriented SRF has been created out of two non-oriented spatiotemporal units with properties similar to those of neurons commonly found in area V1 of primate visual cortex.

Fig. 7 shows the effect of the other parameters in the model upon the shape of the SRF. In Fig. 7a, the preferred speed tuning (v) of the WIM sensor has been increased to 2°/s (cf. Fig. 6i at 1°/s). In Fig. 7b the effect of changing the δ parameter in Eq. (1) is demonstrated. Decreasing δ from 1.25 (as per Fig. 7a) to 0.7 reduces the width of the region of peak sensitivity along the temporal frequency dimension. Fig. 7c demonstrates the effect of the α parameter in Eq. (1). In the first two panels, α is set to 0.0, but in Fig. 7c it is set at 100.0. A larger value of α increases the length of the peak sensitivity ridge and broadens the range of spatial and temporal frequencies to which the sensor responds. However it also increases the 'background' activity of the sensor.

We have demonstrated that it is possible to construct a speed-tuned motion sensor with the requisite oriented SRF properties from just two spatiotemporal filters, one with properties commonly found in V1 'sustained' neurons and another with 'transient' V1 properties. We next demonstrate that the SRFs generated by the WIM model are able to closely emulate those found in MT neurons.

2.10. Simulating the SRF properties of MT neurons with the WIM model

We used the WIM model to fit the rhesus monkey MT neuron data we have previously reported elsewhere (Perrone & Thiele, 2001). In this electrophysiological study, each MT neuron (out of a total of 84) was tested with sine-wave gratings moving in the neuron's preferred direction. The test gratings were made up from six spatial (0.2, 0.4, 0.7, 1.4, 2.8, 5.6 cycles/deg) and five temporal (1, 2, 4, 8, 16 Hz) frequencies. The contour plot shown in Fig. 1b is an example of the response map generated by one of our MT neurons to these 30 different spatial and temporal frequencies. For each set of neuron data, we optimised five parameters in the model to generate the best fit (in a minimum least squares sense) between the model and data. The five parameters were: (1) the peak spatial frequency ($u0$) of the sustained V1 neuron; (2) the degree of band-pass temporal frequency tuning of the transient V1 neuron (ζ, see Appendix A); (3) the speed tuning of the model sensor (v);

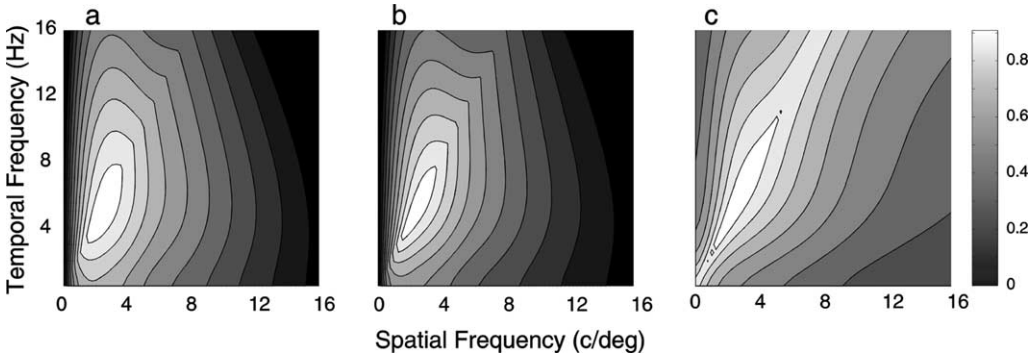

Fig. 7. Contour plots showing the effect of different parameters in the WIM model (Eq. (1)) upon the shape of the SRF. The effect of the changes can be seen by comparing the plots to Fig. 6i. (a) The speed tuning of the mechanism was increased from 1°/s to 2°/s. (b) The δ parameter in Eq. (1) was decreased from 1.25 to 0.7. (c) The α parameter in Eq. (1) was increased from 0.0 to 100.

(4) the α parameter in Eq. (1); (5) the δ parameter in Eq. (1). Although the model can accommodate a larger number of parameters (e.g., the shape of the spatial frequency contrast sensitivity functions), we were interested in discovering the minimum configuration that could adequately account for our MT data.

3. Results

3.1. MT SRF simulations

The left hand column of Fig. 8 shows, in contour plot form, the SRFs of four representative MT neurons in

