PERGAMON

Vision Research 38 (1998) 2769–2786

Vision Research

Neural dynamics of motion processing and speed discrimination

Jonathan Chey, Stephen Grossberg *, Ennio Mingolla

Department of Cognitive and Neural Systems and Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA

Received 25 October 1994; received in revised form 5 October 1995; accepted 30 September 1997

Abstract

A neural network model of visual motion perception and speed discrimination is presented. The model shows how a distributed population code of speed tuning, that realizes a size–speed correlation, can be derived from the simplest mechanisms whereby activations of multiple spatially short-range filters of different size are transformed into speed-tuned cell responses. These mechanisms use transient cell responses to moving stimuli, output thresholds that covary with filter size, and competition. These mechanisms are proposed to occur in the V1 → MT cortical processing stream. The model reproduces empirically derived speed discrimination curves and simulates data showing how visual speed perception and discrimination can be affected by stimulus contrast, duration, dot density and spatial frequency. Model motion mechanisms are analogous to mechanisms that have been used to model 3-D form and figure-ground perception. The model forms the front end of a larger motion processing system that has been used to simulate how global motion capture occurs, and how spatial attention is drawn to moving forms. It provides a computational foundation for an emerging neural theory of 3-D form and motion perception.

Keywords: Vision; Motion; Speed; Velocity; Neural network; MT

1. Introduction

How are estimates of retinal speed of moving objects extracted from continuously changing optic input? Classic computational models of motion detection involving Reichardt-like or motion-energy mechanisms have focused on the recovery of motion direction [1–4]. Primate units of MT exhibit both speed and directional tuning, however [5–8].

The code for speed at a particular spatial location for our model is a distribution of activity in a bank of neural units of multiple scales, whereby units of larger spatial scale code for faster speeds. These multiple spatial scales model the short-range motion process of Braddick [9], including the fact that the short-range motion limit D_{max} depends on the spatial frequency content of the image [10–15]. Such a multiple-scale short-range filter was introduced in earlier versions of the motion model that is developed here, where it was used to stimulate data about long-range apparent motion [16], including beta motion, gamma motion, delta motion, reverse motion, split motion, Ternus motion, reverse-contrast Ternus motion, and Korte's laws [17–19]. Here we show how such a multiple-scale filter can be appropriately combined with other model mechanisms to explain psychophysical and neural data about speed perception.

Heeger [20] has earlier shown how spatiotemporal filtering over a coarse set of units of different spatial extent can encode motion direction and speed. Heeger's model expresses a computational intuition about multiscale speed coding in a set of channels conceived as independent, quasi-linear filters. Later Anadan [21] also employed a multiscale approach in a computational algorithm for robust motion estimation from image sequences. Our concern is to ask how the nonlinear, coupled neural units with limited dynamic range, that have already been used to explain many other psychophysical and physiological motion data, can be organized to display speed-tuning in a distributed multiscale representation.

In the present model, speed is represented through the distributed activity of speed-tuned units, or cell populations. We define speed-tuned cells as those that respond preferentially to a limited, continuous range of

* Corresponding author. Fax: +1 617 3537755; e-mail: steve@cns.bu.edu.

0042-6989/98/$19.00
PII: S0042-6989(97)00372-6

speeds, as opposed to speed-sensitive cells, which vary their response with speed, but do not exhibit a preference for a particular speed. There is considerable neurophysiological evidence for speed-tuned cells in the visual systems of cat and monkey [22–25]. In the model, the speed-tuned cells are arranged in a topographic neural map so that at each location there is a set of cells tuned to a range of speeds. The distribution of activity across the cells in the map then implicitly codes speed estimates at different locations. Cheng et al. [26] have reported that the distribution of cells as a function of speed is not uniform in MT, with most cells peaking at high velocities (32–64°/s). While such a distribution would seem ill-suited to support discrimination among lower speeds, Cheng et al.'s animals were anesthetized, possibly resulting in a need for stronger (i.e. faster) stimuli to drive them than would normally be the case. Also, as we will demonstrate, most of the cells at intermediate stages of our model respond to high velocities, and specific mechanisms operate to ensure that the tuning of cells of subsequent stages is more specific to lower or intermediate velocities. Thus the cells recorded by Cheng et al. may more closely correspond to intermediate than to final stages of our model.

The speed-sensitivities of the model cells arise primarily from their different spatial scales, which determine the size of their input fields, in accord with recent evidence that receptive field sizes of foveal motion sensitive units range from approximately 0.03–1°, as inferred from psychophysical methods [27,28]. In the primate nervous system, speed tuning appears to arise in cortex through the combination of signals from more peripheral cells. Although cells in cat or monkey retina and LGN may exhibit speed-sensitivity, there is no evidence for speed-tuning until striate cortex in both cat [29,30] and monkey [31], in which some, but not all, cells are speed-tuned. Speed and directional tuning both become more prevalent in monkey area MT [6,8,32,33], suggesting that this area is further upstream in a specialized motion processing system that successively refines motion signals. Such speed-tuned cells are usually directionally-selective and exhibit both facilitatory and suppressive interactions within their receptive fields, indicating that peripheral signals may be nonlinearly combined to yield speed tuning.

Our primary goal in developing this model was not to determine a computationally optimal method of extracting speed from optic input, but to simulate important characteristics of human speed perception. Of particular interest are the sensitivities or insensitivities of speed perception to non-speed parametric variations in the stimulus. These sensitivities can reveal details of the operation of the mechanisms that underlie speed perception. In this model we account for and qualitatively simulate data showing changes in speed discrimination resulting from changes in stimulus contrast [34] and duration [35]) as well as changes in perceived speed resulting from changes in stimulus contrast [34] and the density of moving random dot fields [36]. In addition we also account for variations in reaction times at different stimulus speeds [37].

A key model hypothesis is that the spatial scale or input field size of a cell determines its speed-sensitivity, such that larger scales respond preferentially to faster stimuli. We call this covariation the size-speed correlation. Analogously, cells in monkey area MT typically have larger receptive fields, exhibit directional interactions over larger areas and are directionally tuned for a greater range of velocities than cells in V1 [7]. Likewise, in cat visual areas 17 and 18, cells with slow speed preferences are generally unaffected by masking of peripheral portions of their receptive fields, while cells preferring intermediate and high speeds showed reduced responsiveness at higher speeds after the same masking [38].

The theme of multiscale representation occurs in many traditions in the study of vision, notably research on psychophysical channels for pattern perception and in the development of efficient image coding procedures for machine vision. Among the modalities of primate vision, coding at multiple spatial scales is perhaps most familiar in stereo vision [39] (see Ref. [40] for a review). Here sensitivity to amount of disparity is known to covary with sensitivity to size [41–45]). This covariation is often called the size-disparity correlation [41–45].

The visual system is faced with the problem of maintaining sensitivity to a wide range of speeds, using mechanisms with limited operating ranges, without sacrificing speed resolution or spatial resolution, such as when small objects travel very fast. Since a simple 'match filter' scheme using neurons uniquely tuned to every combination of speed, size, contrast, and so forth is hopelessly impractical, units with overlapping sensitivities to spatial and temporal parameters of inputs must be used. These considerations lead, in turn, to issues concerning how speed codes of units sensitive to a range of spatial and temporal frequencies can be properly tuned. The main problem in understanding how the brain represents speed using a multiple scale population code can be succinctly stated: why does not the largest scale always win in response to all input speeds, simply because it has a larger receptive field with which to attain a higher level of activation? This problem arises because each scale is turned ON whenever a contrast passes through the region corresponding to the filter's spatial extent. For this to happen, signals from any changing visual cue input to units of all scales. A continuously moving contrast has a longer dwell time in the domain of a large-scale filter than in the domain of a small-scale filter centered at a corresponding retinal location. Without further pro-

cessing, units corresponding to the largest scales will fire more vigorously. How does one prevent these largest scales from always being the most active, thereby winning the competition for coding a moving feature's speed, regardless of the feature's true speed?

Our work suggests that two simple mechanisms suffice. The first is a scale-proportionate, or self-similar, threshold, which requires units of larger scales to have larger absolute activity (or, equivalently, similar proportions of their maximum possible activity) to transmit a signal. The second is competition, both among units of similar scales and across units of differing scales. These simple ideas suffice to explain many data about speed tuning. They can be reviewed as perhaps the simplest way to realize a size-speed correlation.

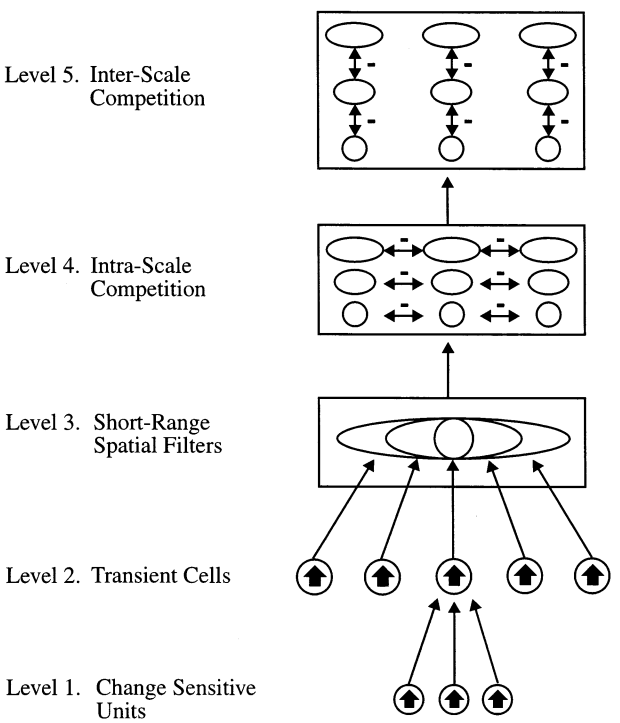

Fig. 1. Schematic outline of model layers. Level 1 consists of change-sensitive units that are transiently activated for fixed time intervals by a moving stimulus. Level 2 transient cells sum and time-average the activities from fixed, non-overlapping sets of the change-sensitive units. Multiple short-range filters occur at Level 3 at each spatial position. Each filter draws input from a set of transient cells, the size of which is determined by the spatial scale of the filter. As depicted, scale 1 receives input from one transient cell, scale 2 from three cells and scale 3 from five cells. The transient cells that input to each filter overlap both between scales at a single position and between positions at a single scale. Thus, the largest scale depicted in the outline draws input from a superset of the transient cells from which the smaller scale draws. The outputs of the short-range filters input to Level 4 intra-scale competition across space. This competition is enacted through a spatial center-surround network. At Level 5, inter-scale competition takes place between all scales at each spatial position, again through a center-surround network. The activities across this final competitive stage form the output of the network and are pooled to define a population measure of speed.

