A neural-based code for computing image velocity from small sets of middle temporal (MT/V5) neuron inputs

John A. Perrone

School of Psychology, The University of Waikato, Hamilton, New Zealand



It is still not known how the primate visual system is able to measure the velocity of moving stimuli such as edges and dots. Neurons have been found in the Medial Superior Temporal (MST) area of the primate brain that respond at a rate proportional to the speed of the stimulus but it is not clear how this property is derived from the *speed-tuned* Middle Temporal (MT) neurons that precede area MST along the visual motion pathway. I show that a population code based on the outputs from a number of MT neurons is susceptible to errors if the MT neurons are tuned to a broad range of spatial frequencies and have receptive fields that span a wide range of sizes. I present a solution that uses the activity of just three MT units within a velocity *channel* to estimate the velocity using a weighted vector average (centroid) technique. I use a range of velocity channels (1, 2, 4, and 8°/s) with inhibition between them so that only a single channel passes the velocity estimate onto the next stage of processing (MST). I also include a contrast-dependent redundancy-removal stage which provides tighter spatial resolution for the velocity estimates under conditions of high contrast but which trades off spatial compactness for greater sensitivity at low contrast. The new model produces an output signal proportional to the stimulus input velocity (consistent with MST neurons) and its input stages have properties closely tied to those of neurons in areas V1 and MT.

Keywords: velocity, 2D motion, MT

Citation: Perrone, J. A. (2012). A neural-based code for computing image velocity from small sets of Middle Temporal (MT/ V5) neuron inputs. *Journal of Vision*, *12*(8):1, 1–30, http://www.journalofvision.org/content/12/8/1, doi:10.1167/12.8.1.

Introduction

Our ability to move our eyes and bodies means that we are constantly exposed to visual motion. Whether we are the predator or the prey, the correct measurement and interpretation of that motion is crucial for our survival. Yet despite many years of effort, we still do not have a complete understanding of how the primate visual system is able measure the velocity of even the most basic visual motion stimulus (e.g., a moving edge).

We do know that in the Fourier (frequency) domain, moving edges are represented by a spectrum that is oriented relative to the spatial and temporal frequency (tf) axes (Watson & Ahumada, 1983) (Figure 1). The slope of the spectral line is proportional to the speed of the edge. The estimation of the edge's velocity is equivalent to determining the orientation of the edge's spatiotemporal energy spectrum in frequency space. We know that humans and many other biological species are able to correctly register the orientation of the edge spectrum (determine the velocity of the edge) under a wide range of conditions (Burr & Thompson, 2011; Clifford & Ibbotson, 2002; Hildreth, 1990; Nakayama, 1985; Smith & Snowden, 1994). What is not known is the exact mechanism for the velocity estimation process and how it is achieved, given our current understanding

of the properties of primate cortical motion sensitive neurons.

Figure 1 shows a Fourier frequency space representation of edges moving right to left at 1, 2, and 4°/s. The locus of spectral energy generated by each of the moving edges is shown by the red shaded radial lines in the plot. In order to estimate the velocity of the $2^{\circ}/s$ edge, we must be able to detect the correct orientation of its edge spectrum and distinguish it from the energy spectra generated by edges moving at 1 and 4° /s as well as other directions. In an extensive review of biological image motion processing, Nakayama (1985, p. 643) postulated the existence of "spatio-temporal filters which share a common velocity" and which are located along the radial velocity lines shown in Figure 1. He then suggested that: "Velocity could be read out by comparing activity in these different higher order radial 'velocity' channels. This could be determined by detecting the mode or the peak of the population profile response possibly with the aid of lateral inhibition" (p. 643).

Despite the passage of more than 30 years, Nakayama's tantalizing proposal has never been successfully implemented and the exact mechanism (mode, peak detection, or something else?) has never been discovered. Theoretical treatments which attempted to register the slope of the edge spectra using spatiotemporal filters (Heeger, 1987; Simoncelli &



Figure 1. Spatiotemporal frequency (Fourier domain) plot showing the spectra generated by edges moving at a range of speeds. The dashed contour represents the amplitude spectrum of a typical sustained type nondirectional V1 neuron with low-pass temporal frequency tuning. The solid contour is for a transient type directional V1 neuron with band-pass tf tuning.

Heeger, 1998; Yuille & Grzywacz, 1988) fall short because of the lack of concordance between the model filters and the known properties of primate motion sensitive neurons (Perrone, 2004). These early models tried to implement the "radial velocity channels" postulated by Nakayama using spatiotemporal energy filters (Adelson & Bergen, 1985; Watson & Ahumada, 1985) directly in their raw form. The proposed energy filters were originally modeled on the properties of a class of neurons in the primary visual cortex (V1) which turned out to be broadly tuned for temporal frequency (Foster, Gaska, Nagler, & Pollen, 1985; Hawken, Shapley, & Grosof, 1996) (see contour lines in Figure 1). These neurons lack the specifications of a filter designed to optimally detect the orientation of the edge spectral lines (Perrone, 2004; Perrone & Thiele, 2002).

A better candidate for the radial velocity channels suggested by Nakayama (1985) are primate Middle Temporal (MT/V5) neurons (Albright, 1984; Maunsell & Van Essen, 1983; Movshon, Adelson, Gizzi, & Newsome, 1983; Zeki, 1980). Rather than velocity being estimated directly from the V1 neurons, it is possible to introduce an intermediate stage in which speed tuned pattern motion detectors (MT neurons) are first constructed from V1 neurons (e.g., Adelson & Movshon, 1982; Albright, 1984; Movshon et al., 1983; Perrone, 2004; Perrone & Krauzlis, 2008a) and then these detectors are used to isolate the edge spectral lines.

Figure 2a shows the spectral receptive field for a macaque MT neuron measured by Perrone and Thiele

(2001; see also Priebe, Cassanello, & Lisberger, 2003). The spectral receptive field is oriented and narrow in the tf dimension, which are desirable properties for a filter that needs to selectively respond to a particular radial edge spectral energy line (velocity). In a recent review of primate velocity computation, Bradley and Goyal (2008) recognized the need for tighter tf tuning in the spatiotemporal V1 stage filters and suggested that velocity estimation at the MT neuron level could be carried out by a "flattened inner tube" configuration with filters that are more compressed in the tf dimension compared to those used in previous models (Simoncelli & Heeger, 1998). Nishimoto and Gallant (2011) have also suggested a spatiotemporal receptive field model that uses more flattened filters. However neither Bradley and Goyal nor Nishimoto and Gallant provided any mechanism for how this "filter flattening" could come about. Our recent work provides a possible solution to this problem. We have proposed a method (Weighted Intersection Mechanism, or WIM) by which broad tf tuned V1 neurons can be converted into "speed tuned" filters that are narrowly tuned in the tf dimension (Perrone, 2004, 2005; Perrone & Thiele, 2002) and which match the properties of neurons in V1 (Priebe, Lisberger, & Movshon, 2006) and MT (Perrone & Thiele, 2001). Figure 2b shows the spectral receptive fields of two of our model MT units. They have been constructed from just two V1 neuron inputs (see Figure 1a) using biologically plausible mechanisms (Perrone, 2004; Perrone & Krauzlis, 2008a; Perrone & Thiele, 2002).

Each of our MT pattern model units are made up from a set of subunits based on V1 directionally selective complex neurons (Figure 2c). The subunits are speed tuned (via the weighted intersection mechanism proposed by Perrone & Thiele) and are analogues of the V1 complex neurons discovered by Priebe et al. (2006). In frequency space they form a set of flying saucers (or a flattened inner tube a la Bradley & Goyal, 2008) and are narrow in the temporal frequency dimension. The MT units inherit their speed tuning from the WIM subunits and so their spectral receptive fields are also narrow in tf (Figure 2b). In the space domain, an MT pattern model unit can be represented as a cluster of WIM subunits (see the flowerettes in Figure 2d) where each arrow represents the speed tuning of the WIM subunits (Perrone, 2004; Perrone & Krauzlis, 2008a).

Despite their apparent suitability as filters for registering the slope of the energy spectra created by moving edges, MT neurons are not an automatic choice for Nakayama's radial velocity channels. There has been some controversy as to the existence and extent of spatiotemporal frequency orientation tuning in primate neurons (Perrone, 2006; Priebe et al., 2003). Initial experiments (Perrone & Thiele, 2001) found clear



Figure 2. MT pattern neurons (actual and model) in the spatiotemporal frequency domain and the spatial domain. (a) Spatiotemporal frequency response map (spectral receptive field) for an MT neuron from Perrone & Thiele (2001). (b) Spectral receptive fields from two model MT units. (c) Frequency space representation showing the spectrum for a moving stimulus (pink plane) and the speed tuned filters (WIM sensors) used as subunits in the model MT pattern neurons. (d) Space domain plot of a model MT pattern unit receptive field. The arrows represent the speed tuning of the WIM subunits. Dashed arrows represent inhibitory (opponent) inputs (Perrone & Krauzlis, 2008a).

evidence for spatiotemporal frequency inseparability in MT neurons (e.g., see Figure 2a) but later studies reported it to be a weak effect (Priebe et al., 2003; Priebe et al., 2006). New research showing the effect of contrast on the speed tuning of V1 and MT neurons (Krekelberg, van Wezel, & Albright, 2006; Pack, Hunter, & Born, 2005; Perrone, 2006; Priebe et al., 2006) suggests that the measurement of spatiotemporal frequency orientation properties in V1 and MT neurons is fraught with difficulty (Perrone, 2006) and the actual proportion of such sf-tf-oriented filters in the primate visual system is currently unknown.

I will sidestep this prevalence debate and work on the assumption that some MT neurons with oriented

spectral receptive fields do exist in the primate visual system. Further, given their velocity sensitive properties (Figure 2a through c), I will adopt these MT neurons as the main building blocks in my new velocity estimation model and use them to construct the radial velocity channels suggested by Nakayama (1985) for the detection of image velocity. We have previously developed a detailed model of the MT neurons (Perrone, 2004; Perrone & Krauzlis, 2008a) and this forms the starting point for my velocity estimation stage. However the MT model treated each MT neuron as an independent unit with no interaction with other neurons. In this paper I show that inhibitory connections with other MT neurons are an essential feature of an effective velocity estimation system.

Signals from MT neurons end up in the Medial Superior Temporal (MST) area (Duffy & Wurtz, 1991; Komatsu & Wurtz, 1988; Perrone & Stone, 1998; Saito et al., 1986; Tanaka et al., 1986; Ungerleider & Desimone, 1986) and there is evidence that some MST neurons respond at a rate proportional to the speed of the input motion and their output is linearly related to the input over a wide range of speeds (Inaba, Shinomoto, Yamane, Takemura, & Kawano, 2007). In contrast, the neurons in antecedent motion areas (V1, MT) tend to be speed tuned (Maunsell & Van Essen, 1983; Perrone & Thiele, 2001; Priebe et al., 2006; Rodman & Albright, 1987). It is the stage after MT that I am attempting to replicate in this paper, the point where the primate visual system transitions from speed and direction tuning to velocity signaling. I aim to emulate the velocity coding response properties of MST neurons (Inaba et al., 2007).

The decision to incorporate speed tuned (sf-tf oriented) filters into the velocity estimation process is only a small step towards the development of a successful neural velocity estimation model. What remains unanswered is how the outputs from these MT filters are combined to produce an output proportional to the velocity of the moving stimulus. Nakayama (1985) speculated as to a number of possible approaches (detection of the mode or peak response possibly with the aid of lateral inhibition) but never put forward a specific mechanism. A large number of proposals have since been suggested for how image velocity can be measured (see Sperling, Neil, & Paul, 2001 and Discussion) but these models either do not include the V1-MT stage of neural processing and assume that the MT output has already been derived or they use a V1 stage that includes elements incompatible with the known properties of motion sensitive neurons in V1 (Perrone, 2004).