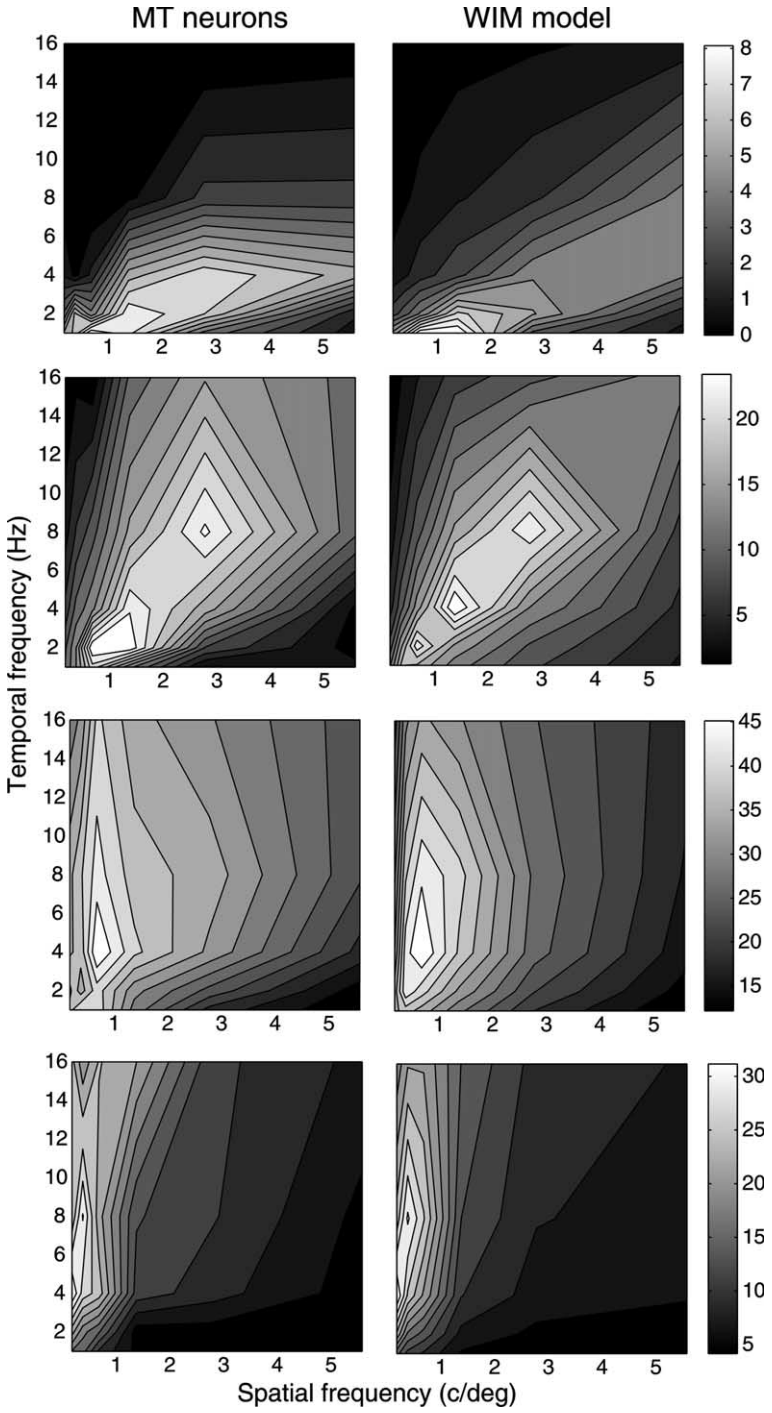

Fig. 8. Comparing the WIM model SRFs with MT neuron data. Left hand column: SRF data from four representative MT neurons (Perrone & Thiele, 2001). The bars to the right of the figure indicate the average response in impulses/s. Right hand column: Best fitting WIM model SRFs when five parameters of the model are allowed to vary (see Appendix A). The model SRFs were sampled at the same 30 spatial and temporal frequencies used to test the MT neurons.

our sample ($N = 84$). The right hand column shows the best fitting SRFs generated using the WIM model. The discrete log sampling and limited range of frequencies used to test the spatiotemporal frequency tuning of the neurons and the model means that the contour plots often lack the smooth peak sensitivity regions seen in the continuous plots of Figs. 6 and 7. However, in all cases the fit between the MT data and the model is excellent ($r = 0.94$, 0.91, 0.91, 0.96 for Fig. 8a, b, c, d, respectively).

We carried out a similar analysis over all of the neurons in our sample. The range of r values was 0.44–0.99 (mean = 0.83, S.D. = 0.12). A large proportion of our MT data was very closely fit using the WIM model. In order to provide a baseline against which the WIM model fits can be compared, we also calculated the degree of fit to the MT data using a non-oriented 2-D Gaussian function (see Perrone & Thiele, 2001). This function uses five parameters and so is directly comparable to the WIM model fits. For the Gaussian, the mean r value across the 84 neurons was 0.80 (S.D. = 0.13). The distributions of r values for the two types of fit are shown in Fig. 9. The WIM model provides a good fit to the MT data relative to the base-line non-oriented 2-D Gaussian function. The two distributions can be compared by applying a Fisher Z transform (Hays, 1963) to the two sets of r values and carrying out a repeated measures t-test. The hypothesis that the WIM model and Gaussian functions provided an identical fit to the MT data can be rejected ($t = 3.71$, 83 df, $p < 0.0001$).

We initially expected only rough qualitative fits between the model and the MT SRF data. Earlier models of the V1–MT processing stage suggest the involvement of many more neural units in the analysis of image speed (e.g., Nowlan & Sejnowski, 1995; Simoncelli & Heeger, 1998). Admittedly these models tackle broader issues such as image segmentation and directional tuning as well, but their speed tuning properties still depend on mechanisms involving more than just two V1-like subunits (see also Simoncelli & Heeger, 2001; Thompson, 1984; Chey, Grossberg, & Mingolla, 1998). Therefore we anticipated that we would need to combine several WIM units across a range of spatial frequencies in order to get good fits to the MT data. However we have shown that the spatiotemporal frequency tuning characteristics of many of the MT neurons in our sample can be captured using an economical mechanism based on the inputs from just two V1-like neurons. We do not of course suggest that all MT neuronal speed tuning is based on just two V1 neurons and we also recognize the need for a greater number of MT V1 afferents in other contexts, e.g., direction estimation. The basic WIM model enables a large range of SRFs to be generated that encompass most of the types we have found in our MT sample (Perrone & Thiele, 2001). The MT data provide the most direct evidence that the visual system may have adopted a scheme similar to the WIM model at some stage between V1 and area MT.

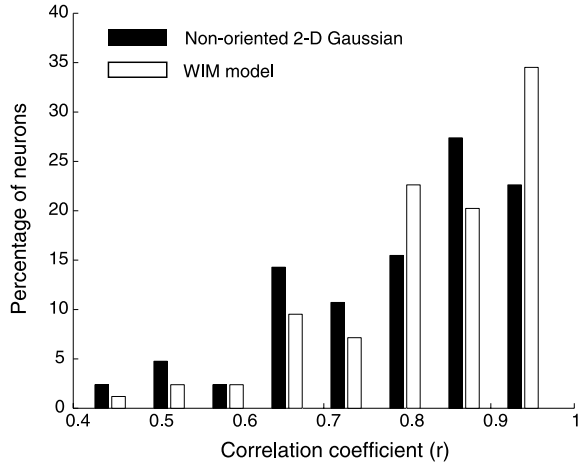

Fig. 9. Distribution of the correlation coefficient (r) for the WIM model–MT data fits (white bars). The black bars are for fits to the MT data using a non-oriented 2-D Gaussian (see Perrone & Thiele, 2001).