2. The speed-sensitive MOC filter

We now outline the architecture of a speed-sensitive filter network and give a functional description of each processing level, illustrated by results of a simulation of the network's behavior. In these descriptions we concentrate on the rationale for each network level without specifying the equations that implement it. Details of the equations and simulation parameters are given in Appendix A. Since the response properties of a cell are determined by the activity of other network cells as well as the network input, each simulation must include a set of cells that contains all neighboring cells whose activity affects the cell or cells at the location reported upon. For simplicity, the model is simulated using one-dimensional stimuli. This allows us to here concentrate on the speed-sensitivity of the network without reference to more complex two-dimensional spatial summation or segmentation. Another report describes how a two-dimensional implementation of a speed-tuned network such as the one presently described can address data on the aperture problem, motion capture, and related effects wherein both motion direction and speed need to be taken into account [46]. Each network simulated here therefore consists of a sequence of neighboring cells. Activity is always reported from the middle cell in this sequence.

A schematic representation of the network is given in Fig. 1. This diagram shows the five processing levels and how cells in each level interact to provide input to the next. The components of the network are: (1) change-sensitive units; (2) transient cells; (3) short-range spatial filters; (4) intra-scale competition; and (5) inter-scale competition.

3. Level 1: Change-sensitive units

In the model (Fig. 1), visual input is initially registered by change-sensitive units that respond briefly to changes in luminance over time at a location. An output pulse of fixed length, independent of input speed, is generated when a moving object enters the receptive field of such a unit, conceived as a simplified photoreceptor. The exact response profile of these cells is not important; for simplicity, a square wave output is assumed. The simulations of our model work despite, not because of, the square waveform, which makes it harder to generate smoothly modulated speed profiles than would be the case if realistic profiles of receptor impulse functions were incorporated. Fig. 2 displays representative activity profiles over time for change-sensitive units.

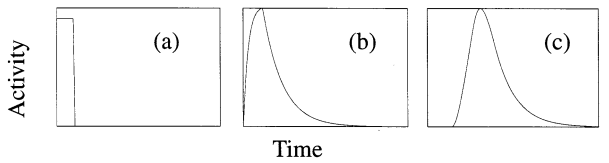

Fig. 2. Representative time-courses of simulated cellular activities from each level of the network: (a) change-sensitive unit; (b) transient cell; (c) thresholded short-range spatial filter. For each level, a single cell's activity is shown as a function of time in response to an input which traverses a range of simulated cells. The input moves at a speed of ten change-sensitive units per simulation time unit. For (a), activity is shown for the first (leftmost) of the series of simulated units. All other activities are taken from cells in the middle of the simulated series. For (c), an intermediate sized scale (5) is chosen for display purposes. These plots are shown on the same horizontal time scale but not on the same vertical scale.

4. Level 2: Transient cells

Model transient cells accumulate inputs from a series of adjacent change-sensitive units and time-average these inputs. Once again, we do not claim to have modeled all important characteristics of the dynamics of such units in vivo, although model cells are assumed to correspond to retinal cells with transient response properties, such as cat Y cells [47] or monkey M cells [48]. Each transient cell responds with an exponentially rising and decaying activity whose duration is determined both by the spatio-temporal parameters of the cell and the response duration and amplitude of the change-sensitive units (see Fig. 2). Due to their time averaging properties, activation of adjacent transient cells may overlap in time even though the input fields of the cells do not overlap in space, and this trait proves important for the ability of subsequent levels of the network, whose units receive input from several adjacent transient cells, to be differentially activated by different motion speeds. At this level, cell responses rise monotonically with input speed due to the temporal summation performed by these cells, whereby more inputs are 'counted' per unit time. (Fig. 3(a)). However, these transient cells are not speed-tuned and, due to the small, fixed size of their input fields, their maximal responses saturate at a low speed.

5. Level 3: Self-similar short-range filters

Level 3 cells utilize short-range spatial filters of a variety of widths to collect and time-average input from a series of adjacent transient cells. The different filter widths, or spatial scales, give rise to different speed sensitivities. At the final level of the network, the scale which responds maximally will covary with input speed. In fact, the basic intuitions behind the network design concerns how to bring about such a state of affairs, considering that the first plausible aggregation of Level 2 outputs by spatial filters at Level 3 cannot achieve such a result. Specifically, at Level 3, larger scale cells always respond at least as vigorously to an input as smaller scale cells, since they draw input from more transient cells (Fig. 3(b)). At slow speeds, however, spatially adjacent transient cell responses (Level 2) show little temporal overlap, nullifying the advantage of a large input field, so large scales at Level 3 respond just a little better than small scales. At fast speeds, there is significant temporal overlap of transient cell responses within larger scales, so large scales respond significantly more vigorously than small.

The model's Level 3 cells use a fixed time averaging rate. This rate affects how vigorously each cell responds to an input, how long it remains active and, therefore, at which input speed it begins to respond vigorously. Since all scales are assumed to respond at the same rate in the current implementation of the model, the spatial extent of the inputs to a cell (i.e. the cell's scale) determines what this speed will be. Fig. 3 shows the

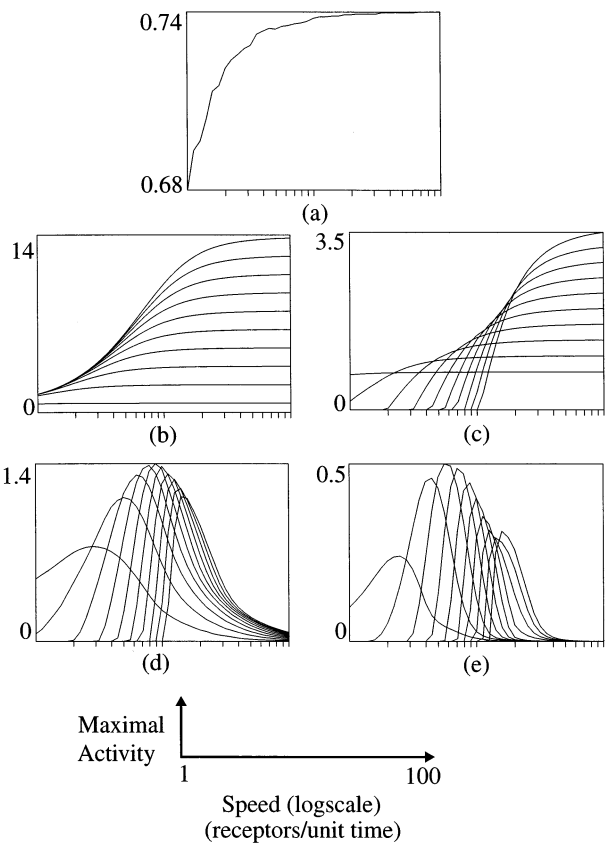

Fig. 3. Maximal responses of cells in the network to a variety of simulated speeds. Plots show (a) transient cells; (b) short-range spatial filters; (c) thresholded short-range filters; (d) intra-scale competition; and (e) inter-scale competition. For levels where there are multiple spatial scales at a single position, activities from all scales are shown as different curves superimposed on the same plot. The smaller scales always respond less vigorously to fast speeds, so their activity profiles always show lower values. This is particularly evident at fast speeds. Vertical axis scales vary with each plot and are indicated next to each.

maximal activities from cells at each level of the network plotted against simulated speed. This allows the speed-tunings of the cells to be observed. These curves were obtained by simulating 41 different input speeds in a logarithmically increasing fashion from speed 1 to speed 100. In the simulation, input 'speed' refers to the number of change-sensitive units traversed in a simulation time unit. Short-range filters are of several spatial scales, such that some receive input originating from a few neighboring change-sensitive units, while others receive input from a greater number. A moving input crosses the modeled receptive fields at a constant speed, thereby creating a series of activity pulses each time it enters the input field of a new Level 1 receptor. Each simulation is run until activity in all network cells falls below a threshold. Maximal activities are recorded during the entire simulation time period.

Fig. 3(b) illustrates the core computational problem faced by a neural multi-scale filter approach to speed detection: How to keep the largest scales from always winning, in the sense of being the most active. Since the final code for speed is presumed to be a (possibly weighted) average of activity at Level 5 across all scales, to a first approximation, which scale is the most active determines the model's 'perceived speed'. Cells of larger scales can be prevented from always winning a cross-scale competition (subsequent to Level 3) if a scale-proportionate threshold is first applied to Level 3 outputs. Note that in Fig. 3(c), the activity curves for the largest scales, which achieve the highest absolute level on the right of the plot, begin to exceed zero at greater levels of input than do the curves for the smallest scales.

The thresholded short-range spatial filter enables cells of different scale to be maximally active in different speed ranges, as in Fig. 3(c). These cells are, however, not truly speed-tuned because large scales always respond better than small, and the difference in response between large and small scales increases monotonically with input speed. True speed tuning requires that different subsets of cells—i.e. cells of different scales in the present model—generate the maximal outputs across all cells as input speed varies. Scale-sensitive selectivity of response is nonetheless achieved despite the tendency of larger spatial filters at early levels to achieve maximal amplitude of activation. This selectivity-amplitude trade-off occurs because these cells realize a property of self-similarity; namely, larger scales require larger total inputs in order to respond, as in Fig. 3(c). This property can be realized, for example, if larger scale spatial filters arise from larger dendritic trees of larger cell bodies. The larger cells require a larger total input in order to fire due to their ability to dissipate membrane potential over a larger cell surface area and volume.

6. Level 4: Intra-scale competition across position

Through short-range spatial averaging, the initially localized moving stimulus is spatially blurred. This blurring process begins to transform the temporal signals from a moving stimulus into a spatial map whose cells respond selectively to different speeds. Competition across space, within each scale, deblurs these activity profiles. This competition locates the maximal activity across space within each scale. For simplicity, a feedforward on-center off-surround network is here used to realize this competition. Within each scale, cells receive excitatory input from cells in the previous network level at spatially proximate locations and inhibitory input from cells at spatially distant locations. In addition to deblurring motion signals, the intra-scale competition plays a key role in achieving the speed-tuning of cells (Fig. 3(d)), through the suppression of homogeneous responses of large-scale filters across spatial locations for high speeds. Now each scale tends to respond unimodally as input speed increases, and the scale of the maximally active population tends to increase with input speed.

The faster the speed of an input, the more homogenous the activity of neighboring spatial filters, which sample from overlapping distributions of units at earlier stages. Since spatial competition tends to enhance differences among already differing units and to suppress regions of homogeneous activity, less activity results for high speeds. At the extreme, when input activity to a competitive network is completely homogenous, the competition completely suppresses responses. Therefore, it is no longer true that all scales respond maximally at high input speeds. Now each scale tends to respond unimodally as input speed increases, and the scale of the maximally active population tends to increase with input speed. Thus, each cell does achieve a measure of speed tuning.