I have rectified these deficiencies and have now developed a neural-based model for estimating the velocity of moving image features in two-dimensional image sequences. I will refer to the system for estimating velocity as a "velocity code model" to retain consistency with the common usage of the term "population code" to characterize the estimation of a stimulus property from a population of neurons. I show how a relatively small number of feed-forward stages based mainly on multiscale filtering and inhibitory interactions can generate an output linearly related to the input image speed (thus emulating MST neuron behavior). A guiding principle in the design of the new velocity code is to reduce the amount of redundant velocity signals that are passed onto the MST stage and I introduce a contrastdependent redundancy removal mechanism for achieving this. I show that the new model is able to accurately

estimate the velocity of moving features while maintaining concordance between the model's component filters and the known properties of motion sensitive neurons. A model that includes multiple neural stages and mechanisms necessarily acquires a high level of complexity and so in order to simplify the description I have divided the overall velocity code model into a number of stages:

- (1) Speed estimation. I show how the speed of motion of a moving image feature such as an edge can be derived from the outputs of a small number of MT neurons with inhibitory connections between them.
- (2) Direction estimation. Because it is based on the outputs of a number of MT pattern neurons the basic speed estimation model is subject to errors when multiple image directions are considered. I introduce a mechanism that uses inhibition from MT units tuned to nearby directions that overcomes this class of error and which enables the direction of a moving edge to be correctly registered.
- (3) Contrast-dependent redundancy removal. The basic velocity code model generates multiple velocity signals across space and these signals are often redundant in that the same signal is created at many adjacent locations. I introduce a mechanism that removes this redundancy and improves the spatial resolution of the velocity estimates at high stimulus contrast levels but not at low contrast.
- (4) Small dot stimuli. Single moving dots create very little motion energy in the early spatiotemporal energy stage of the velocity code model. This has an impact on the effectiveness of some of the other mechanisms used in the model. I present a technique for automatically increasing the gain of the early-stage spatiotemporal filters when the stimulus is small relative to the size of the filters.

An overall summary of these stages will be provided in the Discussion section and each stage will be linked to the particular primate motion sensitive neurons that they are designed to emulate.

Speed estimation

Figure 3 (blue line) shows replotted MT data from a single neuron tested by Maunsell and Van Essen (1983). The data set displays a typical tuning curve found in many MT neurons (Perrone & Thiele, 2001; Rodman & Albright, 1987); the response peaks at some optimal stimulus speed (4°/s in this case) and drops for speeds slower and faster than this value. The black curve shows re-plotted data from a single MST neuron collected by Inaba et al. (2007). In contrast to the MT neuron data, this MST neuron produces an output that



Figure 3. MT and MST neuron responses to a range of stimulus speeds. The Maunsell and Van Essen (1983) data is from their Figure 6a (4°/s unit). The Inaba et al. (2007) data is adapted from their Figure 2e (blue triangles, motion in preferred direction during fixation). It has been normalized relative to the peak response of the cell (approximately 60 spikes/s). In contrast to the MT cell, the MST neuron responds at a rate proportional to the test speed. Note that this is a log-linear plot with the x-axis based on $log_2(V)$. MT data set from "Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation" by J.H. Maunsell & D.C.J. Van Essen, 1983, J. Neurophysiol. 49, 1127-1147. Copyright 1983, The American Physiological Society. Adapted with permission. MST data set from "MST Neurons Code for Visual Motion in Space Independent of Pursuit Eye Movements" by N. Inaba, S. Shinomoto, S. Yamane, A. Takemura, & K. Kawano, 2007, 97, 3473-3483. Copyright 2007, The American Physiological Society. Adapted with permission.

continues to increase as the stimulus speed increases. On a log-linear plot, the relationship between the stimulus input speed and the cell response is surprisingly linear. It is this transformation from the speedtuned response of MT neurons to the linear response of MST neurons that I am attempting to replicate in the new velocity code model.

One reason for attempting to match the MST loglinear function in Figure 3 is that I would like to maintain concordance with the MST data. In addition we have previously presented an argument for a logarithmic sampling of speed based on the distribution of speeds that occur during self-motion through the environment (Perrone & Stone, 1994). I now outline how this MST property can be generated from small sets of MT neurons.

MT model units

The details of the stages leading up to the MT model units have been presented previously (Perrone, 2004, 2005; Perrone & Thiele, 2002). The spatiotemporal filters at the initial filtering stage of the model are based on the temporal (Foster et al., 1985; Hawken et al., 1996) and spatial (Hawken & Parker, 1987) frequency tuning of V1 neurons. One class of spatiotemporal filters are referred to as *sustained* (S) because of their low-pass temporal frequency tuning and the other we refer to as *transient* (T) because they have band-pass temporal frequency tuning.

The combined spatiotemporal amplitude response functions for the S and T filters are shown in Figure 1a in outline (contour) form. The S and T sf and tf functions are specially designed so that the locus of intersection of the two S and T amplitude response functions falls on an oriented line (with a zero intercept) in a sf-tf plot such as that shown in Figure 1a. In brief, the weighted intersection mechanism (WIM) model proposed by Perrone and Thiele (2002) was designed to produce the maximum output from a combination of the two S and T filter inputs whenever their output was both high and equal. This occurs along the oriented locus of intersection and thus speed tuned motion sensors are created with oriented spectral receptive fields similar to the plots shown in Figure 2a and b.

Here I use a slightly modified version of the original WIM model equation:

$$WIM = (S' + T')/(|S' - T'| + \delta),$$
(1)

where δ is the delta term used in the original equation and which controls the bandwidth of the speed tuning of the WIM sensor. It was set to a value of 8.0 in all of the simulations reported in this paper.

We have dropped the logarithm terms used in the original Perrone and Thiele equation because we have now introduced a contrast gain control stage at the point that the spatiotemporal energy is calculated. The spatiotemporal energy (Adelson & Bergen, 1985; Watson & Ahumada, 1985) found from the spatiotemporal filters (S and T) is modified by the following operations:

$$S' = aS/(pS + sc) \tag{2}$$

$$T' = aT/(pT + tc) \tag{3}$$

where a = 6.8, p = .06, sc = .15 and tc = 0.14 (for S and T values in the range 0–60).

This is a form of divisive negative feedback and the resulting transformation produces similar saturating contrast sensitivity functions to that introduced by the use of logarithms in the original WIM equation and which match the contrast response functions of MT neurons (Sclar, Maunsell, & Lennie, 1990).

The WIM model provides an account of how the primate visual system transforms early stage V1 neurons with one quadrant separable spectral receptive fields (see contour lines in Figure 1a) into neurons with inseparable or oriented spectral receptive fields (Figure 2a). These model WIM sensors act as subunits (Figure 2c and d) in our model of MT pattern neurons (Perrone, 2004; Perrone & Krauzlis, 2008a) which form an integral part of my new velocity code model. I also use model MT component units (MTc) based on MT component neurons (Albright, 1984; Movshon et al., 1983).

Basic pattern motion detector (PMD) and component unit design

Each cluster within a PMD is made up from seven positive WIM subunits and five inhibitory (opponent) subunits (Figure 2d). Their direction tuning ranges from $0-330^{\circ}$ in 30° steps. The speed tuning is a cosine function of the difference between the direction tuning value and the optimum overall direction tuning (θ) for the PMD. Therefore, if the overall velocity tuning of the PMD is $\bar{V}_p = (V_p, \theta_p)$, then the speed tuning of the cluster subunits making up the detector in the model is given by $s_i = V_p \cos(\theta - \beta_i)$ where β_i ranges from 0° to 330° in 30° steps. The set of β_i values was designed to sample the range of possible edge orientations that could be present in the receptive field of the PMD and it sets up the speed tuning of each cluster subunit to match the expected speed of the different possible edge configurations. The clusters are spatially separated in a circular array (Figure 2d). The radial separation distance between clusters depends on the spatial frequency tuning (u_0) of the WIM subunits and was set at $12/u_0$ pixels.

The different (s_i, β_i) tuned WIM subunits are weighted prior to their output being summed across all of the clusters. For β values $\pm 30^{\circ}$ on either side of $\theta_{\rm P}$, w_i = 0.87; for values $\pm 60^{\circ}$, w_i = 0.5 and for $\pm 90^{\circ}$, w_i = 0. Subunits that are tuned to directions within the range $\theta_{\rm P}$ - 180° ± 60° (dashed lines in Figure 2d) contribute in an opponent fashion (w = -1.0) and their output is subtracted from the total activity generated in the direction clusters; the output from the opponent units in a cluster is subtracted from the positive activity in the same cluster. The net local output (cluster positive activity - cluster negative activity) from all of the nine clusters in the receptive field is half-wave rectified, then summed (Perrone & Krauzlis, 2008a). This total activity (at frame four of an eight frame sequence) represents the output of the MT model pattern detector and is considered to be the equivalent of the average firing rate (spikes/second) generated by primate MT pattern neurons.

The model component units operate in a similar fashion to the PMDs but do not have the off-axis WIM units tuned to $\theta \pm 30^{\circ}$ and $\theta \pm 60^{\circ}$. Their total output is based solely on the net activity from the θ and $\theta + 180^{\circ}$ WIM units across the nine patch locations.

MT sampling array in spatio-temporal frequency space

Figure 4a is a spatiotemporal frequency space representation with log-log axes. In this type of log plot, the spectra for different edge velocities all fall on parallel lines (shown by the blue lines). The plot only shows the MT unit array for the primary direction (0° in this case). The spectral receptive field (small inset in Figure 4a) indicates the spatiotemporal frequency tuning of a representative MT model sensor.

When an edge moves at 2° /s to the right, it has a spectrum that falls on the 2° /s line in this type of plot (see pink solid line in Figure 4a). The task of estimating the speed of the edge amounts to locating the spectrum along the spatial frequency axis in this log-log plot. To this end, I use a set of five MT pattern units with spatial frequency tunings that span this space. I use units with peak spatial frequency tuning corresponding to 4, 2, 1, and .5 c/°. Four of the units have a peak temporal frequency of 4 Hz (Perrone, 2005), and one has a peak tf frequency of 8 Hz. This set corresponds to speed tunings of 1, 2, 4, 8, and 16° /s.

Because I am using digital image sequences for my simulations I actually test the code using pixels/frame as a speed measure but, for convenience and to retain consistency with the empirical data I simulate, I will report all speeds as degrees per second and assume that the temporal sampling of the input movies is 30 Hz and that a 256 pixel wide image subtends 8.5° such that 1 p/f = 1°/s. The constraints of digital image sampling and the time required to process large images dictate the range of speeds I have chosen to incorporate into the model. With larger input images, the tuning speeds could easily be scaled up by a factor of four to bring them more in line with typical MT optimum speed tuning values (Maunsell & Van Essen, 1983; A. T. Smith & Snowden, 1994).

Consider first a speed estimation system that simply uses the outputs from the full set of MT pattern units shown in Figure 4a. Figure 4b (black solid curve) shows the distribution of activity across the set of five MT units when stimulated with an image sequence ($128 \times 128 \times 8$ frames) of a high contrast (100%) edge moving at 2°/s to the right (0°). The set of MT units was centered on the middle of the image (x = 64, y = 64). The distribution peaks at 2°/s and the form of the curve reflects the underlying speed tuning curves of the MT units.



Figure 4. Basic velocity code and spatial scale problem. (a) Spatiotemporal frequency plot on log-log axes. Blue lines represent the locations for the spectra (e.g., red line) generated by moving edges of different speeds (given by labels at top). The ovals represent the spectral receptive fields of model MT pattern neurons (see inset) tuned to a range of spatial and temporal frequencies. (b) Distribution of outputs from set of five MT units located at center of image (solid line) and 24 pixels to the right of center (dashed line). (c) Possible spatial sampling scheme for MT units in basic velocity code array. (d) Output of a basic weighted vector average (centroid) speed estimation scheme. The actual edge speed was 2°/s (dashed line). Large errors occur for edge locations away from the center of the MT unit array.

A number of strategies have been suggested for how to obtain the speed from an MT population distribution such as the one shown in Figure 4b. One possible technique is to locate the peak of the distribution using some form of winner-takes-all technique (e.g., Perrone, 1992; Perrone & Stone, 1994). Another approach is to use the weighted vector average technique which is a form of centroid estimation (Bracewell, 1978). This technique has been used extensively as the basis for coding a number of stimulus dimensions in a range of biological systems (Dayan & Abbott, 2001; Georgopoulos, Schwartz, & Kettner, 1986) and it has also been used in previous attempts to code image velocity using MT neurons (Churchland & Lisberger, 2001; Krekelberg et al., 2006; Lisberger & Movshon, 1999; Priebe & Lisberger, 2004).

For a distribution such as that shown in Figure 4b the centroid is given by:

$$V_{R} = \frac{\sum_{i=1}^{5} w_{i} M T_{i}}{\sum_{i=1}^{5} M T_{i}}$$
(4)

where V_R indicates a velocity estimate based on the raw MT output values (MT_i) and w is the speed tuning (1, 2° /s etc.) of the MT units. For the 2° /s distribution (black solid line in Figure 4b), the value of V comes out at 4.02°/s which is an overestimate. On a linear x-axis the distribution shown in Figure 4b is highly skewed with a long tail at high speeds and this distorts the speed estimate. The centroid estimate can be improved by using $\log_2(V_i)$ for the w_i values. When this is done the estimate is 2.92°/s which is still inaccurate. This lack of accuracy is not the only reason I decided not to adopt the weighted vector average technique in this basic form or to use a winner-takes-all (peak detection) approach in my velocity code model. Both of these basic methods have a serious limitation: they only work if all of the MT neurons in the set MT_i are centered on the same image location (x, y). These basic techniques do not take into account the problem of spatial scale and how the MT neuron receptive fields sample and tile the visual image.