3.2. MT speed tuning tested with moving bars

As part of the exercise of studying the SRF properties of MT neurons (Perrone & Thiele, 2001) we also tested a subset of our neurons (48/84) with moving bars in order to compare the optimum speed tuning specified with the moving bar to the optimum tuning suggested by the orientation of the SRF. We were able to demonstrate that there was a positive relationship between the two measures. Rather than relying on an estimate of the orientation (based on a 2-D Gaussian fit), we now have at our disposal an estimate of the speed tuning of the MT neurons based on the WIM model fitting procedure. If the WIM model is providing a good description of the MT neuron data, the best fitting v parameter in the model should correspond to the optimum speed tuning of the neuron in response to a broad-band stimulus such as a moving bar. To test this hypothesis, we compared the value of v with the peak speed tuning (v') of the MT neurons that we also tested with moving bars. The value of v' was derived from the mean of the best fitting 1-D Gaussian fitted to the bar speed tuning data (see Perrone & Thiele, 2001).

Fig. 10 shows the result of a regression analysis of $\log v'$ against $\log v$ for 30 cells in our sample. These represent a subset of the total number of neurons we tested with moving bars ($n = 48$) and which produced WIM model fits of $\geqslant 0.8$ and 1-D Gaussian fits to the speed tuning data of $\geqslant 0.8$. The equation of the line is given by $y = 0.91x + 0.56$ ($r^2 = 0.32$, $F = 13.3, 1, 28$ df, $p < 0.001$). Overall, the WIM model provides a good 'prediction' of the speed tuning of the MT neurons in response to moving bars.

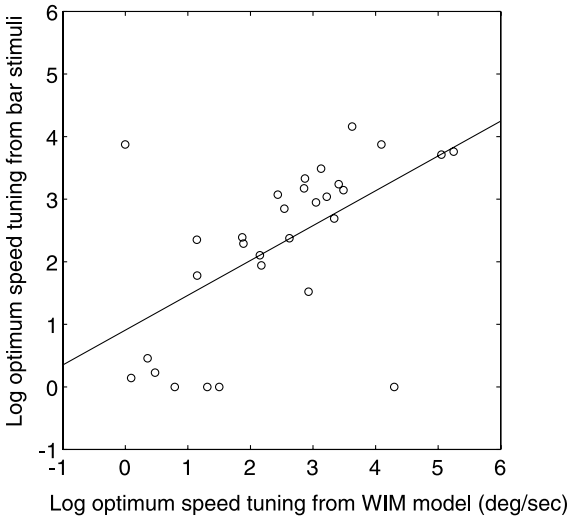

Fig. 10. Predicting the speed tuning of the MT neurons from the speed tuning of the best-fitting WIM model sensor. Plot shows scatter plot of log v against log v' where v is the speed tuning of the WIM model and v' is the speed tuning of the MT neuron derived using moving bars (see Perrone & Thiele, 2001). The data from the moving bar tests were fit with a 1-D Gaussian that could vary in its location (v'), its spread and base (pedestal). Data have only been included in the plot if the fit of the 1-D Gaussian to the speed tuning data was $\geqslant 0.8$ and if the WIM model–MT SRF fit also produced an $r \geqslant 0.8$. This resulted in the regression analysis being carried out on 30 of the 48 neurons tested with moving bars.

3.3. Summary of WIM model parameters

If we assume that a similar mechanism to the WIM model underlies the formation of the SRFs in our MT neuron sample, then we can infer some of the properties of the underlying V1 neurons feeding into each MT neuron.

3.3.1. Sustained and transient neuron spatial frequency tuning

One of the parameters in the model that we varied in order to optimize the fit between the model and MT neuron SRFs was the peak spatial frequency tuning of the sustained V1 neuron ($u0$). For our 84 MT neurons the mean value for the $u0$ parameter was 1.6 cycles/deg (S.D. = 1.8 cycles/deg, median = 0.82 cycles/deg). This is certainly in the range typically reported for macaque V1 (e.g., Foster et al., 1985; Hawken et al., 1988).

We were more concerned however with the spatial frequency properties of the transient V1 neuron required by the model to generate good fits to the MT data. The operation of the model rests on the assumption that the spatial frequency tuning of the transient V1 neuron is different in a very specific way from that of the sustained V1 neuron (see Figs. 4 and 6 above). We therefore determined the peak spatial frequency tuning of the transient V1 neuron in the WIM model that provided the best fit to each set of MT data. This was done by fitting an s-DOG-s function (see Appendix A) to the theoretical values generated from Eq. (A.1) (Appendix A). For our whole set of 84 MT neurons the mean peak spatial frequency tuning of the transient V1 neuron was 1.4 cycles/deg (S.D. = 1.5 cycles/deg, median = 0.8 cycles/deg). This supports the earlier observation (apparent in Figs. 4 and 6) that the transient V1 neuron needs to be tuned to slightly lower peak spatial frequencies than the sustained V1 neuron for the WIM mechanism to work. However it also confirms that the peak spatial frequency tuning of the model V1 transient units falls well within the bounds observed for V1 spatial frequency tuning (Foster et al., 1985; Hawken et al., 1988).

In order to illustrate the difference in the spatial properties of the sustained and transient V1 neurons required by the model, we have plotted space domain versions of the V1 spatial functions (sustained and transient) for one of our MT neuron fits (Fig. 11). For this particular neuron (st0tun3.3-104), the peak spatial frequency ($u0$) of the sustained V1 neuron input was very close to 3 cycles/deg and so the frequency domain plots corresponds closely to those shown in Fig. 6h. It should be apparent that the differences in the spatial receptive field profiles are not large and that the two classes of neurons in the model (sustained and transient) overlap considerably along the spatial frequency dimension.