Note that the suppression of responses from homogeneous patterns is not inconsistent with the perception of a coherently moving texture, such as a field of random dots, or periodic pattern, such as a sine wave grating. A grating, for example, would elicit inhomogeneous responses at the stages described, because the spatial modulation of contrast from peak to trough of each period of the pattern would ensure that neighboring spatial filters of Level 4 always receive differential input.

7. Level 5: Inter-scale competition within position

On the other hand, the total network output from all active cells still tends to become unselective at high input speeds (Fig. 3(d)), with all scales still responding at high speeds. The existence of inter-scale competition

at Level 5, in addition to intra-scale spatial competition at Level 4, overcomes this imperfection of the collective speed-sensitivity of the entire network (Fig. 3(e)).

Similar spatial and inter-scale competition stages have been used to disambiguate the responses of multiple scales to size-disparity correlations during the process of static figure-ground separation [40]. In the present article, the disambiguation is applied to multiple scales that compute a size-speed correlation. The inter-scale competition further sharpens the speed tuning of the cells as follows. Since larger scales tend to respond maximally at larger input speeds, they win the competition and suppress the lesser responses that are distributed across an increasing number of scales as input speed increases. Feedforward competition achieves this result in the present formulation, realized again by an on-center off-surround network, this time in scale space. By this means, each scale receives excitatory input from cells of the same scale from the previous network level and inhibitory input from cells of different scales. Both excitatory and inhibitory inputs are drawn from the same spatial position. Application of a power function to the excitatory and inhibitory inputs biases the competition towards selecting a single winning scale rather than distributing activity across multiple scales. This competition also tends to normalize activity across scales at each location.

The final tuning curves produced by this competition (Fig. 3(e)) are such that each scale responds maximally to a speed that increases monotonically with scale. The maximal activities of the middle scales are somewhat higher than those of the small or large scales, but as we shall see, it is not the absolute response of any one scale that is important for speed tuning, but rather the distribution of activity across the entire set.

8. Relating network activity to perceived speed

The output of the inter-scale competitive level is a spatial map whose distributed activities implicitly represent the speed of the input. In order to interpret this distributed activity pattern for comparison with perceptual data, a linking hypothesis is defined that relates the entire population of active cells to perceived speed. It is here assumed that perceived speed derives from the sum of activities over all scales weighted by the size of the scale. Thus, if large scales are active, speed will be perceived as fast, and if small scales are active, then speed will be perceived as slow, with the level of activity at each scale determining the exact speed percept. We call the calculated number the speed measure.

The maximum speed representable in the network is equivalent to the largest scale and the minimum to the smallest. Fig. 4 shows the speed measure for a range of stimulus speeds. In the model, perceived speed increases approximately linearly with stimulus speed until around speed 20, where it saturates. This saturation is due to the limited range of scales simulated.

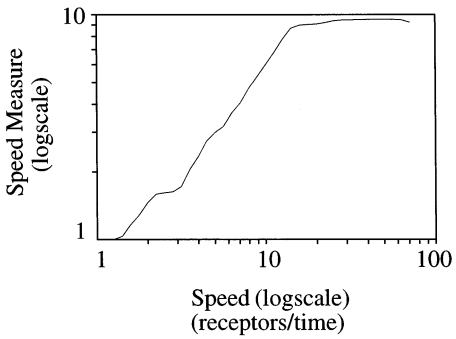

Fig. 4. Simulated speed measure as a function of input speed. The speed measure is obtained from a weighted sum of activities at all scales at the final network level (after the inter-scale competition). The speed measure increases approximately linearly with input speed until it saturates.

9. Computer simulations of psychophysical data

The model stimulates challenging characteristics of human speed perception data. Many phenomena lie outside the domain of speed filtering as a sole mechanism. Experiments were selected that probe speed filtering processes. Such experiments use parametric changes in displays where perceptual grouping effects are not rate-limiting on judgments of speed, as is the case for the reorganization of perceived speed that characterizes the transition from component to plaid motion, or the barberpole effect. See Chey et al. [46] for an analysis of how an extension of the present model handles such cases.

In all simulations reported below, the same network parameter set is used. Network responses are robust across variation of parameters, and network equations were written so as to make the effects of each parameter as clear as possible, rather than to derive the smallest number of parameters by dimensional analysis. As it is, only two parameter values (1 and 10) were used in all the model equations. What is essential for model function is the operation of the conceptually simple and robust processes of a scale-proportionate threshold followed by competition. As long as larger scales have larger thresholds, speed tuning results, though good selectivity depends on maintaining approximate proportionality between the size of the threshold and the size of the maximum possible (unthresholded) activity at a scale. The former generates selectivity at low speeds, and intra-scale competition suppresses responses of larger scales at high speeds, which produce spatially homogeneous activation across a wide region. All parameter choices are listed in the Appendix. The pri-

mary goal of the simulations is to show how the model naturally generates qualitative parametric properties of the data. Quantitative fits were deemed premature, since many simplifications were made in model computations to make them tractable, and additional mechanisms such as directional-selectivity and long-range motion grouping need to be added before the model can be said to be complete.

10. Speed discrimination

Human velocity discrimination can be measured by requiring subjects to judge the relative speeds of two successively presented stimuli moving at different velocities. Using this technique, de Bruyn and Orban [35] found that observers could discriminate random dot speeds ranging from, at least, 0.5–256° of visual angle per second. Discrimination performance within this range varied such that optimal discrimination was achieved at intermediate speeds, with poor discrimination at either extreme (Fig. 5(a)). They found that, at its best, the Weber fraction for perceived speed reached around 5%, a level which remained roughly constant for speeds from 4 to 64°/s. Orban et al. [49] found similar discrimination properties using a moving bar stimulus.

These data were simulated using the hypothesis that two stimuli can be discriminated by their speed if the total difference in activity across all scales in the model at those speeds exceeds a threshold. By total difference in activity is meant the sum of the differences in activity between each corresponding scale at the two speeds (see Appendix A for the exact formula).

In order to obtain a measure similar to a Weber fraction, maximal speed measures from the network were calculated using a variety of stimulus speeds, generated in a logarithmic progression from a base speed. A reference speed was selected from these speeds. Test speeds were then selected, again from the logarithmic series of simulated speeds, just above and below the reference speed. The number of increments or decrements of the test speeds necessary to obtain an above threshold difference in response of the reference from either the higher or lower test speed was then calculated. Since the simulated series of speeds increases in a logarithmic fashion, a constant number of increments or decrements of speed in this series corresponds to a constant ratio of speeds. For example, if the test speed is four increments above the reference speed, then the ratio of the test to reference speed is the same regardless of how the reference speed is chosen.

We call the minimum number of increments or decrements of the test speed required to exceed the difference threshold the discrimination measure. The use of higher or lower test speeds to form the discrimination measure was a convenience adopted solely to reduce discretization aliasing in our simulations. This discrimination measure is plotted against the simulated reference speed in Fig. 5(b.) The discrimination measure has the same form as the discrimination data in Fig. 5(a). The characteristic U-shaped profile in the simulation results from the lack of change in network output at low and high speeds. At very low speeds, the only scale active is the smallest, so changes in input speed do not result in significant changes in network activity. Correspondingly, at high speeds only the largest scale is active, and at very high speeds no scales are active at all. Optimal discrimination is achieved at intermediate speeds where several scales are active simultaneously and any speed change results in a substantial change in active scales.

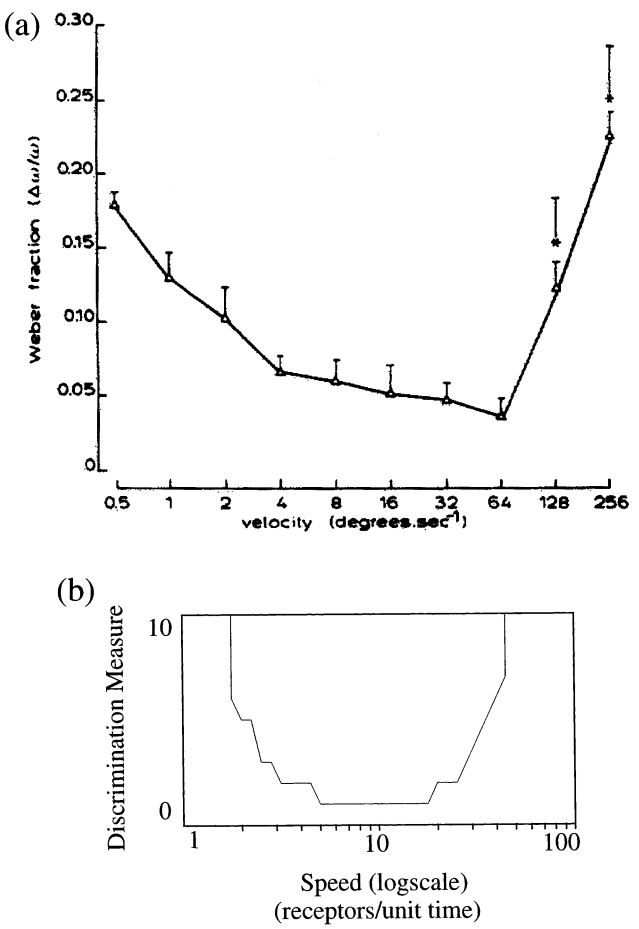

Fig. 5. Plot (a) shows experimentally derived Weber fractions illustrating just noticeable differences in velocity as a function of velocity. Reproduced with permission from Ref. [35]. Plot (b) shows simulated discrimination measures, hypothesized to correspond to the data in (a). These simulated discrimination measures are obtained by determining the number of speed changes (selected from a logarithmic range of simulated speeds) required to generate a difference in speed measures that exceeds a threshold. The simulated discrimination measures reproduce the U-shaped curve of the data.

11. Speed-sensitivity range and self-similar cortical magnification

Humans are capable of perceiving and discriminating many orders of magnitude of speeds. As noted above, de Bruyn and Orban [35] found that subjects could discriminate speeds as slow as 1°/s and as fast as 256°/s, though performance was poor at the extremes. If such discrimination is based on mechanisms whose speed-tuning scales approximately linearly with size, as in our model, then the sizes available must increase exponentially in order to account for the full range of reported sensitivities. Such an exponential increase in size or scale could be caused in part by the cortical magnification factor [50–53].