Problems with spatial scale

The use of multiple spatial channels is obviously beneficial when it comes to locating the edge spectrum along the spatial frequency axis (Figure 4a). However trying to compare the outputs of different sized sensors (spatial scales) introduces a problem. The receptive fields of the different units shown in Figure 4a span a wide range of sizes; the units tuned to $8^{\circ}/s$ are considerably larger than those tuned to 1° /s. In order to sample the visual field with these different sized units and to minimize overlap, the sampling must necessarily be sparser for the larger units than the smaller units. An example of a possible sampling strategy is shown in Figure 4c with the smallest units $(1^{\circ}/s)$ shown in blue at the center and the largest units (8 and $16^{\circ}/s$) shown in green. Only two of the smaller blue and black units have been depicted in the outer regions of the green circle for clarity but it should be obvious that there are regions that are represented by the center of the small (blue and black) units but not by the center of the larger units (green).

The level of output of an MT unit is determined to a large degree by the proximity of the stimulus to the center of its receptive field (e.g., see Xiao, Raiguel, Marcar, & Orban, 1997) and so the opportunity arises for a confounding of stimulus location and speed. The distribution of MT outputs shown in Figure 4b (solid black curve) is derived from a test in which the moving edge in frame four (of an eight frame sequence) is located exactly over the center of the MT spatial array (line A in Figure 4c). Only in this case does the distribution reflect the influence of the speed of the stimulus independent of the location because all of the MT units are being maximally stimulated at the center of their receptive fields. For other edge locations the relative responses depend on both the speed of the edge and its location relative to the center of the receptive fields.

The dashed line in Figure 4b is the MT activity distribution for the array centered on location A but the edge (still moving at $2^{\circ}/s$) is positioned 24 pixels to the right of the center of the array (see line B). The distribution is now skewed to the right and the estimate for the velocity using Equation 4 is greatly overestimated as being 6.5° /s. Even though the edge speed is not optimal for the larger MT units, they still end up responding more than the smaller units, simply because of their larger footprint. An edge at location B still activates the large (green) units but barely triggers any response in the small (blue and black) units at the center of the array. The resulting skewed distribution leads to a large overestimation of the edge speed. Figure 4d shows the velocity estimates using the log scaled version of the centroid equation (Equation 4) for a collection of five MT model units located at the center of the image for different edge locations. The speed estimation error increases rapidly for locations away from the central location of the MT unit receptive field array. A peak detection or winner-takes-all scheme also suffers from this type of error.

New improved velocity code model

The above speed estimation problems stem from the fact that the speed estimation process is dependent on using the outputs from a broad range of MT neuron sizes (spatial scales). The rationale and motivation for many of the design elements in the new velocity code model arise from the constraint that a broad range of MT neuron sizes cannot be used to determine image speed because the cell output is a function of both the stimulus's speed and its location in the receptive field. A wide range of MT receptive field sizes exacerbates this potential confound (Figure 4c) and also leads to biases from truncation effects (Krekelberg et al., 2006).

The use of a threshold to limit the signals from the erroneous larger MT units (Chey, Grossberg, & Mingolla, 1998) makes it difficult to detect low-contrast moving stimuli. Therefore my solution is to minimize the range of MT optimum speeds (sizes) that go into the speed computation. I introduce the concept of a velocity channel whereby only two MT spatial fre-



Figure 5. New velocity code MT unit array in log-log frequency space. (a) The basic array is augmented with four MT component units tuned to half the temporal frequency of the pattern units. The solid ovals connected by shaded lines represent the triad of units making up a velocity channel, in this case one tuned to 2°/s. (b) Red ovals represent overclocked component units also included in the new velocity code model array. The black ovals depict the original tuning of the red units and are not part of the new model array.

quency scales are used in the centroid estimation stage and I use inhibition between channels to overcome the spatial scale problem described above (Figure 4b and d).

In addition to the basic set of five pattern units, I add a row of component units (MTc) that are tuned to half the temporal frequency of the pattern units (Figure 5a). The component units are based on MT component neurons (Albright, 1984; Movshon et al., 1983) and have the same basic speed tuning mechanism as the pattern units but lack the off-axis WIM subunits that make up the model MT pattern units (Perrone, 2004; Perrone & Krauzlis, 2008a). They respond predominantly to the motion component orthogonal to the orientation of the edge. Because their peak temporal frequency tuning is at 2 Hz, they respond optimally to the same speed of edge motion as the pattern unit (tuned to 4 Hz) located along the same iso-velocity line (blue parallel lines in Figure 5a).

In the new model there is also another row of MT component units (red ovals in Figure 5b) that are tuned to 4 Hz temporal frequency. However these four units are derived from a basic set of 2 Hz component units similar to those in Figure 5a but they have had their tf tuning pushed out to 4 Hz by a reweighting of the sustained and transient inputs to the WIM subunits making up the MT component units (see Perrone, 2005). I will refer to this set as *overclocked* component units because they are tuned to twice their natural temporal frequency. The black ovals in Figure 5b are not part of the velocity encoding set but simply represent the original tuning of the overclocked units. The full ensemble of MT units consists of five MT pattern units and eight component units (made up of four standard units and four overclocked units). The overclocked component units are used as part of a contrast sensitive redundancy control mechanism and do not figure directly in the velocity code model. They will be discussed in more detail below when I introduce spatial interactions between MT units.

Instead of using the output from all of the 13 MT units in the centroid calculation I use only a triad of MT units consisting of two pattern units and one component unit: MT_v , MT_{2v} , and MTc_V . Since MTc_V is tuned to .5 V the set is equivalent to MT_{2V} , MT_{V} , and MT_{.5V}. I will refer to these three units as a velocity channel. The main tuning of the channel is determined by the MT_V pattern unit (e.g., $2^{\circ}/s$). The channel also includes a pattern unit tuned to twice the speed as the primary unit and a component unit with the same spatial frequency as the primary unit but tuned to half the speed. The fact that each channel is constrained to just two spatial scales alleviates the spatial scale problem encountered above when all four spatial scales were used to determine the centroid of the MT distribution.

A typical triad making up a 2° /s velocity channel is shown in frequency space in Figure 5a (see solid line ovals connected by gray lines). The spectrum for a moving edge can be conceptualized as a ridge aligned with one of the blue lines in Figure 5a. The three MT units making up a velocity channel straddle this ridge in a symmetrical fashion. Similar velocity channels exist in my model (but not shown) tuned to 1, 4, and 8°/s. The 8°/s channel requires a slightly different arrangement and uses the MT units located along the vertical column to the left of the figure. Note that there is a degree of overlap in the channel structure; an MT unit can act as both the primary unit for a channel as well as one of the secondary units making up the triad in an adjacent channel.



Figure 6. New velocity code model in operation in response to a 2°/s edge. (a) Output of triads making up each of the four velocity channels. Color coding relates to the sizes of the units in the spatial receptive field array shown in Figure 4c. The peak response occurs in the channel two primary (central) unit and the distribution of responses across the three units in the channel determines the speed (Equation 7). (b) Output of second derivative stage of new code that limits the velocity output signal to the channel containing the peak central response. For the 2°/s test case, only channel two produces a positive multiplicative gain signal and so only this channel generates a velocity signal.

Figure 6a shows the output of the three MT triad units making up each of the four channels in response to an edge moving at 2°/s and located such that, at frame four of the eight frame sequence, it is positioned directly above (x, y) as in Figure 4c (line A). The central lines represent the output from the primary pattern unit (MT_V) and the outer two lines are from the MTc units and the MT_{2V} units. The color code continues the convention adopted in Figure 4 with blue representing the highest spatial frequency channel (smallest circles in Figure 4c) and black, red, and green the other sizes.

The most active primary unit is the one in the $2^{\circ}/s$ channel. An estimate of the centroid of the triad of responses in the $2^{\circ}/s$ channel would produce a good estimate of the edge speed (note the symmetry of the responses). However the other channels would produce erroneous estimates of the centroid (and hence the speed). Therefore the new velocity code model uses a centroid calculation on the triad of MT responses within a channel but it also includes a mechanism for constraining the velocity estimates to the channel containing the peak response in the primary unit.

New velocity code model mechanism (Part 1. 2nd derivative stage)

The first stage of the new velocity code model is to introduce a new mechanism (P) that sums the input

from the primary MT unit in the channel triad but which is inhibited by the MT_{2V} and MTc_V units. The reason for this inhibition is two-fold: (a) It limits the output to the channel that is responding the most and (b) It overcomes the spatial scale problem discussed above. Specifically we calculate for each channel V_i:

$$P_{vi} = f^+ (MT_{vi} - .5MT_{2vi} - .5MTc_{vi})$$
⁽⁵⁾

Where

$$f^{+} = \max(f(x), 0) = \begin{cases} f(x) & \text{if } f(x) > 0\\ 0 & \text{otherwise} \end{cases}$$

Given that the optimum speed tuning of an MTc unit is half that of an MT unit of the same spatial scale (= $MT_{.5V}$) it should now be obvious that this new mechanism is the equivalent of finding the second derivative of the MT output distribution. For typical MT speed tuning curves (Figure 3) and distributions (Figure 4b) this stage only produces an output at the peak of the distribution (i.e., when MT_{Vi} is the most active unit). Figure 6b shows the normalized output (see below) of the P mechanism stage for each of the different channels in response to the 2°/s edge stimulus. All channels except the 2°/s one produced values of zero in response to the input stimulus. I use the output of the *P* stage to control which velocity channel is going to produce a signal using a form of multiplicative gain

control. In order to constrain the range of responses of the *P* mechanism to a narrow band of mainly "on" or "off" responses we use the same form of divisive gain control adopted in Equations 3 and 4 and transform the P_{Vi} values using:

$$GP = aP/(pP + gc) \tag{6}$$

where a = 5, p = 0.2 and gc = 0.1.

This transformation produces a log-type function that rises quickly as the P_{Vi} value increases but which saturates and produces very similar output (~ = 5.0) across a wide range of P_{Vi} values. For regions of the MT distribution (Figure 4b) away from the peak, *GP* has a value of 0.0 when half-wave rectified. Velocity channels that do not contain the peak response in their primary unit (MT_V) will therefore have zero gain (Figure 6b) and will not generate a velocity estimate through the mechanism to be outlined in the second stage of the model (see below).

The second benefit of introducing the *P* stage mechanism is that it helps overcome the spatial problem (see section titled Problems with spatial scale) that occurs when a wide range of MT spatial sizes are used to determine the image speed. For example, a 4°/s MT pattern unit may respond because of its large size to an edge moving at 2°/s that is not located centrally above the MT array (see Figure 4c) but it will be inhibited by the large response from the MTc₄ unit which covers the same area but is tuned to 2°/s. For the 4°/s unit, the output of Equation 5 $[P_4 = f^+ (MT_4 - .5MT_8 - .5MTc_4)]$ will therefore be zero.

The above stage of the new velocity code model simply controls which channel will respond but does not actually produce a velocity estimate. The *GP* mechanism does not produce an output proportional to the speed of the input. It could be used in a system that finds the peak in the MT distribution but the resolution of the velocity estimates would be limited to the number of channels (1, 2, 4, or $8^{\circ}/s$). The addition of more channels would improve the resolution but this is not as economical as the system I have developed which is able to interpolate the speed values between the main tunings of the velocity channels.

New velocity code model mechanism (Part 2. Centroid stage)

For each velocity channel we estimate the velocity using the same centroid calculation given in Equation 4 but only the outputs from the three units making up the channel are incorporated into the centroid calculation:

$$C_{i} = \frac{w1MT_{Vi} + w2MT_{2Vi} + w3MTc_{Vi}}{MT_{Vi} + MT_{2Vi} + MTc_{Vi}}$$
(7)

The \log_2 sampling used for the MT channels (Figure 5) helps make the distribution of activity across the triad

of MT units more symmetrical (see Channel 2 in Figure 6a) and it improves the velocity estimates. The weights in Equation 7 are set to generate values of C_i that follow a linear function similar to the MST data (Figure 3). The shape of the speed tuning curves of the individual MT units determines the symmetry of the output distribution across the three units making up the triad. In order to calibrate the system, the weights can be customized to accommodate the different distributions and to optimize the performance of the velocity code model. The weights were set to produce a linear output function such that $C_i = 20 + 20\log_2(V_i)$ for a range of input values ($V_i = .5 - 8^{\circ}$ /s).

The final step in the new velocity code model is to multiply the output of the centroid estimator (Equation 7) for each channel (C_i) by the GP_i gain term for that channel (Stage 1, Equation 6). Therefore the output of the model for the four velocity channels (i = 1, 2, 4, and 8° /s) is given by:

$$V_i = aGP_iC_i \tag{8}$$

where a is an arbitrary scaling value which controls the size of the velocity output. It was set to 0.2 in the model simulations to produce values for V which peak at 80.0 for our highest velocity channel.