The other important parameters in the d-DOG-s spatial frequency tuning function are the amplitudes (A_1, A_2, A_3) of the individual Gaussians making up the overall spatial function. These relate to the contrast sensitivity of the lateral geniculate neurons assumed to

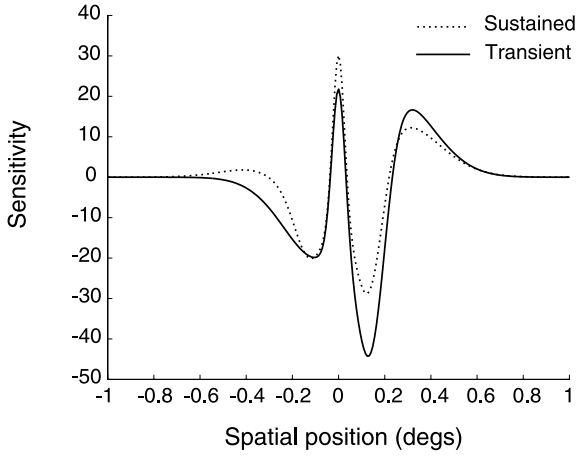

Fig. 11. Space domain plots of the V1 neuron spatial receptive fields used in the WIM model to fit one of our MT neurons' SRF data sets. In the frequency domain, the sustained neuron (dashed line) has a spatial frequency contrast sensitivity curve very similar to that shown in Fig. 2b. In the frequency domain, the transient V1 curve (solid line) looks very much like that shown in Fig. 6h. Compared to the sustained neuron, it is tuned to slightly lower spatial frequencies and has a higher peak sensitivity.

be subunits of the V1 receptive fields (see Hawken & Parker, 1987, and Appendix A). Across our population of MT neuron fits, the means (and standard deviations) for A_1, A_2, A_3 from the transient V1 neurons were: 37.4 (36.2), 77.3 (35.6), and 44.7 (20.6). These all fall within the bounds determined by Hawken and Parker as being realistic for a d-DOG-s spatial function based on lateral geniculate inputs (see Derrington & Lennie, 1984). Therefore we feel confident that the WIM model matches to our MT data did not rely on physiologically implausible properties at the level of the V1 neuron inputs.

3.3.2. Speed tuning parameter (v)

We also examined the values of the other main parameter in the WIM model, namely the preferred speed tuning of the WIM sensor that produced the best match to the MT data. Fig. 12 shows the distribution of the value of the optimum speed tuning of the WIM motion sensor across our population of MT neurons. It is bimodal with a median of 11.5°/s and a mean of 22.67°/s (S.D. = 32.2°/s). If we accept that the WIM model tuning closely reflects the optimum tuning of the fitted MT neuron (see Fig. 10 above) then the average optimum speed tuning of our MT sample was reasonably close to that typically reported for MT neurons (approximately 32°/s, Maunsell & Van Essen, 1983).

3.3.3. The alpha value

The α term in Eq. (1) controls the length of the ridge in the SRF of the WIM model sensor (see Fig. 7c). In effect it controls the range of spatial and temporal frequencies the sensor will optimally respond to when the stimulus speed is held constant. For our model–MT data fits, the mean value for α was 316.3 (S.D. = 368) and the median was 133.8. For comparison, the α value in Fig. 7c was 100. Some of the MT neurons were best fit with model SRFs that were relatively localized for spatial frequency; others exhibited ridges of peak response that extended over a broad range of frequencies. Although the units of α do not correspond directly to firing rate (the units are arbitrary), the α term does relate to the combined background spontaneous firing rate of the sustained and transient V1 neurons. Our results show that if the T and S input units of the model do relate directly to the transient and sustained neurons found in V1, then one would expect to find a wide range of spontaneous activity across a population of these cells. One of the functions of spontaneous activity, according to our modeling results, might be to control the length of the peak sensitivity ridge in the MT neuron's SRFs.

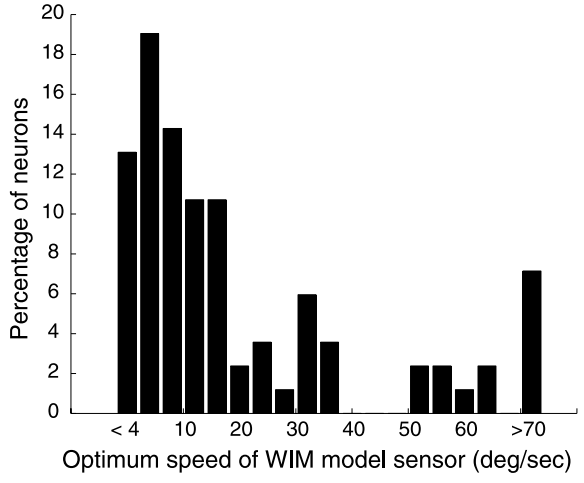

Fig. 12. Distribution of preferred speeds in the WIM model sensors that produced the best fit to our MT neuron data ($N = 84$).

3.3.4. The ζ (zeta) value

The ζ term in the equations for the temporal frequency tuning of the V1 neurons in the model controls the degree of band-pass tuning (see Appendix A). A value of 0.0 indicates low-pass temporal frequency tuning and 1.0 corresponds to band-pass tuning with the low-frequency limb passing through 0.0 Hz (see Watson, 1986). For the WIM model fitting to the MT data, the ζ value for the sustained V1 neuron was always fixed at 0.0. For the transient neuron, the best fits to the MT data were obtained when ζ took on values that ranged from 0 to 1.0. The mean was 0.5 (S.D. = 0.33). This indicates that the majority of transient V1 units in the model did not have to be strongly band-pass in their temporal frequency tuning in order to provide good fits to the MT SRFs. Some of these 'transient' neurons could be categorized as 'low-pass'.