Correspondingly, in the model, at each retinal eccentricity there exists a range of scales whose size increases with the cortical magnification factor. This paper does not seek to reproduce the exact form of cortical magnification or to assess whether this magnification is due to retinal or cortical sampling characteristics. Instead, it is shown below that, using a range of scales whose ratio of largest to smallest remains the same at each location, a greater range of speeds can be discriminated if the entire set of scales is enlarged at peripheral locations. Again, a property of self-similarity obtains, here across position and scale, instead of across threshold and scale, as at the short-range spatial filter. Using this scheme, the total number of scales need not increase across position to achieve a significant expansion of the speed-sensitivity range. An analogous use of self-similar cortical magnification has been used to compensate for larger binocular disparities at larger eccentricities during the computation of planar surface representations in 3-D form perception [40].

Burr and Ross [54] have measured an ability to identify the direction of motion of an 80° bar at speeds up to 1000°/s. This task is rather different from discrimination between two speeds, and may reflect sensitivities of a direction-of-motion, as opposed to a speed-tuned, mechanism. Even this study, however, found a high-speed cut-off for visual sensitivity. Such a cut-off necessarily exists in any system, such as a brain, that is composed of finitely many processing units.

Fig. 6 reproduces the discrimination measures shown in Fig. 5(b), this time using two sets of scales, the second of which is an order of magnitude larger than the first. The range of speeds which can be discriminated is increased by an order of magnitude. This simulation assumes that the input traverses both ranges of scales.

One implication of this scheme is that fast speeds are relatively poorly discriminated in the foveal region and slow speeds poorly in the periphery. The experimental results discussed above were gathered using large stimuli that cover both central and peripheral locations and so do not address these issues. However, Maunsell and Van Essen [6] found that higher speeds tended to be coded at greater eccentricities in velocity-sensitive cells in cat visual cortex. Orban et al. [31] have found the same in monkey areas V1 and V2.

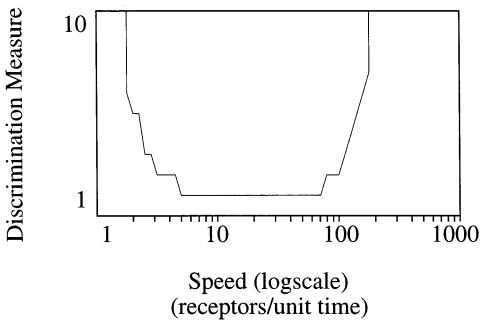

Fig. 6. Simulated discrimination measures using two ranges of scales, hypothesized to occur as a result of cortical magnification. The range of speeds which can be discriminated is increased while maintaining the same qualitative discrimination properties. The second set of scales is five times as large as the first.

12. Contrast effects on speed discrimination

Speed perception can be altered by stimulus contrast. Orban et al. [49] found that discrimination performance deteriorated at lower stimulus contrasts. This deterioration was particularly severe at slow and high speeds (Fig. 7(a)). McKee et al. [55] have disputed this conclusion; they found no contrast effect on velocity discrimination using contrast levels ranging from 5 to 82%. In Orban et al. [49], contrast, defined as $\log(\Delta I/I)$, ranged from -0.65 to 3 and velocity ranged from 0.25 to 256°/s. In McKee et al. [55], contrast ranged from -1.33 to -0.09 and velocity from 1 to 15°/s. Thus, these experiments investigated different, though overlapping, parameter ranges. Our model predicts a limited range of contrast-induced changes in speed discrimination.

The model accounts for contrast changes in perceived speed through the dependence of network output on the spatial and temporal summation of energy provided by receptor responses by the short-range filters of Level 3 (Fig. 1). Stimulus contrast changes translate into changes in receptor activity amplitude in the model, so that high contrast stimuli generate larger receptor activity amplitudes and low contrast stimuli smaller activity amplitudes. Fig. 8 shows the result of altering receptor amplitudes on the maximal activity of different network cells.

Since the model is based on spatio-temporal summation, one might expect that increasing input amplitude would result in a catastrophic failure of speed estimates, by causing large scale cells to respond at very slow speeds where they would normally be inactive. Several

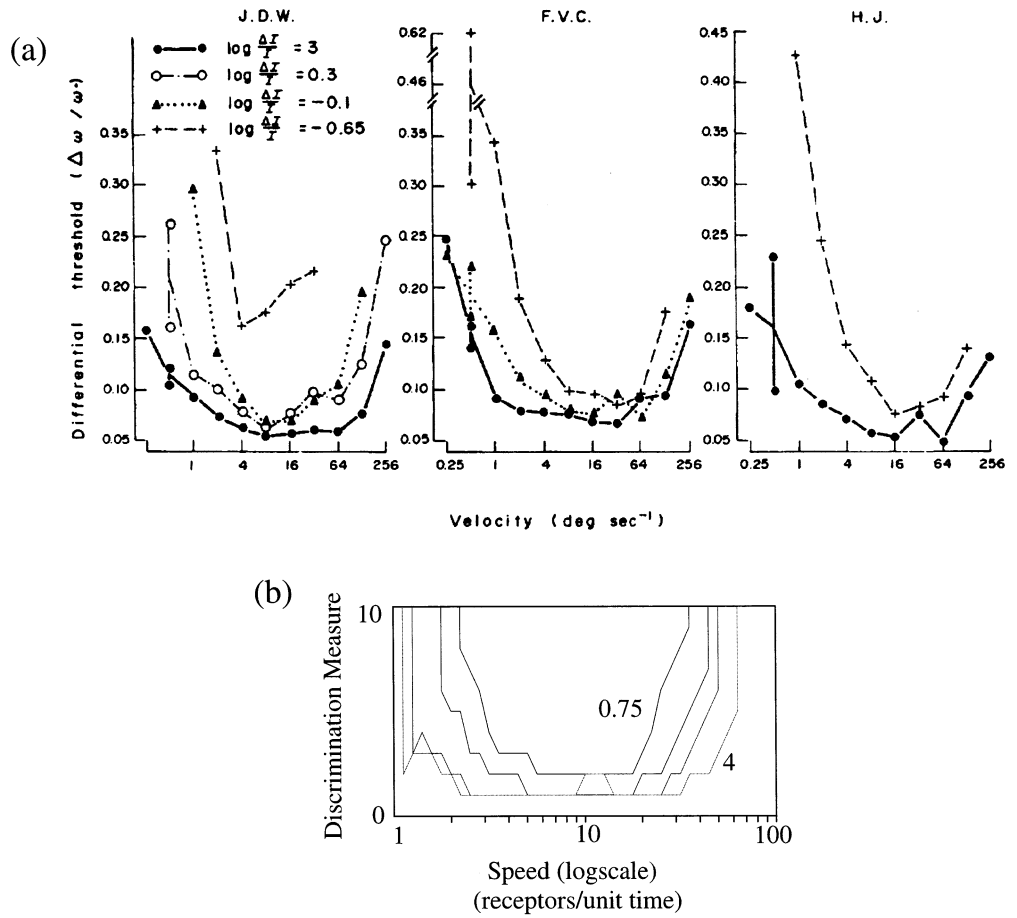

Fig. 7. Plot (a) shows experimentally derived Weber fractions for velocity discrimination under different contrast conditions. Reproduced with permission from Ref. [49]. Part (b) shows simulated discrimination measures using different input magnitudes hypothesized to correspond to different stimulus contrasts. Curves show four input magnitudes: 0.75, 1, 2 and 4. Lower magnitudes have worse discrimination (higher discrimination measure). Note that the curves (especially that for magnitude 4) fluctuate somewhat owing to the coarse quantization of network inputs and scales, which in turn were adopted for computational speed and simplicity.

factors work to ensure that this is not the case. Firstly, the changes in receptor amplitude are limited in their effects by the membrane or shunting properties of the transient cells, which restrict the transient cell output ranges irrespective of their input. Secondly, normalization due to intra-scale competition limits activity at fast speeds, so that activity cannot rise beyond a certain level nor can cells ever respond at speeds beyond some high cut-off.

Since large scales have the highest thresholds and therefore require the greatest input energy to become active, they are the most significantly affected by any changes in input amplitude. Thus, one might expect that, in our model, contrast changes, modeled as changes to this input amplitude, would primarily affect discrimination performance at high speeds where the large scales are active, whereas the data show that performance is diminished at both high and low speeds. However, it is necessary to remember that discrimination performance is based on differential activity between two speeds. The reason that discrimination performance is poor at low and high speeds is that fewer scales are active in these ranges and so any change in input speed results in a smaller change in network activity. As input magnitude is reduced, overall network activity declines (see Fig. 8), resulting in a worsening of discrimination performance at all speeds. Since low and high speed discrimination performance is already poor, it is most significantly affected by this. Fig. 7(b) shows the results of computer simulations with different input amplitudes. The important data properties from Fig. 7(a) are reproduced in the simulation: lowering input magnitude causes a worsening of discrimination at low and high speeds and a shift of the discrimination curve upwards in such a way that the curves tend not to intersect.

13. Contrast effects on perceived speed

Several studies have reported that contrast also affects the perceived speed of moving objects. Thompson

et al. [56] reported that low contrast gratings were perceived to move more slowly than high contrast gratings. Ferrera and Wilson [57] found that contrast influenced the perceived speed of coherent plaid patterns formed from superimposed gratings. Castet et al. [58] found a contrast-induced reduction in perceived speed of translating lines.

The data simulated here are from a study by Stone and Thompson [34] in which subjects compared the speed of two simultaneously presented grating patches, a test and a reference. The contrast of the reference grating was varied and the percentage of test gratings perceived as moving faster were recorded as a function of the test speed for each contrast level. Results (reproduced in Fig. 9(a)) showed shifted psychometric curves such that low contrast gratings were biased towards slow speeds and vice versa.

As indicated in Appendix A, the amplitude of inputs was assumed to covary with contrast. Such would occur if, for example, inputs in our simulations were themselves the output of a shunting center-surround network, which produces peaks whose amplitudes are proportional to contrast near discontinuities of an input pattern [59].