One neural mechanism (Equation 7) results in an output proportional to the centroid location of the triad of MT neurons and another (Equation 6) responds only if the channel contains the peak MT output. The latter mechanism controls the gain of the former and only the centroid estimate from the peak channel is output to the next stage of motion processing.

For the Figure 6 example, only the 2°/s channel had a nonzero *GP* value (4.93) and the centroid estimate (*Ci*) for this channel was 40.53. When the V_i output (39.97) is converted to °/s (see Equation 10) it corresponds to a speed estimate of 1.99°/s which is very accurate.

For an edge speed that falls between the boundaries of the channels (e.g., $1.5^{\circ}/s$) it is possible for two channels to have positive values of *GP* and for each channel to generate a velocity signal. These estimates are usually very similar, if not identical. When calculating the final velocity for assessing the accuracy of the code in this paper I will report the average of the two responses. However the V_i values are seen as the first part of a global motion processing stage which combines many such estimates across the visual field and I will therefore assume that if two channels are active, both of the estimates are passed onto the next stage of neural processing (MST) and any averaging occurs there.

Therefore the final velocity estimate from a mechanism I will label *VMST* is given by: Journal of Vision (2012) 12(8):1, 1-31

$$VMST = V_1 + V_2 + V_4 + V_8 \tag{9}$$

Perrone

The VMST value represents a signal proportional to the stimulus velocity at a particular image location (x, x)v) and many such signals would be summed across the visual field by a unit designed to detect global patterns of image motion such as an MST neuron. Therefore the *VMST* stage is considered to be the local input at a particular image location feeding into an MST neuron that has many such inputs distributed over large regions of the image. When tested with a uniform field of image motion, the MST neuron's response would be a scaled version of the output from the local VMST mechanism. Therefore the VMST output in response to a range of speeds should emulate the behavior of MST neurons such as that shown in Figure 3. In order to test the accuracy of the model and to relate the output to the input, we can derive the actual velocity from the *VMST* firing rate using:

$$VT = 2^{\left(\frac{VMST-20}{20}\right)}$$
(10)

This is not intended to represent a particular neural stage or computation and is only used to provide a comparison of the model estimates to the input values.

Methods

The new velocity code model was tested using edges moving left to right at a range of speeds (.5, .75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 3.5, 4, 5, 6, 7, and 8°/s). The edge contrast was 100% and the image size was 128 \times 128 pixels and eight frames in length. The output of the model at location (64, 64) was recorded at frame four of the sequence. Note that the range of speeds tested is a function of the image size used for the tests and the assumed field of view of the image. The range of speeds can be arbitrarily scaled upwards by increasing the image size.

Results

Figure 7 shows the non-transformed *VMST* output (Equation 9) of the new velocity code model for the tested edge speeds. The output is very linear on a log-linear plot and emulates the behavior of the Inaba et al. (2007) MST neurons (Figure 3). This linearity occurs for input speed values that do not necessarily coincide with the tuning of the velocity channels used in the model. It is clear that the new velocity code model is able to interpolate and produce an output in proportion to the edge velocity even for cases in which the input speed falls between the four channel values. The crosses in the plot are for the cases in which the test speed matches one of the velocity channels. The circles



Figure 7. Test of the new velocity code model. The graph shows the output of the VMST stage (Equation 9) in response to an edge moving left to right at a range of speeds. The crosses are for test speeds that match the tuning of one of the velocity channels, the filled circles are for tests speeds not aligned with the channel tuning. The *x*-axis is $log_2(V)$ but the *y*-axis is linear. The dashed line is generated from $Y = 20 + 20log_2V$ and represents perfect linearity on a log-linear plot. On a linear-linear plot the data fall on a log-type function which asymptotes at high speeds similar to the behavior of some MST neurons (Inaba et al., 2007).

are for test speeds that do not match the primary tuning of the channel yet the new code continues to output values log-linearly related to the input speed.

I have achieved the goal of been able to generate an output from sets of MT neurons that is log-linear in form, similar to some MST neurons (Inaba et al., 2007). What is not apparent in the test example shown in Figure 7 is how well the new code deals with the spatial scale issue that caused problems for a basic centroid (weighted vector average) method or peak detection method when a broad range of MT spatial scales are used to derive the speed estimates (Figure 4d). The Figure 7 tests were generated at a single image location from velocity channels centered on the edge location in the middle of the movie sequence (frame 4). To examine what is happening at other channel locations and directions I need to introduce an extended spatial sampling scheme for the MT unit locations that tiles a larger part of the image. The scheme I have adopted is shown in Figure 8.

The spatial sampling is based on a rotated rectangular array of image locations (x_i, y_i) . The distances between the elements in the original (nonrotated) array



Figure 8. Spatial sampling scheme for MT pattern and component units. (a) Original rectangular array with blue circles corresponding to MT_V unit locations and the red circles the MT_{2V} unit locations. The spacing (d) is scaled depending on the speed tuning of the units. (b). Final diamond lattice array rotated by 45°. An individual velocity channel (tuned to rightwards motion) is represented by the darker red and blue units. Note that adjacent channels share common MT units along their borders (e.g., unit 3 shared by units 1 & 10).

(Figure 8a) is proportional to the speed tuning of the velocity channel: (7, 14, 28, and 56 pixels) for channel 1, 2, 4, and 8°/s, respectively. If Figure 8 represents a 2°/s velocity channel, then the spacing of the 2° /s units (blue circles) is 14 pixels and the spacing of the 4° /s units (red circles) is 28 pixels.

Rather than using the array in its original rectangular layout, I have found a 45° rotated version to work better and it provides a better explanation for the patterns of antagonistic spatial inhibition found in MT neurons (Xiao et al., 1997). When rotated, a single velocity channel tuned to rightwards motion is now represented by the central diamond lattice configuration in Figure 8b. All of the eight surrounding MT units (and one located centrally) are all tuned to V and represent the primary unit of a velocity channel triad. These blue locations are also occupied by the MTc_V units making up the triads. The central location in the diamond (marked with a filled red circle) is occupied by the larger unit of a velocity channel triad (MT_{2V}) and its receptive field covers each of the MT_V and MT_V locations.

The application of the velocity code equations (Equations 5 and 7) is straightforward for the central location of the diamond lattice array (red circle in Figure 8b) and follows the procedure outlined above: the MT_V unit at the central location is inhibited only by the MT_{2V} and MT_{CV} units at the same location. However for the outside positions in the diamond lattice array, the MT_V units are inhibited not only by the MT_{2V} unit at the center of the array but also by the MT_{2V} units at the center of adjacent diamond arrays. For example, for the unit 3 in Figure 8b, Equation 5 becomes:

$$P_{V3} = f^+[MT_{V3} - .5(.MT_{2V1} + MT_{2V10}) - .5MTc_{V3}]$$

where the numbers after the V symbol indicate the location specified in Figure 8b. For a P_V unit at location 2, the inhibition comes from the MT_{2V} units at all four red locations surrounding it.

Although it complicates the triad calculations (Equations 5 and 7), this sharing of the outside MT_V units by the central MT_{2V} units enables adjacent diamond arrays to be closely packed together without overlap and it provides an economical sampling of the 2D image. In order to explore how well the new velocity code deals with the spatial problem outlined above, the model was run at a range of image locations specified by the diamond array (Figure 8b) covering a 128 × 128 pixel image area and using a range of different directions for the velocity sensors (0–330° in 30° steps). To better visualize the output of the model, the firing rate of the *VMST* stage has been converted to °/s values using Equation 10 and has been plotted in the form of vectors.

Figure 9a is a vector plot showing the transformed outputs of the different velocity detectors across space and across a range of angles in response to an edge moving at 2°/s to the right. The edge extended the full 128 pixel height of the image and was located at x = 64 in frame 4 of the eight frame movie sequence.

Compared to the basic velocity code model test shown in Figure 4d, the velocity estimates are constrained to locations close to the edge location and there is no evidence of the high velocity overestimation errors at distant locations that were evident in the Figure 4d test. Although the spatial scale problem has been overcome, another problem has manifested itself in this test: The *VMST* velocity estimators tuned to other directions besides 0° are responding to the edge and signaling a higher speed in that direction. My solution to this direction problem is to make use of the



Figure 9. Results of test of new velocity code model produced using an edge moving at 2° /s to the right (image size = 128×128 pixels). Vector plot showing output of velocity stage transformed to $^{\circ}$ /s (Equation 10). Vectors have been scaled up (×4) to make them easier to see. The gray area indicates the position of the edge in the middle of the movie sequence. (a) Output without direction inhibition mechanism. (b) Application of direction inhibition mechanism.

signals from the overclocked component units also present in the model array (Figure 5b). These units are tuned to the same speeds as the pattern units that form the primary elements in the velocity estimation channels but they respond primarily to motion 90° to the edge orientation. They enable us to constrain the velocity estimates to directions orthogonal to the edge and to remove the spurious outputs visible in the Figure 9a vector plot.

Direction estimation

The broad directional tuning of MT pattern neurons means that an edge moving in the 0° direction will activate an MT model pattern unit tuned to 0° but it will also stimulate MT pattern units tuned to directions on either side of 0° (Perrone, 2004; Perrone & Krauzlis, 2008a). An MT unit tuned to 4°/s and 60° direction will be well stimulated by an edge moving at $2^{\circ}/s$ in a 0° direction because the 60° MT pattern unit contains a WIM subunit tuned to 4 $\cos(60^\circ) = 2^\circ/\text{s}$ in the 0° direction. Therefore for velocity channels tuned to directions $\pm 60^{\circ}$ and $\pm 30^{\circ}$ relative to the edge direction the channel triad MT distributions will be skewed to higher speeds and the VMST units will estimate higher values for the speed in these directions as can be seen in the Figure 9a vector plot. These additional estimates of the velocity of the moving edge add noise to the next stage of motion processing (MST) that is concerned with finding the overall direction and speed of the edge. We would like to constrain the estimates to mainly lie in the direction normal to the edge (0° in this case).

In the new model, a moving edge is characterized not only by the responses of the model MT pattern neurons but by the fact that the overclocked component units (MToc) tuned to the direction orthogonal to the edge orientation also respond at a high rate. However it is not simply the case that if an MT pattern unit is responding in a particular direction (θ) without a corresponding response from the MToc unit tuned to θ , then one could conclude that the MT pattern response is spurious and should be eliminated. There are bona fide cases in which the pattern unit tuned to direction θ is responding strongly but there is no response from the MToc unit tuned to θ . Moving plaid patterns (the sum of two gratings with different orientations) create such a state of affairs for example; the MToc component units tuned to the plaid components do not align with the plaid direction or the MT pattern unit that responds maximally to the plaid. A characteristic of this situation however is that the directions of the active component units are symmetrically oriented around the pattern unit direction. I therefore introduce a mechanism that produces no inhibition when the MToc responses are symmetrically distributed (in terms of their directions) but which generates strong inhibition when an asymmetry occurs in the direction response profile.

For each angular direction θ_j , (where $j = 0-330^\circ$ in 30° steps) I define an inhibitory direction component DI such that:

$$DI_{\nu,\theta_j} = |MT_{oc}(V,\theta_j + 30^\circ) - MT_{oc}(V,\theta_j - 30^\circ)| + |MT_{oc}(V,\theta_j + 60^\circ) - MT_{oc}(V,\theta_j - 60^\circ)|$$
(11)

Where $MToc(V, \theta_j)$ is the output from an MT overclocked component unit tuned to speed V and direction θ_j . I use the absolute value of the differences for mathematical convenience and compactness but this operation could be implemented biologically by half-wave rectification and summing of the signals from $MT_{oc}(V,\theta_j + 30^\circ) - MT_{oc}(V,\theta_j - 30^\circ)$ and $MT_{oc}(V,\theta_j - 30^\circ) - MT_{oc}(V,\theta_j + 30^\circ)$ and their $\pm 60^\circ$ equivalents.

At the stage where the second derivative operation is applied (post MT) I subtract off the direction inhibition, DI_{V,θ_j} . Therefore Equation 5 can be modified to this form:

$$P(V_i, \theta_j) = f^+ \Big(MT(V_i, \theta_j) - .5MT(2V_i, \theta_j) - .5MTc(V_i, \theta_j) - DI_{V, \theta_j} \Big)$$
(12)

It should be apparent that if the direction inhibition is high, the value of P will be low and hence GP will be low (Equation 6) and this will turn off the velocity channel for direction θ_j and prevent the velocity signal from being output.