3.4. Other potential parameters affecting the fits

We restricted the spatial frequency tuning of the V1 neurons to just one of the d-DOG-s functions reported by Hawken and Parker (1987). It is the function that they studied the most (their Fig. 6a) and it does appear to be typical in shape when compared to their other published figures. However they did report a variety of other spatial tuning functions (see their Fig. 13) and we may have obtained better fits to our MT data if we had allowed some of the other parameters of the d-DOG-s function to vary (besides the peak spatial frequency). Similarly, the basic shape of the V1 temporal frequency tuning functions remained constant throughout our fitting procedure. We only varied one parameter related to these functions and that was the degree of band-pass tuning of the V1 transient neuron. Actual V1 neurons exhibit a variety of temporal frequency tuning functions (Foster et al., 1985; Hawken et al., 1996). Even better fits may have been achieved if this aspect of the V1 neuron properties had also been allowed to vary.

The close fits between the WIM model and the MT data mean that we now have at our disposal a powerful

tool for better analyzing the SRFs of MT neurons. We no longer have to use abstract theoretical fitting functions such as the 2-D Gaussian in order to extract features like the preferred speed tuning of the MT neuron (see Perrone & Thiele, 2001). As can be seen from the analyses above, we are now also in a position to make predictions about the properties of the MT neuron's V1 afferents.

4. Discussion

In this paper we have demonstrated that a neuron with an oriented SRF can be generated from just the combination of a sustained and transient V1 neuron. This neural mechanism is tuned for a particular speed of edge motion. We have previously verified that there exist neurons in extrastriate area MT (a target area for V1 neurons) that have oriented SRFs (Perrone & Thiele, 2001). In this paper, we further demonstrated that the MT SRFs can be mimicked using the WIM model neural mechanism.

4.1. Alternative models of speed estimation

The principles underlying the construction of the WIM model speed-tuned mechanism are fundamentally different from what has been proposed previously in speed estimation models. Many motion models have been proposed which attempt to estimate the slope of the edge spectral line in order to obtain a *direct* estimate of the edge speed. Early on it was recognized that the ratio of the output from a transient spatiotemporal tuned neuron to the output of a sustained neuron produces a result that is approximately linearly related to the speed of the edge (Adelson & Bergen, 1986; Harris, 1986; Thompson, 1982; Tolhurst et al., 1973). Unfortunately there are currently no physiological data to support the idea that the visual system is using the transient and sustained V1 neurons' outputs in this particular way. If direct velocity estimation were being performed, one would expect to find a signal proportional to the edge speed at some stage of the motion pathway, i.e., a neuron that responds a little for slow speeds and a lot for high speeds with a linear response pattern in between (rate coding). The best candidate for this role is an MT neuron, but the majority of these are speed tuned (Lagae et al., 1993; Maunsell & Van Essen, 1983; Perrone & Thiele, 2001). Evidence for a neural signal directly related to the image speed has yet to be found. Admittedly other speed estimation schemes that are not based around the spatiotemporal aspects of the stimulus are still possible (e.g., tracking edge features over time; see Del Viva & Morrone, 1998).

A model developed by Simoncelli and Heeger (1998) uses the interaction between a number of spatiotemporal filters to isolate the edge spectral line and hence to determine the speed of the moving edge (see also Grzywacz & Yuille, 1990; Heeger, 1987). The problem with the Simoncelli and Heeger approach is that the spatial and temporal frequency tuning properties of the filters in their model do not match those found in V1 neurons. Their model 'V1 neurons' have spatiotemporal SRFs whose main axes are not parallel to the spatial and temporal frequency axes (i.e., they are inseparable). This conflicts with the SRF properties of actual V1 neurons (Fig. 1c). In a recent commentary on our work (Simoncelli & Heeger, 2001) these authors sketch out an alternative method for constructing speed-tuned MT neurons based on V1 afferents that are separable (see their Fig. 2). They argue that a typical MT neuron field can be built up by summing the SRFs from three V1 neurons, each tuned to a different peak spatial frequency and a different peak temporal frequency. First off, this mechanism is inefficient compared to the WIM model, because it requires a broad range of V1 spatial frequency channels feeding into each MT neuron for it to work. The WIM model only requires one spatial channel (ignoring the small difference between the sustained and transient peak spatial tuning). The Simoncelli and Heeger scheme also requires that there be a large number of different temporal frequency channels in V1, some with very tight bandwidths and some with extremely high peak temporal frequency tuning. This is hard to reconcile with the evidence pointing to just two broad temporal frequency channels (one low-pass, the other moderately band-pass peaking around 6–12 Hz) in V1 neurons (Foster et al., 1985; Hawken et al., 1996; Tolhurst & Movshon, 1975) and from human psychophysics (Anderson & Burr, 1985; Hammett & Smith, 1992; Watson & Robson, 1981).

Nowlan and Sejnowski (1995) have developed a selection model of motion processing in MT which uses more traditional motion energy filters (Adelson & Bergen, 1985) for the V1 stage units. These do have spatial and temporal frequency tuning properties consistent with what is found in V1. However the V1–MT connectivity in this model is established through the use of a neural network learning algorithm and so it is difficult to infer the exact role of the individual spatiotemporal filters (out of a set of 36) in the final velocity estimation process.

4.2. Speed tuning versus direct speed estimation

The assumption behind many previous motion models is that the visual system is attempting to derive a direct estimate of the edge velocity (the velocity 'flow-field') for the recovery of higher-level information such as self-motion and depth estimation (Heeger & Jepson, 1992; Koenderink & van Doorn, 1975). An alternative

view exists which is better supported by the physiological data; i.e., the visual system uses speed-tuned neurons for the processing of higher-level motion (Perrone, 2001). Neurons with these properties are abundant in area MT (Maunsell & Van Essen, 1983; Perrone & Thiele, 2001). We have previously shown that self-motion information and depth recovery can be derived from networks made up of speed- and direction-tuned sensors with the properties of MT neurons, without the need for direct speed estimation (Perrone, 1992; Perrone & Stone, 1994, 1998). Therefore the type of speed tuning achieved with the model is adequate for high-level motion processing tasks such as self-motion and depth estimation.