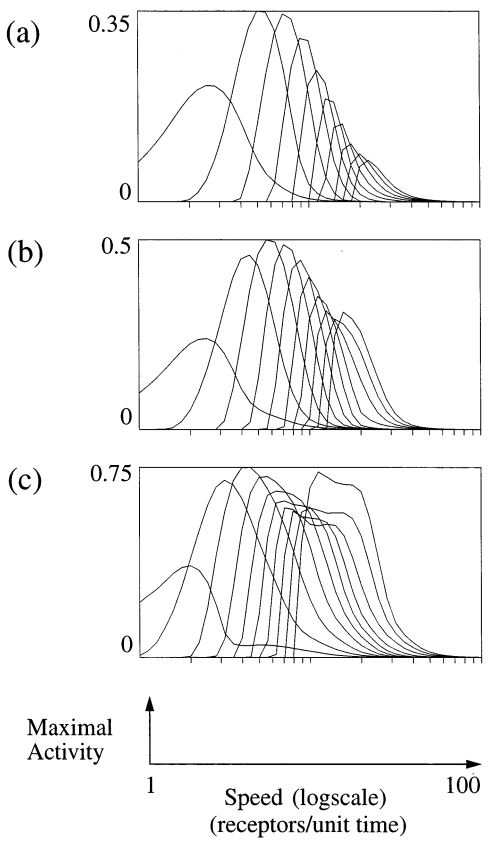

Fig. 8. Maximal simulated cellular activities over speed using different change-sensitive unit activity amplitudes, hypothesized to correspond to different stimulus contrasts. Plot (a) shows a low amplitude unit activity (0.75), plot (b) shows an intermediate amplitude (1) and plot (c) shows high amplitude (2). Increasing unit activity amplitude causes large scales to respond more vigorously and at lower speeds. This biases the network to provide higher speed estimates.

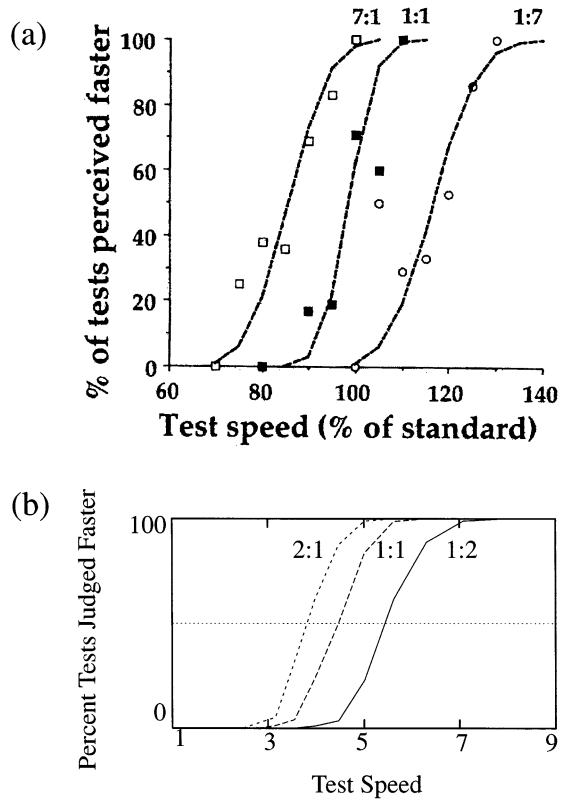

Fig. 9. Plot (a) shows how perceived speed is affected by stimulus contrast. Data reproduced with permission from Ref. [34] shows the percentage of trials on which a test grating was judged to be moving faster than a standard grating as a function of the test grating speed for three different contrast levels expressed as the ratio of the test speed to the reference speed. Plot (b) shows simulated contrast effects on perceived speed in the model. Simulated results were obtained by passing the difference between two speed measures, a test and standard, through an error function. Three different test input contrasts were simulated by varying the change-sensitive unit activity amplitude. The ratio next to each curve indicates the ratio of the test speed to the reference.

The model simulates the change in relative speed judgments due to contrast variations using the previously defined speed measure. Speed judgments were calculated from speed measures obtained from two inputs that were simulated separately. The difference between the two speed measures was passed through a sigmoidal 'error function' to obtain a simulated probability of an observer judging one speed as faster than the other (see Appendix A for details). The results of this process are shown in Fig. 9(b). For two identical inputs of unit amplitude, we obtained a sigmoidal curve. Changing the simulated contrast by varying the receptor activity amplitude causes a shift in the curve similar to that observed in the data, with a greater shift occurring due to increases than decreases of input magnitude. The sigmoidal shape of the curves in Fig. 9(b) results from the error function. The key result of

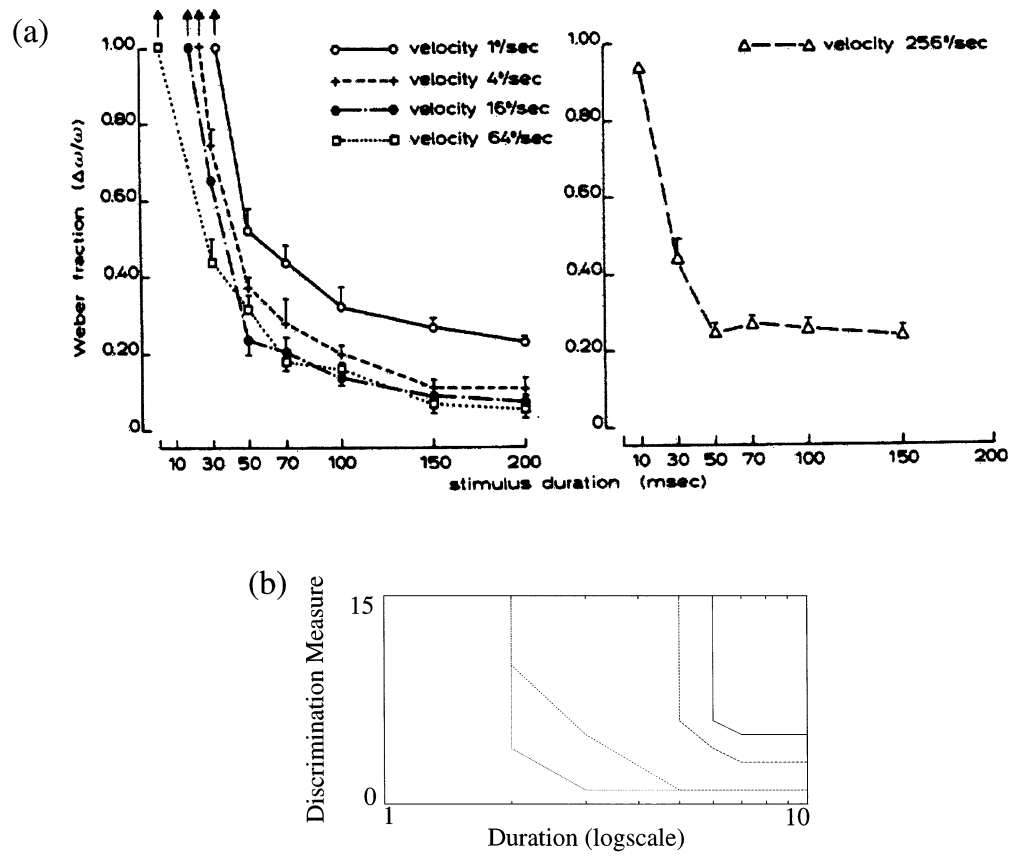

Fig. 10. Plot (a) shows discrimination performance as a function of velocity under different stimulus durations. Data reproduced with permission from Ref. [35]. Plot (b) shows simulated discrimination measures using different input durations. Four curves are shown, each corresponding to a different input speed. From left to right these speeds decrease (10, 5, 2.24, 1.78). Qualitative properties of the data are reproduced: discrimination performance worsens at short durations and this effect is more pronounced for slow speeds.

the simulation is the horizontal shift of model output as stimulus amplitude is varied.

As previously noted, the effects of increasing input magnitude are bounded by the shunting properties of the transient cells. Thus the model predicts that there should exist a saturation level beyond which there exists no discernible effect of increased stimulus contrast. As contrast decreases, perceived speed also decreases until the input energy is sufficiently low as to cause the stimulus to be no longer visible, or visibly in motion. At the same time as the speed measure decreases, the total activity in the network also decreases. We predict that at very low activity levels, speed measures obtained from the network are indistinguishable from noise. Thus, it may not be the case that decreasing stimulus contrast always results in slower perceived speeds; that is, there may be a network energy threshold below which the speed measures are no longer relevant. In summary, the range of stimulus contrasts under which a contrast-induced speed change can be effected in the model is bounded below by the energy present and above by the shunting properties of the transient cells. Stimuli outside this range may not result in contrast-induced changes in speed perception.

14. Duration effects on speed discrimination

De Bruyn and Orban [35] showed an influence of stimulus duration on speed discrimination performance (Fig. 10(a)). At short durations, discrimination performance worsened. This effect was particularly noticeable for slow velocities, as would be expected if the relevant variable is spatio-temporal integration of unit responses.

In any model that requires inputs to traverse a certain distance in order to activate motion-sensitive units, there will be a minimum duration required for the stimulus to traverse this distance. In our model, a number of factors complicate the determination of this minimum distance. Firstly, the spatial scale of cells covaries with their speed tuning. Thus, the minimum duration will not necessarily decrease with input speed. Secondly, it is not necessary for an input to completely traverse a receptive field to activate the cell of a particular scale, as it is in a two-point correlational model such as a Reichardt detector. For example, a scale may be partially activated by an input which partially traverses its receptive field. Because of these complexities, it is not possible to explicitly calculate the minimum

input durations required at certain speeds. Instead, we use simulations to observe the effect of this parameter on model performance.

Fig. 10(b) shows the effect of changing model input duration at various simulated speeds has the same qualitative properties as data in the Fig. 10(a). At low speeds, the input traverses a small distance in a given time, so any reduction of input duration can prevent activation of mid- to large-scale units, causing a deterioration in discrimination performance. At high speeds, input duration is of less importance, as the input will traverse the input field of each scale even at very short durations.

15. Dot density effects on perceived speed

Watamaniuk et al. [36] reported that increasing the density of a field of moving dots increases the perceived speed of these dots (Fig. 11(a)). In the same paper, this result was modeled using motion coherence theory [60],

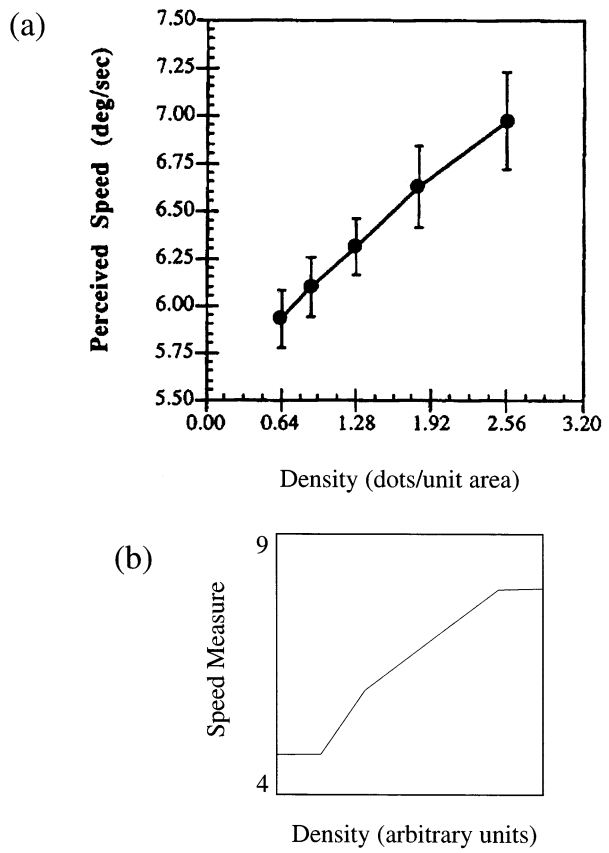

Fig. 11. Plot (a): perceived speed of moving random dot patterns is affected by dot density. Data reproduced with permission from Ref. [36] shows that the perceived speed of a random dot pattern (comparison stimulus) increases with its density. Plot (b): decreasing mean arrival time between successive inputs, hypothesized to correspond to increasing density, increases simulated speed measures in the network. In the simulation, this effect can be observed only over a range of densities.

in which a smoothing process integrates the responses of primitive motion detectors across space. The presence of additional dots causes this smoothing stage to provide a higher estimate of the image speed by providing more evidence for that speed in the smoothing stage. In our model, the presence of additional dots provides more inputs to the network, causing an increase in perceived speed before this information is integrated across later mechanisms.