The operation of the direction inhibition stage is demonstrated in Figure 9b. All velocity outputs except those orthogonal to the edge direction have been removed. Although it has been presented first, the direction inhibition stage is actually implemented at a late stage of the velocity estimation process. The local inhibitory mechanisms between MT units described below occur prior to the direction inhibition stage. However it is easier to depict the local spatial inhibition effects without the presence of the erroneous direction vectors.

Contrast-dependent redundancy removal

The vector flow field plot (Figure 9b) reveals another problem with the basic velocity code model. Even though the outputs at the position of the edge are relatively accurate and in the correct direction, they are spatially redundant in the sense that the same signal is being generated by all the velocity sensors located along the edge. There are also signals being generated on either side of the actual edge location. If these signals were being used by a higher level system designed to extract global motion patterns by integrating the local estimates over wide regions of the visual field, (e.g., MST neurons designed to detect heading; Perrone, 1992; Perrone and Stone, 1994, 1998), the additional edge outputs simply add to the noise and make it more difficult to recover the correct rotation or heading signal. As was the case for the direction estimates, we would like to minimize the number of these redundant signals.

On the other hand, this redundancy problem is something that one would only want to correct at high contrast levels; when the stimulus contrast is low, all the velocity signals along the edge would be useful to the next stage of processing and should not be eliminated. This will require a mechanism that is sensitive to and can adjust itself according to the contrast level of the stimulus. At low contrast, any motion signals are useful to the later stages and should be retained. I therefore developed a spatial inhibitory mechanism that only works at high contrast levels. It also makes use of the overclocked component MT units (Figure 5b).

Contrast-dependent speed retuning

Perrone

The overclocked MT component units respond differently to contrast compared to the standard pattern and component units. The reason for this can be seen in the response to contrast of the sustained and transient V1 inputs to the WIM subunits that make up the MT sensors. In the WIM subunits that make up the overclocked MT units the T spatiotemporal energy is weighted by 0.5 relative to that of the S units in order to increase the peak temporal frequency tuning of the WIM unit (see Perrone, 2005). This means that the contrast gain control mechanism (Equations 3 & 4) produces a different output (T') for the T units compared to the S units. This is illustrated in Figure 10. In the standard units, the S' and T' responses as contrast changes are almost identical (Figure 10a). However in the WIM units feeding into the overclocked MT units, as contrast drops the S' and T' spatiotemporal energy outputs are no longer equal; some parts of the T' contrast response curve are higher than the S' curve (Figure 10c). The consequences of this can be seen in the temporal frequency tuning curves of the respective units (Figure 10b & d). These were generated by examining the S' and T' levels in response to a moving edge at a range of speeds. For a particular spatial frequency channel (2 $c/^{\circ}$ in this case) the speeds can be converted into temporal frequency using tf = 2V. For the standard units (Figure 10b), as contrast drops the relative contribution of the S' and T' remains the same and the point at which the S and T curves cross remains at 4 Hz (arrows in Figure 10b). However for the overclocked MT units, as contrast drops the relationship between the S' and T' inputs to the WIM stage does not remain the same. At low contrast the S' values fall by a greater amount compared to the T'values (see vertical dashed line in Figure 10c) and so the crossover point shifts to a lower temporal frequency (Figure 10d).

This has an important impact on the speed tuning of the WIM units and consequently the speed tuning of the MT units. As the contrast of the input stimulus drops, the preferred speed of an overclocked MT component unit is reduced to a slower speed. This speed retuning mechanism is a critical part of the spatial inhibition stage of the new velocity code model and so I now present evidence of a similar mechanism in MT neurons.



Figure 10. Contrast response curves (left) and temporal frequency tuning curves (right) for V1-stage model neurons when tested with a moving edge (2°/s) at two different contrast levels (100% and 10%). (a) Contrast response function for sustained (black line = S') and transient (gray line = T') standard model V1 spatiotemporal energy units feeding into the standard MT pattern and component units. (b) Temporal frequency tuning curves for standard units obtained from channel two (2 c/°). The cross-over point of the two temporal frequency tuning curves remains at the same temporal frequency (4 Hz) when the contrast of the edge drops. (c) Contrast response curves for V1 units feeding into overclocked MT component units in model. (d) Temporal frequency tuning curves for overclocked units. The crossover point shifts to lower temporal frequencies at low contrast. Therefore the peak speed tuning of the WIM units and overclocked MT units drops to lower values at low contrast.

Figure 11a shows replotted data from Krekelberg et al. (2006) for a single primate MT neuron in response to a range of random dot speeds at a range of contrast values (see also Pack et al., 2005). At high contrast, this cell preferred a speed of approximately 16°/s. As the contrast dropped, the peak tuning dropped to lower values and once the contrast dropped to 5%, the cell ended up tuned to approximately half the speed preferred at the 70% contrast level. Figure 11b is the response of one of our model overclocked MT component units tuned to an image speed of 4°/s. When tested with similar dot patterns to those used by Krekelberg et al., the MToc unit shows the same shift in peak tuning as the MT neuron when the contrast of the dots dropped. This shift can be directly attributed

to the contrast responses of the overclocked WIM units making up the MToc unit (Figure 10c and d) and the contrast mechanism outlined above. The Krekelberg et al. data show that there exist cells in the primate brain with a speed retuning feature that I will incorporate into the new velocity code model. Similar contrast related changes to spatiotemporal frequency tuning are also apparent in V1 data (Priebe et al., 2006) and have been modeled using our WIM model (Perrone, 2006).

Contrast-dependent spatial inhibition

Consider now a contrast-dependent mechanism that uses the outputs from sets of the MToc units. Figure



Figure 11. Changes in peak speed tuning with contrast. (a) Replotted data from Krekelberg, van Wezel, and Albright (2006) (their Figure 3a) showing changes in the peak speed tuning of an individual MT neuron as the contrast of the stimulus drops from 70% to 5%. (b) Output from a model overclocked MT component unit.

12a (top curve) shows the output from a single MToc unit tuned to 2°/s in response to a vertical edge of 100% contrast moving at 2°/s located at a range of x image locations relative to the center (x = 64) of the frame. This MToc₂ unit is located at the central location of the diamond array spatial patches shown in Figure 8b (red central circle). The red curve is the output from an MToc₄ unit at the same location as the MToc₂ unit. The red curve has been inverted to reflect its subtractive role in the new spatial inhibitory mechanism. Figure 12b is for the same set of units and speed but the edge contrast was only 20%.

I postulate a spatial inhibitory (SI) mechanism that outputs the half-rectified difference between the $MToc_2$ unit and the $MToc_4$ unit or more generally:

$$SI_v = f^+(MT_{ocV} - .5MT_{oc2V})$$
 (13)

The values of SI for 2°/s and at the location of the MToc unit shown in Figure 12a are plotted in Figure 12c. This is the output I will use for my spatial inhibitory mechanism. The point of interest is that for the low contrast case (Figure 12d), the value of SI is zero.

The reason why the *SI* values are positive at high contrast but zero at low contrast can be directly attributed the speed retuning behavior of the overclocked MToc units (Figure 11b). At high contrast, the MToc₂ units are tuned to 2° /s and so respond well to the moving 2° /s edge at all locations. The larger MToc₄ unit is tuned to 4° /s and so responds (approximately 50%) less to the 2° /s edge. This results in a positive value for the difference between the MToc₂ and (.5 weighted) MToc₄ unit ($SI_2 > 0$). At low contrast however, the MToc₂ unit now prefers a speed of 1°/s and the MToc₄ unit prefers a speed of 2°/s (see Figure 11b for this case). Now the smaller MToc₂ unit is responding less than the MToc₄ unit and so no output results from the *SI* units (Equation 13 & Figure 12d).

I capitalize on the spatial inhibitory behavior offered by the mechanism of Equation 13 to control the responses of the MT pattern units across space and hence limit the amount of redundant velocity signals being generated. Specific primary pattern units (MT_V) are inhibited within the diamond lattice spatial receptive field array making up a channel (see Figure 8b) using the output from the *SI* units located at the center of the array.

The activity of particular MT pattern neurons is reduced prior to the second derivative and centroid stages listed above. I use different patterns of spatial inhibition to achieve a number of different outcomes. The first type is shown in Figure 13 and it is intended to remove redundant velocity output along edges. In Figure 9b the velocity units along the edge are all outputting the velocity 90° to the edge orientation. Given that the center unit of each MT receptive field array is relaying this velocity value, it is redundant for the nearby units in the array above or below that location to also output the same velocity, particularly at high contrast.

Redundancy reduction 1: Edge thinning

Therefore my first class of spatial inhibition will be referred to as *edge thinning* and it is applied to the MT



Figure 12. Contrast-dependent inhibitory mechanism. (a) Output from the model MT overclocked units (MToc) in response to a high contrast (100%) edge moving at 2°/s and located at different positions relative to the diamond MT spatial array (see Figure 8b). The black curve is for an MToc unit located at the center of the array. The red curve is the output from an MToc unit at the same location but tuned to twice the speed (4°/s) (b) $MToc_2$ and $MToc_4$ outputs at low contrast. (c) Rectified difference between $MToc_2$ and .5 ($MToc_4$) units. (d) Difference at low contrast. No inhibitory signals are generated.

neurons feeding into the velocity channels, i.e., this inhibition is between MT neurons and is expected to manifest itself at the level of the MT units. The different possible cases are shown in Figure 13. To accommodate edges moving at 0°, the MT pattern unit tuned to 0° at the bottom of the diamond array (Figure 13a) is inhibited by an *SI* unit (made up of MToc units as per Figure 12) located at the center of the array (red dot). The positive MToc unit (black curve in Figure 12a) making up the *SI* unit is tuned to 0° and to the same speed as the MT pattern unit. The red arrow indicates the direction of the inhibition. The inhibition is intentionally designed to be asymmetric, because the diamond array is surrounded by other similar arrays and the top unit in the diamond array would be inhibited by the *SI* unit at the center of the array above it.

In a similar fashion, the MT unit tuned to 180° at the bottom of the diamond is inhibited by the *SI* unit tuned to 180° . Other possible directions of motion are catered for by the inhibition patterns shown in Figure 13b through d. Because of the contrast sensitive nature of the *SI* inhibitory units (Figure 12) this edge thinning operation will not be operational at low contrast.

Redundancy reduction 2: Spatial sharpening

In the test output shown in Figure 9b the velocity units on either side of the moving edge produce a



Figure 13. Patterns of spatial inhibition used to remove redundant velocity signals at high contrast. The diamond array of dots represents the spatial layout of the primary MT units within a velocity channel (see Figure 8b). This class of inhibition is designed to thin the velocity signals along edges. The red arrow indicates the MT pattern unit inhibited by the SI component units (red circles). The black line indicates the edge location that triggers the most inhibition. (a) Inhibition for MT units tuned to 0 and 180° (b) Inhibition for 90 and 270° MT units. (c) Inhibition for MT units tuned to 120, 150, 300, and 330° (d) Inhibition for MT units tuned to 30, 60, 210, and 240°

similar output to the central unit. The spatial resolution of the velocity vectors is unnecessarily broad in the direction of motion of the edge (0° in this case). The units to the left and right of center are generating redundant signals given that the central units are responding to the edge. This redundancy could be reduced at high contrast by inhibiting the MT pattern neurons in the outside regions of the diamond array (units 3, 5, 7, and 9 in Figure 8b) with activity from SI units located at the center of the array (see Figure 14a). The units to the far left and right (two and six in Figure 8b) do not need to be inhibited because their output is very low when the edge is located over the central unit (this is one of the advantages of using a diamond rather than a square lattice). For other edge locations, the two and six units are inhibited by the edge thinning operation (outlined in the section Redundancy reduction 1: Edge thinning) from the red units above them.

Once again, at low contrast the *SI* units at the center of the array (red circles) will have zero output (Equation 13) and the patterns of inhibition shown in Figure 14 will be turned off.

Tests of the redundancy reduction inhibitory mechanisms

Methods

In order to demonstrate the above mechanisms, tests were carried out using a larger image size $(256 \times 256 \text{ pixels})$ and the diamond lattice sampling array described above (Figure 8). An edge moving at 2°/s moved left to right and two edge contrasts were used (100% and 20%).