4.3. V1 neuron spatial frequency tuning functions

The WIM model can explain a number of unusual properties that have been discovered in V1 neurons. For example, the 'complexity' of the spatial function needed by Hawken and Parker (1987) to fit their V1 neurons' spatial contrast sensitivity data is surprising. If the sole task of a V1 neuron was to selectively pass a narrow band of spatial frequencies, then one would expect a simple inverted U-shaped function such as the Fourier transform of a Gabor function, or a single difference of Gaussians to suffice. Yet the nine parameter d-DOG-s function was needed to account for the wide variety of spatial contrast sensitivity functions that Hawken and Parker found in their population of V1 neurons. It was also needed in the WIM model to simulate the MT data. The WIM model supports the claim by Hawken and Parker (1987) and others (Stork & Wilson, 1990; Young, 1987) of the need for a complex function such as the d-DOG-s and provides a possible reason for the complexity. It suggests that the flexibility of form found in V1 neurons' spatial functions is a natural consequence of the visual system's attempt at constructing a speed-tuned neural mechanism.

4.4. V1 neuron temporal frequency tuning

In the examples presented, the sustained temporal frequency contrast sensitivity function was simulated using a purely low-pass function. However the WIM model does not rely on this assumption; as long as the sustained function is less band-pass than the transient function, the intersection mechanism will work. The two neuron types can actually possess a wide variety of temporal tuning curves consistent with the variation seen in some V1 data (Hawken et al., 1996). The transient neurons' temporal frequency tuning functions do not have to be strongly band-pass in order to achieve good fits of the WIM model SRF to the MT data. In many instances, the model fit to the MT data was achieved when the 'transient' neuron had a very low transience factor and it could easily be mistaken for a temporally low-pass ('sustained') tuned unit.

4.5. 'Parvocellular' and 'magnocellular' pathways

Sustained and transient temporal responses have also been associated with the two classes of neurons found in the primary visual pathway: parvocellular and magnocellular (Lennie, 1980; Livingstone & Hubel, 1988). The WIM model analysis shows that the distinction between 'transient' and 'sustained' temporal tuning need not be large and the differences in spatial frequency tuning predicted by the WIM model are often very subtle and may be difficult to detect. It suggests that there may be a lot of overlap between the parvocellular and magnocellular neuron classes along the sustained–transient dimension. The WIM model adds to the arguments put forward by others that motion processing does not belong exclusively to the magnocellular system (Galvin, Williams, & Coletta, 1996; Merigan & Maunsell, 1993) and suggests a means by which the parvocellular and magnocellular systems may interact for the purpose of motion processing.

4.6. Component versus pattern motion processing

Spatial frequency space actually needs to be considered in two dimensions (u_x and u_y) because edges move in a variety of directions and have many orientations. The above discussion of the WIM model has concentrated on the speed tuning aspects of the motion sensitive neurons and assumes that the direction tuning of the neurons coincides with the direction of the edge normal. In this respect it is a model of how MT-like 'component' neurons could be constructed from V1 neurons rather than a model of MT 'pattern' neurons (Adelson & Movshon, 1982).

4.7. Coarse-to-fine perceptual discrimination

The properties of the final speed-tuned mechanism generated by the WIM model contrast sharply with those of the two underlying neural components. Two neurons, each with very broad temporal tuning and with no orientation in spatiotemporal frequency space, are able to work together to produce a speed-tuned mechanism that is precisely matched to the input generated by a moving edge. The proposed weighted intersection mechanism could form the basis of a general principle by which biological systems obtain very fine perceptual discriminations from the broadly tuned neural processors common to many of the senses.

Acknowledgements

The authors thank Drs. Brent Beutter, Rich Krauzlis and Lee Stone for their helpful comments on earlier drafts. Special thanks to Michael Hawken for providing additional V1 temporal frequency tuning data. Supported by NASA grant NAG 2-1168 to JP and a Human Frontier Science Program fellowship to AT. Some of the research reported in this paper was carried out during tenure by JP as a Sloan Visiting Scientist at the Salk Institute. Electrophysiological data were recorded in T.D. Albright's laboratory at the Salk Institute.

Appendix A

A.1. Temporal frequency contrast sensitivity tuning

The V1 temporal frequency data were modeled and fit using the following function:

$$z(\omega) = \sqrt{m_1^2 + (\zeta^2 + m_2^2) - 2\zeta m_1 m_2 \cos(\vartheta_1 + \vartheta_2)},$$

where $m_1 = ((2\pi\omega\tau_1)^2 + 1)^{-9/2}$, $m_2 = ((2\pi\omega\tau_2)^2 + 1)^{-10/2}$, $\vartheta_1 = -9\tan^{-1}(2\pi\omega\tau_1)$ and $\vartheta_2 = -10\tan^{-1}(2\pi\omega\tau_2)$.

We used the above equation to model the contrast sensitivity (z) of the neuron for each value of the stimulus temporal frequency (ω) measured in Hz. This particular function was based on the temporal filter developed by Watson (1986) and by Watson and Ahumada (1985) in their motion model. Other functions such as the difference of exponentials (Derrington & Lennie, 1984) could also be used, but the Watson and Ahumada function can be constrained to only three parameters (ζ, τ_1, τ_2). The key parameter is the transience factor (ζ) which changes the function from low-pass or sustained ($\zeta = 0$) through to band-pass or transient ($\zeta = 1$). The other two parameters are time constants, measured in seconds. The model simulations described in this paper used sustained neuron temporal function parameter values of $(0, 0.0072, 0.0043)$ for (ζ, τ_1, τ_2). These were derived from a least squares minimization fitting procedure (see below) applied to the V1 temporal data shown in Fig. 1a (open circles, data taken from Foster et al., 1985). The transient neurons' temporal functions used a range of ζ values, but τ_1 and τ_2 were set to 0.0059 and 0.0115, respectively. These values were derived from the V1 data in Fig. 1a (solid and open squares, data taken from Foster et al., 1985, and Hawken, personal communication). The two model temporal frequency contrast sensitivity functions (sustained and transient) are plotted in Figs. 2a (log–log plot) and 3a. The amplitude of the transient temporal function relative to that of the sustained function was calculated so that the cross-over point occurred at ω_0 Hz such that $\omega_0 = vu0$, where v is the preferred speed tuning of the WIM sensor and $u0$ is the peak spatial frequency of the sustained V1 neuron (see below).