To simulate such a random dot paradigm using a single one-dimensional network, we assumed that a series of dots provide input to the network as they traverse the change-sensitive units. An increase in dot density is simulated by a decrease in the mean time between the appearance of these dots, which decreases the mean time between the responses of the change-sensitive units. Given a dot density, a fixed width for each receptor, and the input speed, the expected time between the arrival of two successive dots at a receptor was calculated. Units were then caused to provide input pulses with a frequency corresponding to that expected time. Note that since our simulations are one-dimensional, they do not code the distinction between a dot moving horizontally and a vertical bar moving horizontally. Chey et al. [46] implement a two-dimensional extension of the present model that simulates data in which direction is crucial, such as the barberpole illusion and plaid coherence data, without undermining the simulations reported herein.

More frequent receptor responses (greater numbers of dots) increase the responses of the larger scales by providing more input energy and thus lead to an increase in perceived speed, just as greater receptor amplitude (greater contrast) did. Again, however, several factors limit these effects. Saturation of transient cell responses limits the energy obtainable from any one transient cell, just as it did in simulations using greater receptor activation amplitude. However, in these simulations, it is not the activity at any one transient cell that increases input energy, but rather the increase in the number of simultaneously active transient cells. This effect is limited by the inter-scale competition, which suppresses activity when cells are active at the same time as their neighbors. Fig. 11(b) shows that, as a result, there is a range over which changes in density affect speed judgments in the simulation. On either side of this range the speed measure asymptotes. At low densities, increasing density has no effect, as activity from previous dots has already dissipated by the time the next dot arrives. At high densities, additional dots do not increase response, as transient cell response saturation and inter-scale competition limit activity. Thus, simulated dot density effects, like simulated contrast effects, are observable only in a limited range. Watamaniuk et al. [36] reported results from a range where density effects could be observed. It is not clear from their data what the limits of this range are.

16. Spatial frequency effects on perceived speed

It is difficult to isolate the effect of spatial frequency on speed estimates, since this factor may well help to maintain velocity constancy through its influence on perceived depth. Therefore, it is not surprising that a number of different claims have been made regarding changes in speed perception in response to spatial frequency variations. Diener et al. [61] reported that increasing the spatial frequency of sinusoidal gratings increased perceived velocity. However, this result was obtained only with very low spatial frequencies (below 0.07 cd). In a later study, Smith and Edgar [62] found the opposite effect; namely, that increasing spatial frequency led to decreases in perceived velocity. The model of Johnston et al. [63] predicts this effect. Campbell and Maffei [64] reported that the perceived speed of rotating gratings increased with spatial frequency up to 4 cd and then decreased, and Ferrera and Wilson [57] found that the spatial frequency of sinusoidal gratings and perceived speed co-varied. Since each of these results was obtained under a different experimental paradigm, it is difficult to compare them. In our model, increasing spatial frequency, while maintaining a constant stimulus size, causes more frequent receptor responses and therefore increases network response in a manner similar to increasing dot density; that is, there is saturation as very high spatial frequencies are used.

There is currently, however, no explanation in our model for a decrease in perceived speed at high spatial frequencies unless the width of the stimulus also changes. This may be because the model currently simulates only how ON cell responses are processed, since these are sufficient to explain many speed perception data. An extended version of the model simulates data for which OFF cell responses are also important [65], such as second-order motion percepts. To the extent to which OFF cell responses to high spatial frequency stimuli inhibit larger scales, then a decrease in perceived speed due to relatively greater activation of smaller scales could be explained. Such a decrease would depend on a change in the overall balance of ON and OFF cell responses with spatial frequency.

17. Reaction time

Reaction time to stimulus onset is a decreasing function of stimulus velocity [37,66,67]). Mashhour [37] fit this function with an expression of the type:

$$RT = c/V^n + RT_0$$

where RT is the reaction time, RT_0 is the asymptotic reaction time (for fast velocities), V is the stimulus velocity and both c and n are empirically derived constants. This is an exponentially decreasing function that asymptotes at RT_0.

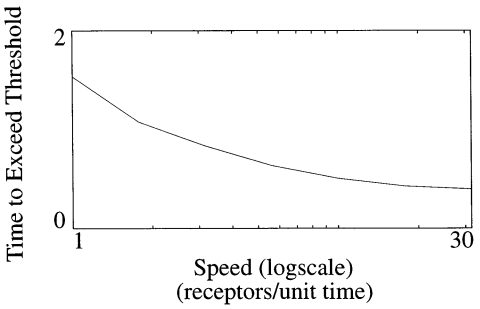

Fig. 12. Time taken for activity to exceed a threshold as a function of stimulus speed, hypothesized to correspond to the speed variable component of reaction time to motion onset. This is a decreasing function of speed that asymptotes at fast speeds.

We simulated the time-variant component of reaction time by computing the time taken for activity at any scale to exceed a threshold after a stimulus starts moving. As stimulus speed increases, transient cell responses have greater temporal overlap. This causes an increase in input amplitude to the short-range spatial filters and speeds their averaging rate. Faster short-range filter averaging leads to a decrease in time taken for a filter to exceed threshold. Thus, we predict a decrease in simulated reaction time in the model. Fig. 12 shows that the simulated reaction does decrease in an exponential manner with stimulus speed. The asymptotic behaviour of this measure is caused by the shunting properties of the transient cells and the limited overlap possible between transient cell responses.

18. Discussion

Several other methods by which the human visual system might extract speed estimates have been proposed. Correlational models, such as the Reichardt detector, incorporate speed-selectivity into directionally-selective motion sensors. Human speed perception has been modeled by the elaborated Reichardt detector [2,3], which differs from the original Reichardt formulation in that preliminary spatial filtering is performed by the receptors before their outputs are multiplied. A related model is the motion energy model [1] in which temporal filtering takes the place of the delay. Watson and Ahumada [4] proposed that the temporal response patterns of directionally-selective sensors are used to derive speed estimates.

We call the model of early motion mechanisms described in this article the MOC filter, because units are responsive to motion of oriented contrasts. The MOC filter model differs from other formulations in that it starts with unoriented transient cell responses and builds directionality and speed sensitivity from these. Fay and Waxman [68] also used such a system. They measured speed from the shape of convected activation

profiles generated by transient detectors. A fundamental difference between all of these models and the MOC filter model is that the MOC filter explains contrast-dependent and other non-speed parametrically induced variations in perceived speed, whereas other formulations are designed to minimize such effects. Adelson and Bergen [1] introduce an additional contrast normalization that will produce contrast-dependent effects in certain situations. In addition, as noted in the Introduction, the MOC filter simulates many other data about motion perception that the correlational models cannot explain.

The MOC filter model postulates that speed tuning is an emergent property of spatio-temporal network interactions across a series of network processing stages; it is not explicitly defined by any one operation. Spatial and temporal averaging across multiple spatial scales enable the network, as a whole, to begin the transformation from local temporal properties of moving inputs to a global spatial map that computes a variety of motion properties, including input speed. Spatio-temporal averaging alone is, however, not sufficient to generate true speed tuning, since the larger scales then always respond more to all speeds. Self-similar thresholds and competition within and across scales lead to true speed-tuned receptive fields. This approach to modeling motion and speed perception avoids the danger that miscalibrated delays across motion detectors could yield biases in individual speed estimates. Instead, the collective responses of multiple receptive field sizes generates a spatial pattern of activation that may be used to vote for the most robust speed estimate. Such a multiple-scale mechanism also enables the cortical magnification factor to naturally be used to explain the large range over which humans can discriminate input speeds.

Another major difference between the MOC filter and correlational models is that the former does not presume that only two spatial locations are used in the correlational process. Instead, all intermediate locations participate in determining the final speed percept. This mechanism provides for more robust computation in the presence of cortical noise, and may help to explain how sampling at additional spatial locations improves the quality of apparent motion percepts, as Nakayama and Silverman [13] have reported.

A key feature of the MOC filter model is that, in addition to moment-by-moment speed estimates, it provides a continuously evolving set of speed estimates at every location. For example, when a fast stimulus starts its motion across a simulated series of cells, the initial readings are of slow motion and only over time do the larger scales become active and signal faster speeds. Although data are referenced above describing a loss of discrimination at short durations, these data do not address a speed bias that may be present at such durations.

The data discussed in the context of our simulations support the hypothesis that spatial parameters of speed-tuned units co-vary with their speed tuning. Others have attempted to measure the spatial (and temporal) characteristics of elementary motion detectors in different ways. For example, van Doorn and Koenderink [69,70] looked for the spatial or temporal points that defined the transition between coherent and incoherent motion with spatially or temporally alternated motion directions, respectively. The points at which this transition took place were taken to reflect the characteristics of the detector and varied with stimulus velocity, suggesting different spatial and temporal characteristics of different velocity-tuned mechanisms. The MOC filter currently contains no explicit variation of temporal properties (such as averaging rate) with scale. Different scales nonetheless respond at different rates due to their spatial properties and their interactions with neighboring units (see Fig. 13). Thus, one does not need to vary the temporal processing rates of the cells simply to achieve speed tuning, and the wide range of observed speeds can be obtained through use of cortical magnification to control spatial scale as a function of eccentricity.