Results

Figure 15a shows the output of the model in response to the 100% contrast edge without the spatial inhibition mechanism. Only the direction inhibition (Equation 11) has been applied at this stage and just the signals orthogonal to the edge remain. Figure 14b shows the output when the spatial redundancy mechanisms (sections Redundancy reduction 1: Edge thinning and Redundancy reduction 2: Spatial sharpening) were applied. The velocity is mainly being reported in the central location of each of the MT diamond shaped



Figure 14. Patterns of inhibition used for improving the spatial resolution of the velocity signals. (a) Case for MT units tuned to 0 and 180°. (b) Inhibition for 90 and 270° (c) Inhibition for 120, 150, 300, 330° MT units. (d) Inhibition for 30, 60, 210, and 240° MT units

a b c c

Figure 15. Contrast-dependent spatial inhibition test results.(a) Output of velocity code model with direction inhibition mechanisms but without the spatial inhibitory mechanisms. Vectors represent transformed output (Equation 11) scaled by a factor of five to make them easier to see. The same velocity signal is being output at many locations along the edge and on either side of the edge resulting in high levels of redundancy and poor spatial resolution. (b) Output of the model when both direction and spatial mechanisms are applied. The top vector is not removed because of edge effects and because there was no inhibitory unit above it. (c) Output when the contrast of the edge was low (20%) and the inhibitory signals were inactive.

receptive field arrays (Figure 8c) and the edge location is very tightly localized. Figure 14c shows the results for the case in which the edge was at 20% contrast. At low contrast the spatial inhibition mechanisms are turned off and many more units in the receptive field lattice respond along the edge. The drop in contrast also produces a loss of the direction inhibition (Equation 11) feeding into the other velocity units tuned to different directions and so more directions now respond compared to the high contrast case (Figure 15c).

At low contrast the output from the MT units is low for locations away from the moving edge and so the



Figure 16. Tests of the velocity code model with multiple orientations and speeds a) Moving cross made up of two 16 pixel wide bars oriented at 60 and 120°. Each bar moved at 4°/s. (b) Output of velocity code model for input movie shown in (a). (c) Moving cross in which the 60° bar moves at 2°/s instead of 4°/s. (d) Vector output of model in the non-rigid case.



Figure 17. Results of tests of velocity code model (surface moving in depth). (a) Movie sequence used for tests. (b) Model estimates in the form of a vector flow field (x 5 for clarity). The spatial inhibition mechanism was turned off. (c) Model estimates after spatial inhibition is applied. Many redundant velocity signals are now absent. (d) Output of array of heading templates (Perrone, 1992; Perrone & Stone, 1994, 1998) tuned to a range of horizontal heading directions (-60° to 60° in 5° steps). The actual heading was at 0° but the velocity flow field without the redundant vectors removed caused the heading to be biased to the left (-15°). (e) When the spatial inhibition was in place, the heading estimate was correct and the total activity was reduced.

number of MT units responding on either side of the edge is greatly reduced. There is no need for the spatial inhibitory mechanisms in this case because the redundancy problem does not really exist. In fact the number of velocity signals being generated can be so low as to prevent later motion processing stages from registering any motion at all. The removal of the direction and spatial inhibitory mechanisms at low contrast alleviates this to some extent with the trade off being a loss of directional precision. The contrast-dependent spatial mechanism enables this trade off to occur depending on the contrast level present in the stimulus.

Tests with multiple orientations and speeds Methods

The previous tests of the model have all used a single moving edge in order to simplify the discussion. However these simple edge tests do not indicate how well the new model is able to register the motion of multiple edges at different orientations and different speeds. In order to demonstrate the direction inhibition and redundancy mechanisms in a more challenging context, tests were carried out using two moving bars oriented at 60° and 120° . In one case the bars both moved at 4° /s and in the other, one of the bars (60°) moved at 2° /s. Perceptually these scenarios tend to lead to two different percepts; in one case (Figure 16a) the bars appear as a rigid cross moving to the right and in the other (Figure 16c) the bars seem to slide past each other and the figure appears nonrigid (Adelson & Movshon, 1982).

Results

Figure 16b and d shows the output of the velocity code model in response to these two inputs. The vectors have been scaled (×4) for clarity. For the case when the bars moved at the same speed the model has determined the overall velocity to be to the right at a speed of 5.2° /s which is a good match to the rigid cross direction and speed (0°, 4.6° /s). For the non-equal speed condition, the velocity code model produced outputs from both the 2°/s channel (red vectors) and the 4°/s channel (black vectors), whereas for the Figure 16a condition the outputs were all generated in the same velocity channel (4°/s). In Figure 16d, the vectors tend to point in the direction normal to the edge orientation (330° for the 2°/s bar and 30° for the 4°/s bar). This is consistent with the percept that the figure is nonrigid and that the two bars are moving independently of each other.

Tests with expansion patterns

Methods

In order to demonstrate the importance of the redundancy reduction mechanisms in my new velocity code model I have included a test using a more complex 3D motion stimulus. This is shown in the movie of Figure 17a. The $(256 \times 255 \times 8 \text{ frames})$ movie simulates motion towards a black rectangle (1.7 m wide, 6 m high) offset 1.5 m to the right of the heading direction (0°, 0°). The velocity units were tiled across the image as per the diamond lattice array used in previous tests (Figure 8b).

Results

Figure 17b is the output of the new velocity code model in response to this movie sequence without the application of the spatial inhibition mechanisms. Figure 17c is the output when the spatial mechanisms are in place. Below each figure is the output of a set of heading templates (Perrone, 1992; Perrone & Stone, 1994, 1998) tuned to a range of possible heading directions spanning $\pm 60^{\circ}$ azimuth in 5° steps (elevation was fixed at 0°). Each of these templates represents the output of a set of neurons (MST) that integrate the velocity information generated at image locations across a wide part of the visual field (Britten & Van Wezel, 2002; Duffy & Wurtz, 1991; Saito et al., 1986; Tanaka et al., 1986).

For the Figure 17b case, the heading direction was incorrectly signaled as being 15° to the left of the correct location $(0^{\circ}, 0^{\circ})$. When the spatial inhibition is applied, the velocity flow field (Figure 17c) is sparser and the heading direction is correctly indicated by the heading template array (Figure 17e). In addition to the incorrect heading estimate in the Figure 17b case the total output in Figure 17d is higher than in Figure 17e (see different *y*-axis scale). The redundant information along the edge of the rectangle is not only biasing the heading estimate but it also adds a significant level of additional noise against which the heading signal must be extracted. The peak sits on a larger pedestal of activity compared to the Figure 17e case. The performance of the heading detection stage (MST) is enhanced by the presence of the redundancy removal mechanisms.

Perrone

Tests of the model using small moving dots revealed a problem with the basic velocity code model. The shape and location of the S and T contrast response functions (Figure 9a and c) are determined by the parameters in the contrast gain stage (Equations 2 and 3). However the shape of the functions is also determined by the total amount of S and T energy generated by the (V1) spatiotemporal energy stage. When the energy from a particular V1 stage filter is very low because of the presence of a very small stimulus (e.g., a single moving dot) the contrast gain control stage is insufficiently powerful to produce the type of saturating functions that the WIM stage relies on to produce good speed tuning. It is difficult to set up values for the parameters that control the speed tuning (Equation 1) so that they are suitable for moving edges as well as for dots. One type of stimulus (edges or dots) ends up with poor speed tuning and so the behavior of the later MT stages is compromised. Since dot stimuli are commonly used in electrophysiological and psychophysical visual motion studies I have developed a solution to this problem to increase the generality and usefulness of the model.

A signature of small moving dot stimuli is that there is spatiotemporal energy at other orientations besides the one parallel to the direction of motion; this is particularly so for the orientation 90° to the motion path. A dot moving in the 0° direction produces a significant amount of energy in the sustained (static) filters that are oriented at 0° (parallel to its motion path). Edges on the other hand, have little or no energy in the sustained filters tuned to an orientation parallel to its motion direction because of the inhibitory flanks surrounding the filter. I use this property to modify the gain term (sc and tc) in the contrast-gain normalization mechanisms (Equations 2 and 3) applied to our early stage S and T filters.

I make the size of the sc and tc values in Equations 2 and 3 (which controls the size of S' and T') be based on the relative sizes of the outputs from the S_{θ} and $S_{\theta+90}$ filters. Specifically at each S filter location we calculate a ratio value:

$$R = \frac{S_{\theta+90}}{(S_{\theta}+\delta)} \tag{14}$$

Where S_{θ} is the spatiotemporal energy from the sustained V1 model units (used in the WIM stage) tuned to direction θ . $S_{\theta+90}$ is the energy from the V1 unit at the same location but with an orientation parallel to the direction of motion. The δ value prevents division by 0 and is set to .0005. For edges, the value of $S_{\theta+90}$ is close to zero and so *R* is close to 0. For small punctate points such as a moving dot, $S_{\theta+90}$ tends to be





Figure 18. Small dot stimuli mechanism. (a) Contrast response curves for the S and T early stage V1 model units with (solid curves) and without (dashed lines) the small stimulus gain mechanism (Equations 14 and 15). The dot was 4×4 pixels and moved at 2°/s to the right. (b) Movie used to test the velocity code model with the dot-gain mechanism in place. (c) Vectors (×4 for clarity) representing transformed output (Equation 10) of the model in response to the movie sequence.

closer to, or even larger than S_{θ} and so *R* is greater than 1.0.

We therefore set the values of the gain terms (*sc* and *tc*) in Equations 2 and 3 using:

$$sc = 0.15/\exp(wR) \tag{15}$$

and $tc = 0.93 \ sc$, where w is a weight dependent on the spatial frequency channel of the S unit: for the 1, 2, 4, and 8°/s channels, w was set at 1, 1.1, 2, and 3, respectively.

The values of w are chosen to produce contrast sensitivity functions that match those shown in Figure 9 and typical V1, MT data (Sclar et al., 1990). The values are also set to produce the best speed tuning at the WIM stage.

Note that this variation to the basic code works automatically and is not customized for just dots or just edges. The R output (Equation 14) dictates the gain level and this is determined by the particular stimulus moving over the filters. When used in conjunction with the mechanism specified in Equation 14, Equation 15 produces values of sc and tc that are relatively large (0.15) when an edge is moving across the S filters and small when a dot moves. A smaller value of sc and tc in the normalization stage (Equations 2 and 3) increases the gain and the moving dots end up with comparable S' and T' spatiotemporal values to those obtained with edges. Figure 18a shows the S' and T' spatiotemporal energy when a single 6×6 pixel dot moving at 2°/s was tested at a range of contrast values. The dashed lines show the energy output without the small dot gain mechanism (Equations 14 and 15). The solid lines are the outputs when the mechanism is in place. The contrast response curves now demonstrate saturation at high contrast



Figure 19. Natural image test (Forest scene). (a) Movie sequence. 1° /s planar motion to the right. (b) Velocity code model output in vector flow field form (scaled by factor of 8). No channel one motion sensors were located in the outside regions of the movie (within 14 pixels of the frame) to prevent wraparound effects. (c) Output of MST-like planar motion detectors. When transformed (Equation 10), the peak corresponds to 1.08° /s motion to the right and slightly downwards (-2°).



Figure 20. Natural image test: Hand-waving (after Watson & Ahumada, 1985). (a) Movie sequence used for test (b) Velocity code model output in vector flow field form (scaled by factor of four) superimposed over the middle frame of the sequence. Yellow is for downward pointing vectors and light green is for upward pointing vectors.

levels and are similar to the curves generated by edge stimuli (see Figure 10a). The later stages of the velocity code model (MT and the centroid stages) therefore work as designed and the code is able to extract the velocity signals from the moving dots (see below).

Tests with moving dots

Methods

In order to demonstrate this feature of the model I tested it with the movie sequence shown in Figure 18b. This is the simulated image motion that results from an observer moving parallel to a plane containing the dots that is slanted at 75° to the line of sight of the observer. The image dot speeds are .74, 1.4, 2, 2.6, and 3.3° /s when considering the dots from top to bottom.

Results

Figure 18c shows the velocity estimates (in vector form) derived from the new velocity code model for this movie sequence. The average estimated speeds for each dot from top to bottom were .82, 1.1, 1.7, 2.6, and 2.8° /s which is reasonably accurate. Without the additional small point gain mechanism (Equations 14 and 15) the velocity estimates were 0.0, 0.9, 1.5, 1.8, and 2.0° /s. With the addition of this feature the scope of the new model has been broadened and it can now be tested with the types of small dot stimuli commonly used in a wide range of electrophysiological and psychophysical studies of motion processing.

Natural scene image tests

The previous tests of the model used synthetic (computer generated) stimuli which lacked the variations in texture and contrast present in natural scenes. Therefore in order to test all aspects of the new velocity code model (multiple textures, low contrast, multiple directions and small features) two tests were run using actual video sequences.

Forest scene

A test was carried out using a movie sequence generated from an outdoors scene (forest). The movie sequence shown in Figure 19a simulated an eye rotation (or a camera pan) across the scene by extracting a 256 \times 256 region from a larger image and shifting the location of the subregion 1 pixel to the left every frame. The resulting image motion is globally 1° /s to the right (0°). Figure 19b shows the model outputs in vector form (scaled by a factor of eight). Figure 19c is the output of a set of planar motion detectors previously proposed as a method for estimating global patterns of uniform motion (Perrone, 1992) and which are designed to mimic the behavior of MST planar cells (Duffy & Wurtz, 1991). Each MST-like detector is tuned to a direction α (ranging from $0-360^{\circ}$ in 1° steps) and samples the velocity code model estimates across the whole image. For each image location corresponding to the position of the velocity channels, the detector sums the output generated from the velocity code model (Equation 9) for a particular direction θ and weighted by $\cos(\alpha - \theta)$.