A.2. Spatial frequency contrast sensitivity tuning

The spatial frequency contrast sensitivity tuning of the V1 neurons was modeled using the magnitude part of the Fourier transform of a d-DOG-s function, identical to that used by Hawken and Parker (1987). The data from one of their neurons and the best fitting d-DOG-s function are shown in Fig. 1b. The magnitude part of a Fourier transformed d-DOG-s function has the following equation:

$$z(u) = \sqrt{r_1^2 - 2r_1 r_2 \cos(2\pi uS) + (r_2\cos(2\pi uS))^2 + ((1-2g)r_2\sin(2\pi uS))^2},$$

where $r_1 = p_1 - q_1$ and $r_2 = p_2 - q_2$, with

$$p_1 = A_1\exp(-xc_1(\pi u)^2); \quad q_1 = A_2\exp(-xs_1(\pi u)^2);$$

$$p_2 = A_3\exp(-xc_2(\pi u)^2); \quad q_2 = A_4\exp(-xs_2(\pi u)^2).$$

The above equation generates the contrast sensitivity (z) for each value of the stimulus spatial frequency (u) measured in cycles/deg. This function has 10 parameters that relate to the three individual differences of Gaussian (DOGs) that together make up the space domain spatial receptive field. The x terms are the space constants of the individual Gaussians with xc_1 and xc_2 controlling the size of the central part and xs_1, xs_2 the surrounds of the DOGs. There is also a parameter to control the separation of the DOGs in the space domain (S), and one that controls the symmetry of the overall spatial receptive field (g). The amplitudes of the Gaussians (A_1, A_2, A_3, A_4) relate to the contrast sensitivity of the LGN neurons assumed to be subunits of the receptive field (see Hawken & Parker, 1987). The total number of parameters is reduced to nine by applying a constraint such that $(A_1 - A_2) - (A_3 - A_4) = 0$. This follows Hawken and Parker's treatment and is based on their assumption that the combined sensitivity of all sub-regions of the receptive field sums to zero. The parameter values for the spatial function shown in Fig. 1b are: 43, 43, 41, 41, 2.22′, 15.3′, 4.97′, 17.41′, 0.25, 8.23′ for A_1, A_2, A_3, A_4,

xc_1, xs_1, xc_2, xs_2, g and S, respectively. Different values of $u0$ (the peak spatial frequency of the V1 neurons) were generated by scaling the x terms in the above equation by a scale factor of $3.0/u0$ (since the peak spatial frequency for the Hawken and Parker neuron was 3.0 cycles/deg).

A.3. Equating the transient and sustained V1 spatiotemporal contrast sensitivities

Let the sustained temporal frequency contrast sensitivity function (dashed line on right wall of Fig. 4a) be represented by $z = p(\omega)$ and the transient function (solid line on right wall of Fig. 4b) by $z = m(\omega)$. Also let the sustained spatial frequency function (dashed line on back wall of Fig. 4a) be defined as $z = f(u)$ and the transient spatial function (solid line on back wall of Fig. 4b) as $z = f'(u)$. The combined contrast sensitivity of the sustained neuron, $S(u, \omega)$, is given by $f(u)p(\omega)$ and the sensitivity of the transient neuron is $T(u, \omega) = f'(u)m(\omega)$. Set $T(u, \omega)$ and $S(u, \omega)$ to be equal and let $\omega = vu$. Therefore we have $f'(u)m(vu) = f(u)p(vu)$. The values of $p(vu)$ and $m(vu)$ are known and can be found from the transient and sustained temporal frequency functions based on V1 data (see Fig. 2a). The value of $f(u)$ is also known (see the spatial frequency tuning function of the sustained V1 neuron in Fig. 2b). The only term required to make this equality hold is $f'(u)$, the spatial frequency tuning of the transient V1 neuron. Rearranging terms, it is easy to solve for the transient spatial frequency sensitivity using

$$z = f'(u) = \frac{p(vu)}{m(vu)} f(u), \tag{A.1}$$

i.e., $z = Rf(u)$ where R corresponds to the sustained neuron's temporal sensitivity at $\omega = vu$ Hz divided by the transient neuron's temporal sensitivity at $\omega = vu$ Hz.

It is convenient to demonstrate the concept behind the intersection mechanism by assuming separability and there is some evidence to support this assumption (Foster et al., 1985; Tolhurst & Movshon, 1975). However the same end result could be achieved by manipulating the individual contrast sensitivity functions of the T and S V1 neurons, i.e., they could each be slightly inseparable and yet still be made to intersect along the $\omega = vu$ line.

A.4. Function and model fitting procedure

Whenever a temporal function or a d-DOG-s function was fit to data, the sum of squared differences between the fitted function and the data was minimized using fminsearch in MatLab (MathWorks, Natick, MA). For the spatial fitting, the parameters were limited to positive values, the maximum spatial frequency ($u0$) was limited to 10 cycles/deg, and the g value was constrained to lie between 0 and 1.0. The same minimum least squares fitting procedure was used to fit the 30 WIM model values to the 30 MT neuron spatial frequency responses (see Perrone & Thiele, 2001). In this case the α value in Eq. (1) was constrained to be less than 1000 and all parameter values were limited to positive values only.

References

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

Adelson, E. H., & Bergen, J. R. (1986). The extraction of spatiotemporal energy in human and machine vision. In *Proceedings of IEEE workshop on motion: Representation and analysis, Charleston, SC* (pp. 151–156).