Johnston and Clifford [71] have developed an alternative approach to modeling motion perception. This model is based upon formal Taylor series of image brightness around a point of interest. These expansions are used in conjunction with integral operations to provide 'a least squares estimate of image speed based on measures of how the image brightness and its derivatives are changing with respect to space and time'. The present approach directly develops a neural model of the magnocellular brain mechanisms that subserve motion perception. It is not yet clear how the two approaches can be linked. The two models do share a key property, however: both attempt to explain key first-order and second-order motion percepts using a single processing stream. Baloch et al. [65] show how mechanisms of the present model can be used to simulate both first-order and second-order motion percepts.

A more complex model of speed perception will necessarily include mechanisms for grouping and com-

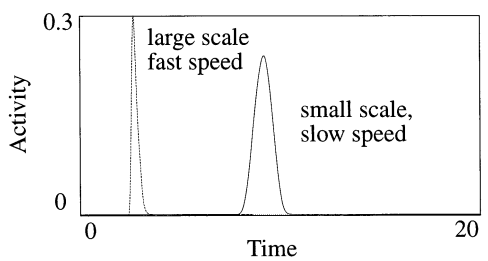

Fig. 13. Responses of two different cells at different scales (left is scale 1 and right is scale 5) and different speeds (left is speed 2.24 and right is speed 10) show that response period or apparent rate can vary without explicit variation of the time-averaging rates of the cells.

pleting motion direction and speed signals across space. Only after such global processes occur do coherent percepts of object direction and speed emerge. Typical examples include how speed and direction percepts both change when plaid patterns are perceived either as independently moving components or as a coherently moving plaid [57,72,73]. Chey et al. [46] have modeled how output signals from our MOC filter input to such a global motion grouping network. This extended model proposes a solution of the global aperture problem, wherein unambiguous feature tracking signals are used to capture and transform ambiguous motion signals to generate a coherent representation of an object's direction and speed of motion. The model's central problem is to understand what type of feature tracking process can select an unambiguous object direction without distorting the estimates of object speed that are computed by the MOC filter. The resultant model, which is called the motion boundary contour system, or BCS, simulates data concerning the conditions under which components of moving stimuli cohere or not into a global direction of motion, as in barberpole and plaid patterns (both Type 1 and Type 2), and how the perceived speed of lines moving in a prescribed direction depends upon their orientation, length, duration, and contrast.

The motion grouping, or capture, network of the motion BCS involves a feedback process that allows attention to prime a desired motion direction. In other words, motion capture, which seems to be an automatic and preattentive process, may be carried out by the same circuit that permits top-down attention to selectively focus on a desired motion direction [74–76]. Baloch and Grossberg [77] have suggested how this process can help to explain the interplay of preattentive and attentive processes during percepts like the line motion illusion [78,79] and motion induction [80–82]. Thus the simple multiple-scale filtering and competition circuits simulated here have already provided a foundation for building a more general neural theory of motion perception.

It is of considerable interest that similar multiple-scale filtering and competition mechanisms model the size-disparity correlation that is used to explain how three-dimensional forms pop-out from their backgrounds [40]. Taken together, these results suggest that the brain may utilize a similar multiple-scale filtering strategy in both the V1→V2 and V1→MT cortical processing streams. This hypothesis is consistent with data showing that cells of primate MT exhibit sensitivity to disparity [32,83,84].

The multiple-scale organization of both streams has elsewhere been used to model how the V2→MT cross-stream interaction gives rise to percepts wherein representations of object form that emerge in the V1→V2→V4 stream are seen to move in the V1→MT→MST stream [17,77]. The present model of motion speed filtering thus has the dual advantages of simplicity and embeddability with a larger theory of 3-D form and motion perception.

Acknowledgements

JC supported in part by DARPA (ONR N00014-92-J-4015), the National Science Foundation (NSF IRI-90-00530) and the Office of Naval Research (ONR N00014-91-J-4100). SG supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0490), DARPA (ONR N00014-92-J-4015), and the Office of Naval Research (ONR N00014-91-J-4100, ONR N00014-95-1-0657, ONR N00014-95-1-0409). EM supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0334) and the Office of Naval Research (ONR N00014-94-1-0597, ONR N00014-95-1-0409). The authors thank Diana Meyers for her valuable assistance in the prepartion of the manuscript.

Appendix A. Network equations and parameters

The model is defined using differential equations. Each equation specifies the time varying activity of a cell, or cell population, within a level. The activity of each such cell is represented by a variable whose letter indicates the level in which that cell is located and whose subscript indicates the cell's position within that level and, if necessary, its scale. Spatial locations are indexed in numerical order within each level, so that a cell's number indicates its absolute position in a one-dimensional grid in its level. Simulations were conducted by numerically integrating these equations using Euler's method with a time step of 0.01.

A.1. Level 1: Change-sensitive units

Assume the stimulus to be moving at velocity v. Change-sensitive units are activated for a fixed time period when an input enters their receptive field. The ith change-sensitive receptor activity x_i obeys the equation:

$$x_i = \begin{cases} \eta & \text{if } \dfrac{i}{v} \le t \le \left(\dfrac{i}{v} + \varepsilon\right) \\ 0 & \text{otherwise} \end{cases} \tag{A1}$$

The magnitude of the activity, η is assumed to co-vary with the stimulus contrast (see Ref. [65] for details). The duration of the activity is given by the constant ε.

Eq. (A1) is not intended to be a realistic model of the temporal impulse response of photoreceptors, and is employed only for simplicity. Our model works despite, not because of, the waveform it generates, which makes it harder for our model, or any model, to generate smoothly modulated speed profiles.

A.2. Level 2: Transient cells

Transient cells space-average and time-average signals from the change-sensitive units through a membrane shunting equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} y_i = -y_i + (1 - y_i) \sum_{j \in X_i} x_j \tag{A2}$$

The ith transient cell activity y_i has several important features: its response rate $(1 + \Sigma_{j \in X_i} x_j)$ increases with total receptor activity, and its amplitude is bounded by 1 for any receptor activity. The set X_i denotes change-sensitive units from which the ith subunit draws its input. Such a set consists of a series of adjacent cells that do not intersect any other set X_j, $j \neq i$.

A.3. Level 3: Self-similar short-range spatial filter

Each scale of the short-range spatial filter space-averages and time-averages transient cell activity over a different range. Each filter scale is represented at every position. For example, scale 1 spatial filters draw input from a transient cell at the same position in the previous level, scale 2 draws input from the same cell and the two adjacent cells, scale 3 from those and the next two adjacent and so on. The activity, z_{si}, of the short-range spatial filters obeys:

$$\dot{z}_{si} = 10\left(z_{si} + \sum_{j \in Y_{si}} y_j \right) \tag{A3}$$

Here the input set Y_{si} varies with both location, i and scale, s. The output threshold of each filter increases with scale:

$$o_{si} = [z_{si} - s]^+ \tag{A4}$$

This thresholded activity forms the input to the next level.

A.4. Level 4: Intra-scale competition

Intra-scale competition occurs across spatial positions within each scale. The center-surround mechanism is implemented as follows:

$$\dot{u}_{si} = 10\left(-u_{si} + \sum_{j \in C_i} o_{sj} - \sum_{k \in S_i} o_{sk} \right) \tag{A5}$$

The sets C_i and S_i define the excitatory center and inhibitory surround, respectively.

A.5. Level 5: Inter-scale competition

A similar equation is used for the inter-scale competition. The inputs from the previous level, are thresholded (to keep them from being negative) and raised to the third power, which causes the stimulations of Eq. (A6) to equilibrate more rapidly than would otherwise be the case.

$$\dot{w}_{si} = -w_{si} + (1 - w_{si}) \left(\sum_{r \in D_s} [u_{ri}]^+ \right)^3 - (1 + w_{si}) \left(\sum_{r \in E_s} [u_{ri}]^+ \right)^3 \tag{A6}$$

The center and surround are defined by the sets D_s and E_s.

A.6. Speed measure

The estimated speed is obtained from the weighted sum over the entire set of scales S:

$$\mathrm{speed}_i = \frac{\sum_{s \in S} s[w_{si}]^+}{\sum_{s \in S} [w_{si}]^+} \tag{A7}$$

A.7. Discrimination measure

The discrimination measure is obtained by finding the minimum number of speed increments or decrements that cause the total difference between activity at each scale to exceed a threshold. Let the superscript indicate a speed at which an activity is obtained, then the total difference in activity between two speeds is defined as:

$$\mathrm{difference}_{\mathrm{p,q}} = \sum_{s \in S} \left| w_{si}^p - w_{si}^q \right| \tag{A8}$$

A.8. Error function

The error function used for the simulations in Fig. 9 is erf (speed$_i$), where:

$$\mathrm{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-(1/2)x^2} \, \mathrm{d}x \tag{A9}$$

A.9. Simulation parameters

All parameters were kept the same during the simulations reported in this paper. Only input magnitude was varied. That is, η in Eq. (A1) was set to 1 for all simulations except that of Fig. 9, where values of 0.5, 1, and 2 were used. Every simulation used 15 spatial locations and activities were measured from the middle location unless explicitly stated otherwise. Ten scales were used (so s, which gives the value of the thresholds

used in Eq. (A4) as well as indexing equations for the various scales, ranged from 1 to 10). In Eq. (A1), ε was set to 1 for all simulations.

The various sets that determine which units form input to cells were established as

$$X_i = \{j:(i-1)\times 10 \le j < i\times 10\} \quad \text{(A10)}$$

$$Y_{si} = \{j:|i-j| \le s\} \quad \text{(A11)}$$

$$C_i = \{j:|i-j| \le 2\} \quad \text{(A12)}$$

$$S_i = \{j:|i-j| \le 3\} \quad \text{(A13)}$$

$$D_s = \{s\} \quad \text{(A14)}$$

$$E_s = \{t:t \ne s\} \quad \text{(A15)}$$

References

[1] Adelson EH, Bergen JR. Spatiotemporal energy models for the perception of motion. J Opt Soc Am 1985;2(2):284–99.

[2] van Santen JPH, Sperling G. Temporal covariance model of human motion perception. J Opt Soc Am 1984;1(5):451–73.

[3] van Santen JPH, Sperling G. Elaborated Reichardt detectors. J Opt Soc Am 1985;2(2):300–21.

[4] Watson B, Ahumada AEJ Jr. Model of human visual-motion sensing. J Opt Soc Am 1985;2(2):322–42.

[5] Allman J, Miezin F, McGuinness E. Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). Perception 1985;14:105–26.

[6] Maunsell JHR, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus duration, speed, and orientation. J Neurophysiol 1983;49(5):1127–47.

[7] Mikami A, Newsome WT, Wurtz RH. Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. J Neurophysiol 1986;55(6):1328–39.

[8] Zeki SM. Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. J Physiol (London) 1974;236:546–73.