The total activity across the image is normalized by the number of active velocity channels. This was found from the sum of the GP values across the image that only have positive values when a velocity signal is created (Equation 6).

Figure 19c shows the distribution of activity across the set of MST-like detectors in response to the forest movie. The detector unit tuned to an overall rightward motion (358°) responded the most and the peak activity level in this unit was 22.1. When converted to °/s using Equation 10, this value is equal to 1.08°/s. The model has performed well with this naturalistic test case and has established the global motion accurately despite the complexity of the features in the movie.

Moving hands

As an example of a natural scene with multiple moving directions, the movie shown in Figure 20a was tested using the velocity code model. This video clip contains two objects (hands) moving in opposite directions.

The velocity estimation model has successfully determined the correct overall direction of each hand. The resolution of the output could be increased by increasing the size of the input movie and it is currently limited by the spacing of the MT unit lattice array.

Discussion

I have presented a new velocity code theory explaining how the primate visual system can transform the outputs of sets of speed tuned MT neurons to a signal proportional to the speed of the stimulus such as is observed at the level of MST (Inaba et al., 2007). Because we have an existing model of MT neurons able to be tested with image sequences (Perrone, 2004; Perrone & Krauzlis, 2008a) the new velocity code model includes all of the stages from V1 spatiotemporal energy processing through to the MST stage via a series of MT neuron interactions. I have presented a number of problems that arise when attempting to use a population-based code for extracting a velocity signal from sets of MT neurons. The majority of these problems arise from the fact that MT neurons have different sized receptive fields and are tuned to a range of spatial frequencies. Combining outputs from these different sized MT units is not trivial because the location of the moving feature can influence the distribution of responses across the population of MT neurons. A significant part of the new velocity code model therefore involves mechanisms for overcoming these spatial problems.

Another core design principle underlying my new model is the need to reduce redundancy at the next stage of processing where the velocity signals are generated (between MT and MST). This requires inhibitory interactions between MT neurons prior to the estimation of velocity. However such inhibition is counterproductive at low contrast levels because it limits the signal available to the next levels of motion processing. Therefore I developed a contrast-dependent spatial inhibition mechanism that is automatically switched off at low contrast levels.

The resulting velocity code model is able to accurately estimate the velocity of complex moving objects under a range of contrast conditions while maximizing the spatial resolution of the velocity outputs. Unfortunately the additional power of the new code has come at the cost of added complexity. Therefore I will now summarize the functional units making up the velocity code model and relate them to neurons in the motion processing pathway of the primate visual system.

Summary of stages

Perrone

The first stage of the velocity code model uses the spatiotemporal energy from two types of filters (sustained and transient). Prior to the next stage, the S and T spatiotemporal energy outputs are transformed by a contrast gain control mechanism that uses a divisive feedback mechanism (Equations 3 and 4). I have also included a mechanism that increases the gain of the S and T filter responses when small punctate stimuli such as moving dots are present (Equations 14 and 15). The transformed S and T energy outputs are combined using Equation 1 to produce the WIM-stage speed-tuned responses. All of these S and T units are considered to be analogues of primate V1 neurons. The S and T units have spatial and temporal frequency tuning that are modeled directly on the properties of complex V1 neurons (Foster et al., 1985; Hawken & Parker, 1987; Hawken et al., 1996). The speed tuned WIM units correspond to the speed tuned complex V1 neurons discovered by Priebe et al., 2006, and have comparable speed tuning properties (Perrone, 2006).

The next stage of the velocity code model involves a specific set of connections between groups of the WIM units to form MT pattern neuron model units and MT component units. These could be connections within V1 or between V1 and MT. The pattern units have been well described previously (Perrone, 2004; Perrone & Krauzlis, 2008a). The component units are simpler in terms of their WIM inputs (only a single speed and direction is represented) but here I have introduced the new concept of units tuned to a lower temporal frequency (2 Hz) as well as overclocked units tuned

to 4 Hz. The concept of extending the temporal frequency tuning of the WIM units has been presented previously (Perrone, 2005) but it has never been applied to MT component units before.

Prior to the estimation of velocity I include inhibitory interactions between MT units whereby the MT pattern units are inhibited by the outputs from particular component units located nearby. This local inhibition solves the redundancy problems outlined above and it is contrast dependent. This inhibition is assumed to take place within area MT and would manifest itself in the form of asymmetric antagonistic surrounds as demonstrated in actual MT neurons by Xiao et al. (1997). It is also consistent with the contrastdependent inhibitory surround property noted in many MT neurons (Pack et al., 2005).

The subsequent stage of the velocity code model uses the outputs from a triad of MT units (two pattern types and one component type) to generate a velocity estimate using a centroid (weighted vector average) mechanism (Equation 7). The triad of MT units form a velocity channel and I use four such channels in my model (tuned to 1, 2, 4, and 8°/s). Each channel is controlled by a gain signal derived from another intermediate process which incorporates a second derivative mechanism whereby the output from the central unit of each channel (MT_v) is summed and the (.5 weighted) output from the units tuned to 2 V and .5 V is subtracted (see Equation 5). The inhibition from the units tuned to slower and higher speeds than the primary unit (MT_v) is important for overcoming the spatial scale problem (discussed in the section Problems with spatial scale). Spurious velocity signals from locations well away from the moving edge are suppressed by this second derivative mechanism. Included in the inhibition from the 2 V and .5 V units is also inhibition from other MT units tuned to speed V but tuned to different directions (Equation 11). This is designed to overcome the direction problem whereby other directions besides the correct one are signaling incorrect velocity estimates (see section Direction estimation). This direction inhibition is assumed to occur between MT and MST and so would not manifest itself at the level of individual MT neurons.

The final velocity estimation stage (Equation 9) is assumed to take place somewhere between MT and MST. Some individual MST neurons output a signal that is proportional to the speed of the input stimulus over a wide range of speeds (Inaba et al., 2007). Therefore the Equation 9 estimates would need to feed into a single MST neuron. It is known that MT neurons project directly to MST (Ungerleider & Desimone, 1986) and so the second derivative and centroid mechanisms (Equations 5–7) have to be based in the connections between MT and MST. Therefore the VMST stage (Equation 9) is considered to be the local MST input at a particular image location with each MST neuron having many such inputs distributed across wide regions of the image.

Depending on its heading or planar motion tuning, an MST neuron would also include cosine weighting of the VMST input at a particular location to establish the dot product between the VMST stage preferred direction tuning and the direction dictated by the heading or planar tuning (Perrone, 1992; Perrone & Stone, 1994). Therefore despite the inclusion of the direction inhibition stage in the model (Equation 11) the direction tuning of the MST planar units is still broad (see Figure 19c). The direction inhibition mechanism does not create MST units with direction tuning properties that are inconsistent with actual MST properties (Kawano, Shidara, Watanabe, & Yamane, 1994).

Similarly our MT model units do not end up with unrealistically sharp direction tuning curves because the direction inhibition signal is applied between MT and MST (Equation 12) and would not be apparent in the activity of the actual MT units themselves. Without this constraint, our MT pattern unit direction tuning curves would be much tighter than has been observed in actual MT neurons (Albright, 1984; Movshon et al., 1983; Pack & Born, 2001; Snowden, Treue, Erickson, & Andersen, 1991).

We have already demonstrated the power of units that integrate motion signals over wide areas of the visual field and which act as templates for specific fullfield motion patterns (Perrone, 1992; Perrone & Stone, 1994, 1998) as well as providing the ability to remove the effects of extraretinal signals (Perrone & Krauzlis, 2008b). Our previous models of MST processing have used the output of MT neurons directly to determine self-motion parameters such as heading. They relied heavily on the direction of motion to discriminate different global motion patterns and used a form of winner-takes-all to take into account the speed of motion. These models will now need to be modified to make use of the velocity signals generated by the new velocity code model.

It is possible to use the output from the MST units (e.g., the planar motion detectors in Figure 19c) to modify the velocity signal output (the vector flow field) such that anomalous directions and speeds are eliminated. However I have not incorporated this top-down feature into the velocity code model at this stage and am currently only concerned with the feed-forward process.

The component units in the new model form a key part of the second derivative stage (Equation 5) and the centroid stage (Equation 7) as well as playing an important role in the inhibitory (direction and spatial) mechanisms. The new velocity code model therefore ascribes a role to MT component neurons that has previously been unspecified. Despite their prevalence in the primate brain there has never been a good explanation provided as to their function. The ratio of component to pattern MT neurons is reported to be around 1.6 (Smith, Majaj, & Movshon, 2005). This fact was not the basis of my decision to include a particular number of pattern and component units into the velocity code model but the reported ratio maps nicely onto that produced by the eight component units and five pattern units (ratio = 1.6) in the model MT unit array (Figure 5).

In the development of the velocity code model I have been particularly mindful of ensuring that the properties of each stage of the model line up with the properties of the primate neurons they are supposed to be emulating. The V1 spatiotemporal energy stage units have temporal and spatial frequency tuning based directly on V1 neurons. The MT neurons have speed and direction tuning that matches those of actual MT neurons and the velocity estimation stage matches the linear velocity output of some MST neurons (Figures 3 & 7). This alignment with the neuron properties along the primate visual motion pathway is not trivial to achieve.

Many other velocity encoding schemes derive a velocity estimate immediately after the spatiotemporal energy (V1) stage (Adelson & Bergen, 1985; Johnston, McOwan, & Buxton, 1992; Thompson, 1984; Watson & Ahumada, 1985), effectively bypassing the speed tuning properties of MT neurons. They are left with the difficult question as to why the visual system, after having derived a velocity output proportional to the input at the V1 level, then resorts back to speed tuning at the following MT stage. Similarly, a popular model of primate velocity encoding (Rust, Mante, Simoncelli, & Movshon, 2006; Simoncelli & Heeger, 1998) is able to simulate some later higher-level aspects of motion processing without an MT stage that possesses the correct speed tuning properties (Maunsell & Van Essen, 1983; Perrone & Thiele, 2001). The primate visual system has a complex chain of neuron properties prior to displaying any evidence of a linear velocity code in MST (Inaba et al., 2007). Not only do these properties need to be matched in an effective velocity code model, but their order of occurrence is very important as well.

Many computer-vision algorithms have been suggested for image velocity estimation (e.g., Barranco, Diaz, Ros, & del Pino, 2009) but these do not incorporate the constraints imposed by the electrophysiological data from motion sensitive neurons. My code is primarily designed to be neurally-based in the strictest sense of the term. The properties of the input filters are based directly on known electrophysiological data from V1 and MT neurons and are not just inspired by the biology.

A number of alternative velocity population codes based on the outputs from sets of MT neurons have been postulated (e.g., Chey et al., 1998; Lisberger & Movshon, 1999; Priebe & Lisberger, 2004) but these models do not include a detailed V1-MT stage of neural processing and assume that the MT output has already been derived. Others have incorporated a V1-MT stage (Chey et al., 1998) but they use a V1 stage that includes elements that are incompatible with the known properties of motion sensitive neurons in V1 (Perrone, 2004). The bulk of these models have also ignored the problem of spatial scale and contrast that become apparent in real images containing objects of different

sizes and intensity levels.

Perrone

The spatial scale problem was recognized in a previous attempt to extract speed using a multiscale filter approach (Chey et al., 1998) and the solution was to use a "scale-proportionate threshold" to avoid the errors introduced by the skewed MT distributions (Figure 4b). A cutoff threshold set just above the dashed curve in Figure 4b would prevent the erroneous centroid estimate from being generated but it would also limit the ability of the system to respond to low contrast stimuli. The same threshold would block centroid estimates from legitimate MT distributions at low contrast. This is why I opted to use the second derivative mechanism instead.

Many of the concepts used in the model are not new and have been used in other contexts. For example contrast adaptive systems and redundancy removal is common throughout descriptions of many sensory systems and inhibitory surrounds are integral to many filtering schemes (Gallant & Prenger, 2008). My redundancy reduction mechanism is specific to motion processing and is motivated particularly by the need to prevent extra velocity signals being passed onto the MST stage of motion processing. Of note is that the new model's particular spatial mechanisms lead to (and can explain) the inhomogeneous patterns of antagonistic zones found outside the classical receptive field of MT neurons (Xiao et al., 1997).