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*(9 December), 523–525.

Anderson, S. J., & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detecting system. *Vision Research, 25*, 1147–1154.

Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research, 38*, 2769–2786.

Cleland, B. G., Dubin, M. W., & Levick, W. R. (1971). Sustained and transient neurones in the cat's retina and lateral geniculate nucleus. *Journal of Physiology, 217*, 473–496.

Del Viva, M. M., & Morrone, M. C. (1998). Motion analysis by feature tracking. *Vision Research, 38*, 3633–3653.

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

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

Felleman, D. J., & Kaas, J. H. (1984). Receptive-field properties of neurons in the middle temporal visual area (MT) of owl monkeys. *Journal of Neurophysiology, 52*, 488–513.

Foster, K. H., Gaska, J. P., Nagler, M., & Pollen, D. A. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *Journal of Physiology, 365*, 331–363.

Galvin, S. J., Williams, D. R., & Coletta, N. J. (1996). The spatial grain of motion perception in human peripheral vision. *Vision Research, 36*, 2283–2295.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

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.

Hammett, S. T., & Smith, A. T. (1992). Two temporal channels or three? A re-evaluation. *Vision Research, 32*, 285–291.

Harris, M. G. (1986). The perception of moving stimuli: a model of spatiotemporal coding in human vision. *Vision Research, 26*, 1281–1287.

Hawken, M. J., & Parker, A. J. (1987). Spatial properties of neurons in the monkey striate cortex. *Proceedings of the Royal Society of London B, 231*, 251–288.

Hawken, M. J., Parker, A. J., & Lund, J. S. (1988). Laminar organization and contrast sensitivity of direction selective cells in the striate cortex of the old world monkey. *Journal of Neuroscience, 8*, 3541–3548.

Hawken, M. J., Shapley, R. M., & Grosof, D. H. (1996). Temporal frequency selectivity in monkey visual cortex. *Journal of Neuroscience, 13*, 477–492.

Hays, W. L. (1963). *Statistics for psychologists*. New York: Holt, Rinehart and Winston.

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., & Jepson, A. D. (1992). Subspace methods for recovering rigid motion I: algorithm and implementation. *International Journal of Computer Vision, 7*, 95–177.

Koenderink, J. J., & van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to the observer. *Optica Acta, 22*, 773–791.

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient neurones in the human visual system. *Journal of Physiology, 232*, 149–162.

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

Lennie, P. (1980). Parallel visual pathways: a review. *Vision Research, 20*, 561–594.

Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science, 240*, 740–749.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London B, 208*, 385–387.

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

Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience, 16*, 369–402.

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

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology, 283*, 79–99.

Nakayama, K. (1984). Biological image motion processing: a review. *Vision Research, 25*(5), 625–660.

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

Perrone, J. A. (1992). Model for the computation of self-motion in biological systems. *Journal of the Optical Society of America, 9*, 177–194.

Perrone, J. A. (1994). Simulating the speed and direction tuning of MT neurons using spatiotemporal tuned V1-neuron inputs. *Investigative Ophthalmology and Visual Science (Supplement), 35*, 2158.

Perrone, J. A. (1998). Precise visual speed estimation from broadly tuned V1-like spatiotemporal filters. *Investigative Ophthalmology and Visual Science, 37*, 2359.

Perrone, J. A. (2001). A closer look at the visual input to self-motion estimation. In J. M. Zanker, & J. Zeil (Eds.), *Motion vision. Computational, neural, and ecological constraints* (pp. 169–179). Heidelberg: Springer.

Perrone, J. A., & Krauzlis, R. (2002). Simulating the time course of MT neuron responses with a model based on V1 neuron properties. Vision Sciences Society meeting abstract.

Perrone, J. A., & Stone, L. S. (1994). A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research, 34*, 2917–2938.

Perrone, J. A., & Stone, L. S. (1998). Emulating the visual receptive field properties of MST neurons with a template model of heading estimation. *Journal of Neuroscience, 18*, 5958–5975.

Perrone, J. A., & Thiele, A. (2000). Predicting the optimum speed tuning of MT neurons from their SRFs. *Society for Neuroscience Abstracts, 26*, 399.3.

Perrone, J. A., & Thiele, A. (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nature Neuroscience, 4*(5), 526–532.

Reisbeck, T. E., & Gegenfurtner, K. R. (1999). Velocity tuning mechanisms in human motion processing. *Vision Research, 39*, 3267–3285.

Robson, J. G. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America, 56*, 1141–1142.

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

Simoncelli, E., & Heeger, D. J. (2001). Representing retinal image speed in visual cortex. *Nature Neuroscience, 4*(5), 461–462.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of the neuronal responses in visual area MT. *Vision Research, 38*, 743–761.

Stone, L. S., & Thompson, P. (1992). Human speed perception is contrast dependent. *Vision Research, 32*, 1535–1549.

Stork, D. G., & Wilson, H. R. (1990). Do Gabor functions provide appropriate descriptions of visual cortical receptive fields? *Journal of the Optical Society of America, 7*, 1362–1373.

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research, 22*, 377–380.

Thompson, P. (1983). Discrimination of moving gratings at and above detection threshold. *Vision Research, 23*, 1533–1538.

Thompson, P. (1984). The coding of velocity of movement in the human visual system. *Vision Research, 24*, 41–45.

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

Tolhurst, D. J., Sharpe, C. R., & Hart, G. (1973). The analysis of the drift rate of moving sinusoidal gratings. *Vision Research, 13*, 2545–2555.

Watson, A. B. (1986). Temporal sensitivity. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1) (pp. 6.1–6.42). New York: Wiley.

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

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

Watson, A. B., & Robson, J. G. (1981). Discrimination at threshold: labelled detectors in human vision. *Vision Research, 21*, 1115–1122.

Young, R. A. (1987). The Gaussian derivative model for spatial vision: I. Retinal mechanisms. *Spatial Vision, 2*, 273–293.