Xiao et al. (1997) looked at the spatial distribution of the antagonistic surround of MT neurons. In one of their series of tests they stimulated and recorded from an MT neuron with small patches of moving dots while simultaneously presenting another small patch of moving dots to locations around the MT receptive field location. They found antagonistic (inhibitory) zones that were non-homogenous in space. The expectation based on earlier work (Allman, Miezin, & McGuinness, 1985) was that the surround inhibition would be symmetric and completely surround the central excitatory region. Instead the majority of the Xiao et al. neurons exhibited asymmetric antagonistic surrounds similar to the inhibition patterns I have included in the velocity code model (Figures 13 and 14). I believe that the Xiao et al. (1997) MT data supports the spatial inhibition mechanisms I have proposed as a

core part of the new code and we will present a direct comparison of the model and Xiao et al. data in a future paper.

An unexpected result while testing the model with a range of different stimuli was that small moving dots can be problematical to a system that relies on spatiotemporal filtering (energy calculation) using a wide range of filter sizes. The small dot stimuli provide only minimal stimulation for large filters and this has an impact on later normalization (divisive contrast gain) stages. Edge stimuli tend to fall along the whole length of the spatiotemporal energy filters and activate them strongly while the dot stimuli only activate the central regions. However the dot stimuli also activate sustained (static) filters that are oriented parallel to the dot motion path whereas edge stimuli do not. I was able to capitalize on this difference to develop a mechanism which increases the gain on the energy filters when the sustained V1 filter oriented 90° to it is also active.

While no direct evidence for such a mechanism in V1 neurons is available there have been previous suggestions for an interaction between the outputs from neurons tuned to a particular direction (θ) and from neurons oriented parallel to the direction of motion (θ + 90°). Geisler (1999) has presented a model suggesting that such an interaction can help with motion direction estimation. I am suggesting that cooperative mechanisms between the θ° and θ + 90° neurons could also help overcome the low output problem caused by small localized stimuli such as moving dots that do not cover a significant part of the receptive fields of the V1 neurons.

The development of a system that is able to extract velocity information from actual image sequences opens up the possibility of a huge number of possible tests concerning motion phenomena from a multitude of studies stretching back many years. I have limited the neural tests of the model in this paper to those that demonstrate some of the neuron properties I consider to be essential to the working of the model. A number of features in the model were included in order to "make it work" and there is currently no electrophysiological data on which to base decisions regarding the inclusion (or removal) of these particular features. The model generates a number of testable hypotheses regarding the properties of MT and MST neurons (e.g., MT component neurons should change their speed tuning with contrast but not pattern neurons) and we have begun a program designed to test these.

The main motivation for generating an image-based velocity code is to enable us to begin simulating the self-motion and depth extraction processes that begin to occur at the MST stage of the visual motion pathway. The velocity signals feeding into these MST neurons provide a strong basis from which to recover the three-dimensional layout of the world in front of the moving observer (Perrone & Stone, 1994) and our next endeavor is to model this 3D reconstruction stage.

More than 30 years ago, Nakayama (1985) suggested the concept of velocity channels and that velocity could be read out by comparing the activity in these different channels. He also mentioned using a population profile response as well as lateral inhibition. The new velocity code model has made some of these suggestions explicit and states that a velocity channel needs to be made up of a small set of MT neurons (just three in my model); too many MT neurons produce a population profile that can be distorted by spatial sampling and lead to erroneous velocity estimates. I have also demonstrated that Nakayama's suggestion of lateral inhibition is critical in an effective velocity code and that inhibition between MT neurons and between velocity channels is essential for overcoming a number of problems associated with extracting a velocity signal from MT neuron outputs (e.g., spatial and direction errors). Many of these problems only manifest themselves when the challenge of image velocity measurement is considered in the domain of real image sequences containing stimuli of various sizes and contrasts.

The biggest development since Nakayama conducted a review of biological motion processing in 1985 is the greater preponderance of information available concerning the properties of MT neurons (e.g., Krekelberg et al., 2006; Pack et al., 2005; Perrone & Thiele, 2001; Priebe et al., 2003). The many electrophysiological studies carried out on MT since 1985 have enabled detailed models of these neurons to be developed (e.g., Perrone, 2004; Perrone & Krauzlis, 2008a; Perrone & Thiele, 2002). The remaining challenge was to explain how these velocity-tuned neurons are able to convert their outputs into the linear velocity outputs seen in neurons at the following motion processing stage (MST). This paper presents one possible neural solution to this problem.

Acknowledgments

Thanks to Rich Krauzlis for his helpful feedback on previous drafts and to the two anonymous reviewers for their useful comments. Supported by the Marsden Fund Council from Government funding, administered by the Royal Society of New Zealand.

Commerical relationships: none. Corresponding author: John Perrone. Email: jpnz@waikato.ac.nz. Address: School of Psychology, The University of Waikato, Hamilton, New Zealand.

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(2), 284–299.
- Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300, 523–525.
- Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52(6), 1106–1130.
- Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, 14(2), 105– 126.
- Barranco, F., Diaz, J., Ros, E., & del Pino, B. (2009). Visual system based on artificial retina for motion detection. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics, 39*(3), 752–762.
- Bracewell, R. N. (1978). *The Fourier transform and its applications*. New York: McGraw-Hill.
- Bradley, D. C., & Goyal, M. S. (2008). Velocity computation in the primate visual system. *Nature Reviews Neuroscience*, 99, 686(610).
- Britten, K. H., & Van Wezel, R. J. (2002). Area MST and heading perception in macaque monkeys. *Cerebral Cortex*, 12(7), 692–701.
- Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985–2010. Vision Research, 51(13), 1431– 1456.
- Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, *38*(18), 2769–2786.
- Churchland, M. M., & Lisberger, S. G. (2001). Shifts in the population response in the middle temporal visual area parallel perceptual and motor illusions produced by apparent motion. *The Journal of Neuroscience*, 21(23), 9387–9402.
- Clifford, C. W. G., & Ibbotson, M. R. (2002). Fundamental mechanisms of visual motion detection: Models, cells and functions. *Progress in Neurobiology*, 68(6), 409–437.
- Dayan, P., & Abbott, L. F. (2001). Theoretical neuroscience: Computational and mathematical modeling of neural systems. Massachusetts: MIT Press.
- Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of

response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6), 1329–1345.

- 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.
- Gallant, J. L., & Prenger, R. J. (2008). Neural mechanisms of natural scene perception. In R.H. Masland & T.D. Albright (Eds.), *The senses: Comprehensive reference (Vol. 1)*. Oxford: Elsevier.
- Geisler, W. S. (1999). Motion streaks provide a spatial code for motion direction. *Nature*, 400, 65–69.
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233, 14161 419.
- 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., Shapley, R. M., & Grosof, D. H. (1996). Temporal frequency selectivity in monkey visual cortex. *Journal of Neuroscience*, 13, 477–492.
- Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A*, 4, 1455–1471.
- Hildreth, E. C. (1990). The neural computation of the velocity field. In B. Cohen & I. Bodis-Wollner (Eds.), *Vision and the brain*. (pp. 139–164). New York: Raven Press, Ltd.
- Inaba, N., Shinomoto, S., Yamane, S., Takemura, A., & Kawano, K. (2007). MST neurons code for visual motion in space independent of pursuit eye movements. *Journal of Neurophysiology*, 97(5), 3473–3483, doi: 10.1152/jn.01054.2006.
- Johnston, A., McOwan, P. W., & Buxton, H. (1992). A computational model of the analysis of some firstorder and second-order motion patterns by simple and complex cells. *Proceedings of the Royal Society B: Biological Sciences*, 259, 297–306.
- Kawano, K., Shidara, M., Watanabe, Y., & Yamane, S. (1994). Neural activity in cortical area MST of alert monkey during ocular following responses. *Journal of Neurophysiology*, 71(6), 2305–2324.
- Komatsu, H., & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements I. Localization and visual properties of neurons. *Journal of Neurophysiology*, 60, 580– 603.
- Krekelberg, B., van Wezel, R. J. A., & Albright, T. D. (2006). Interactions between speed and contrast tuning in the middle temporal area: Implications

for the neural code for speed. *The Journal of Neuroscience*, *26*(35), 8988–8998, doi: 10.1523/ jneurosci.1983-06.2006.

- Lisberger, S. G., & Movshon, J. A. (1999). Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *The Journal of Neuroscience*, 19(6), 2224–2246.
- Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal* of Neurophysiology, 49(5), 1127–1147.
- Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1983). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross, (Eds.), *Study week on pattern recognition mechanisms* (Vol. 54. pp. 117–151). Civitas Vaticana: Ex Aedibus Academicis.
- Nakayama, K. (1985). Biological image motion processing: A review. Vision Research, 25(5), 625–660.
- Nishimoto, S., & Gallant, J. L. (2011). A threedimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *Journal of Neuroscience*, 31(41), 14551–14564.
- Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409(6823), 1040–1042.
- Pack, C. C., Hunter, J. N., & Born, R. T. (2005). Contrast dependence of suppressive influences in cortical area MT of alert macaque. *Journal of Neurophysiology*, 93(3), 1809–1815, doi: 10.1152/jn. 00629.2004.
- 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. (2004). A visual motion sensor based on the properties of V1 and MT neurons. *Vision Research*, 44(15), 1733–1755.
- Perrone, J. A. (2005). Economy of scale: A motion sensor with variable speed tuning. *Journal of Vision*, 5(1):3, 28–33, http://www.journalofvision.org/ content/5/1/3, doi: 10.1167/5.1.3.
- Perrone, J. A. (2006). A single mechanism can explain the speed tuning properties of MT and V1 complex neurons. *Journal of Neuroscience*, 26(46), 11987– 11991.
- Perrone, J. A., & Krauzlis, R. J. (2008a). Spatial integration by MT pattern neurons: A closer look at pattern-to-component effects and the role of speed tuning. *Journal of Vision*, 8(9):1, 1–14, http://

www.journalofvision.org/content/8/9/1, doi: 10. 1167/8.9.1.

- Perrone, J. A., & Krauzlis, R. J. (2008b). Vector subtraction using visual and extraretinal motion signals: A new look at efference copy and corollary discharge theories. *Journal of Vision*, 8(14):24, 1–14, http://www.journalofvision.org/content/8/14/24, doi: 10.1167/8.14.24.
- Perrone, J. A., & Stone, L. S. (1994). A model of selfmotion 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. (2001). Speed skills: Measuring the visual speed analyzing properties of primate MT neurons. *Nature Neuroscience*, 4(5), 526–532.
- Perrone, J. A., & Thiele, A. (2002). A model of speed tuning in MT neurons. *Vision Research*, 42, 1035– 1051.
- Priebe, N. J., Cassanello, C. R., & Lisberger, S. G. (2003). The neural representation of speed in macaque area MT/V5. *The Journal of Neuroscience*, 23(13), 5650–5661.
- Priebe, N. J., & Lisberger, S. G. (2004). Estimating target speed from the population response in visual area MT. *The Journal of Neuroscience*, 24(8), 1907– 1916, doi: 10.1523/jneurosci.4233-03.2004.
- Priebe, N. J., Lisberger, S. G., & Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *The Journal of Neuroscience*, 26(11), 2941–2950, doi: 10.1523/jneurosci.3936-05.2006.
- Rodman, H. R., & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Vision Research*, 27(12), 2035–2048.
- Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, 9(11), 1421– 1431.
- Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., & Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience*, 6(1), 145–157.
- 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.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of

the neuronal responses in visual area MT. Vision Research, 38, 743–761.

- Smith, A. T., & Snowden, R. J. (Eds.). (1994). Visual detection of motion. London; San Diego: Academic Press.
- Smith, M. A., Majaj, N. J., & Movshon, J. A. (2005). Dynamics of motion signaling by neurons in macaque area MT. *Nature Neuroscience*, 8(2), 220–228.
- Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *The Journal* of Neuroscience, 11(9), 2768–2785.
- Sperling, G., Neil, J. S., & Paul, B. B. (2001). Motion perception models. *International Encyclopedia of the Social & Behavioral Sciences*. (pp. 10093– 10099). Oxford: Pergamon.
- 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. *The Journal of Neuroscience*, 6(1), 134–144.

Thompson, P. (1984). The coding of velocity of

movement in the human visual system. Vision Research, 24(1), 41-45.

- Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of area MT in the macaque. *Journal of Comparative Neurology*, 248, 190–222.
- Watson, A. B., & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2(2), 322–341.
- Watson, A. B. & Ahumada, A. J., Jr. (1983). A look at motion in the frequency domain. In N. I. Badler & J. K. Tsotsos (Eds.), *Motion: Representation and perception*. (pp. 1–10). New York: Association for Computing Machinery.
- Xiao, D. K., Raiguel, S., Marcar, V., & Orban, G. A. (1997). The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cerebral Cortex*, 7(7), 662–677, doi: 10.1093/cercor/7.7.662.
- Yuille, A. L., & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature*, 333(5), 71–74.
- Zeki, S. M. (1980). The response properties of cells in the middle temporal area (area MT) of owl monkey visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, 207, 239–248.