[9] Braddick OJ. A short-range process in apparent motion. Vis Res 1974;14:519–27.

[10] Anderson SJ, Burr DC. Spatial and temporal selectivity of the human motion detection system. Vis Res 1985;25:1147–54.

[11] Anderson SJ, Burr DC. Receptive field properties of human motion detection units inferred from spatial frequency masking. Vis Res 1989;29:1343–58.

[12] Burr DC, Ross J, Morrone MC. Smooth and sampled motion. Vis Res 1986;26:643–52.

[13] Nakayama K, Silverman GH. Temporal and spatial characteristics of the upper displacement limits for motion in random dots. Vis Res 1984;24:293–9.

[14] Nakayama K, Silverman GH. Detection and discrimination of sinusoidal grating displacements. J Opt Soc Am A Optics Image Sci 1985;2:267–73.

[15] Petersik JT, Pufahl R, Krasnoff E. Failure to find an absolute retinal limit of a putative short-term process in apparent motion. Vis Res 1983;23:1663–70.

[16] Kolers. Aspects of Motion Perception. Oxford:Pergamon Press, 1972.

[17] Francis G, Grossberg S. Cortical dynamics of form and motion integration: Persistence, apparent motion, and illusory contours. Vis Res 1996;36(1):149–73.

[18] Grossberg S, Rudd ME. A neural architecture for visual motion perception: Group and element apparent motion. Neural Networks 1989;2(6):421–50.

[19] Grossberg S, Rudd ME. Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. Psychol Rev 1992;99(1):78–121.

[20] Heeger D. Model for the extraction of image flow. J Opt Soc Am 1987;A4(8):1455–71.

[21] Anadan P. A computational framework and an algorithm for the measurement of visual motion. Int J Comput Vis 1989;2:283–310.

[22] Duysens J, Orban GA, Verbeke O. Velocity sensitivity mechanisms in cat visual cortex. Exp Brain Res 1982;45:285–94.

[23] Goodwin AW, Henry GH, Bishop PO. Direction selectivity of simple striate cells: Properties and mechanisms. J Neurophysiol 1975;38:1500–23.

[24] Goodwin AW, Henry GH. Direction selectivity of complex cells in a comparison with simple cells. J Neurophysiol 1975;38:1524–40.

[25] Orban GA, Kennedy H, Maes H. Response to movement of neurons in areas 17 and 18: Velocity sensitivity. J Neurophysiol 1981;45:1043–58.

[26] Cheng K, Hasegawa T, Kadharbatcha SS, Tanaka K. Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. J Neurophysiol 1994;71(6):2269–80.

[27] Anderson SJ, Burr DC. Spatial summation properties of directionally selective mechanisms in human vision. J Opt Soc Am 1991;A8(8):1330–9.

[28] Anderson SJ, Burr DC. Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision. J Opt Soc Am 1991;A8(8):1340–51.

[29] Goodwin AH, Henry GH. The influence of stimulus velocity on the responses of single neurons in the striate cortex. J Physiol (London) 1978;277:467–82.

[30] Movshon JA. The velocity tuning of single neurons in the striate cortex. J Physiol (London) 1975;249:445–68.

[31] Orban GA, Kennedy H, Bullier J. Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: Influence of eccentricity. J Neurophysiol 1986;56(2):462–80.

[32] Maunsell JHR, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Binocular interactions and sensitivity to binocular disparity. J Neurophysiol 1983;49(5):1148–67.

[33] Snowden RJ, Treue S, Andersen RA. The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. Exp Brain Res 1992;88:389–400.

[34] Stone LS, Thompson P. Human speed perception is contrast dependent. Vis Res 1992;32(8):1535–49.

[35] De Bruyn B, Orban GA. Human velocity and direction discrimination measured with random dot patterns. Vis Res 1988;28(12):1323–35.

[36] Watamaniuk SNJ, Grzywacz NM, Yuille AL. Dependence of speed and direction perception on cinematogram dot density. Vis Res 1993;33(5/6):849–59.

[37] Mashhour M. Psychophysical Relations in the Perception of Velocity. Stockholm: Almquist and Wiksell, 1964.

[38] Duysens J, Orhan GA, Cremieux J. Velocity selectivity in the cat visual system. II. Independence from interactions between different loci. J Neurophysiol 1985;54(4):1050–69.

[39] Marr D, Poggio T. A computational theory of human stereo vision. Proc R Soc London B 1979;204:301–28.

[40] Grossberg S. 3-D vision and figure-ground separation by visual cortex. Perception Psychophys 1994;55:48–120.

[41] Julesz B, Schumer RA. Early visual perception. Annu Rev Psychol 1981;32:572–627.

[42] Schor CM, Wood I. Disparity range for local stereopsis as a function of luminance spatial frequency. Vis Res 1983;23:1649–54.
[43] Schor CM, Wood I, Ogawa J. Binocular sensory fusion is limited by spatial resolution. Vis Res 1984;24:661–5.
[44] Tyler CW. Spatial organization of bionocular disparity sensitivity. Vis Res 1975;15:583–90.
[45] Tyler CW. Sensory processing of binocular disparity. In: Schlor CM, Cuiffreda KJ, editors. Vergence Eye Movements. Boston: Butterworths, 1983:199–295.
[46] Chey, J, Grossberg, S, and Mingolla, E. Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. J Opt Soc Am, A 1997;14:2570–94.
[47] Enroth-Cugell C, Robson JG. The contrast sensitivity of retinal ganglion cells of the cat. J Physiol (London) 1966;187:517–52.
[48] Gouras P. Identification of cone mechanisms in monkey ganglion cells. J Physiol 1968;199:533–47.
[49] Orban GA, Wolf JD, Maes H. Factors influencing velocity coding in the human visual system. Vis Res 1984;24(1):33–9.
[50] Daniel PM, Whitbridge D. The representation of visual field on the cerebral cortex in monkeys. J Physiol 1961;159:201–21.
[51] Hubel DH, Wiesel TN. Functional architecture of macaque visual cortex. Proc R Soc London Ser B 1977;198:1–59.
[52] Schwartz EL. Computational anatomy and functional architecture of visual cortex: A spatial mapping approach to perceptual coding. Vis Res 1980;20:645–69.
[53] Schwartz EL. Spatial mapping and spatial vision in primate striate and inferotemporal cortex. In: Wooten BR, editor. Sensory Experience, Adaptation and Perception. Hillsdale, NJ: Lawrence Erlbaum, 1984:73–104.
[54] Burr DC, Ross J. Contrast sensitivity at high velocities. Vis Res 1982;22:479–84.
[55] McKee SP, Silverman GH, Nakayama K. Precise velocity discrimination despite random variations in temporal frequency and contrast. Vis Res 1986;26:609–19.
[56] Thompson P. Perceived rate of movement depends on contrast. Vis Res 1982;22:377–80.
[57] Ferrera VP, Wilson HR. Perceived direction of moving two-dimensional patterns. Vis Res 1990;30(2):273–87.
[58] Castet E, Lorenceau J, Shiffrar M, Bonnet C. Perceived speed of moving lines depends on orientation, length, speed and luminance. Vis Res 1993;33(14):1921–36.
[59] Grossberg S. The quantized geometry of visual space: the coherent computation of depth, form, and lightness. Behav Brain Sci 1983;6(4):625–57.
[60] Yuille AL, Grzywacz NM. A computational theory for the perception of coherent visual motion. Nature 1988;333:71–4.
[61] Diener HC, Wist ER, Dichgans J, Brandt T. The spatial frequency effect on perceived velocity. Vis Res 1976;16:169–76.
[62] Smith AT, Edgar GK. The influence of spatial frequency on perceived temporal frequency and perceived speed. Vis Res 1990;30(10):1467–74.
[63] Johnston A, McOwan PW, Buxton H. A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. Proc R Soc London B 1992;250:297–306.
[64] Campbell FW, Maffei L. The influence of spatial frequency and contrast on the perception of moving patterns. Vis Res 1981;21:713–21.
[65] Baloch, A, Grossberg, S, Mingolla, E and Nogueira, C. A neural model of first-order and second-order motion perception and magnocellular dynamics. Technical Report CAS/CNS-96-30. Boston, MA: Boston University, 1996.
[66] Brown K, Sekuler R. Models of stimulus uncertainty in motion perception. Psychol Rev 1980;87:435–69.
[67] Tynan PD, Sekuler R. Motion processing in peripheral vision: Reaction time and perceived velocity. Vis Res 1982;22:61–8.
[68] Fay DA, Waxman AM. Neurodynamics of real-time image velocity extraction. In: Carpenter GA, Grossberg S, editors. Neural Networks for Vision and Image Processing. Cambridge, MA: MIT Press, 1992:221–46.
[69] van Doorn AJ, Koenderink JJ. Spatial properties of the visual detectability of moving spatial white noise. Exp Brain Res 1982;45:189–95.
[70] van Doorn AJ, Koenderink JJ. Temporal properties of the visual detectability of moving spatial white noise. Exp Brain Res 1982;45:179–88.
[71] Johnston A, Clifford CWG. A unified account of three apparent motion illusions. Vis Res 1995;8:1109–23.
[72] Adelson EH, Movshon JA. Phenomenal coherence of moving visual patterns. Nature 1982;300:523–5.
[73] Ferrera VP, Wilson HR. Perceived speed of moving two-dimensional patterns. Vis Res 1991;31(5):877–93.
[74] Groner R, Hofer D, Groner M. The role of anticipation in the encoding of motion signals-sensitization or bias. In: Klix F, Hagendorf, editors. Human Memory and Cognitive Capabilities. Amsterdam: Elsevier, 1986.
[75] Sekuler R, Ball K. Mental set alters visibility of moving targets. Science 1977;198:60–2.
[76] Stelmach LB, Herdman CM, McNeil R. Attentional modulation of visual processes in motion perception. J Exp Psychol: Hum Perception Performance 1994;20:108–21.
[77] Baloch, A and Grossberg, S. A neural model of high-level motion processing: Line motion and formotion dynamics. Technical Report CAS/CNS-96-20. Boston, MA: Boston University, 1996. Vis Res 1997;37:3037–59.
[78] Hikosaka O, Miyauchi S, Shimojo S. Focal visual attention produces illusory temporal order and motion sensation. Vis Res 1993;33(9):1219–40.
[79] Hikosaka O, Miyauchi S, Shimojo S. Voluntary and stimulus-induced attention detected as motion sensation. Perception 1993;22:517–26.
[80] Faubert J, von Grünau M. The extent of split attention and attribute priming in motion induction. Perception 1992;21:105.
[81] Faubert J, von Grünau M. The influence of two spatially distinct primers and attribute priming on motion induction. Vis Res 1995;35(22):3119–30.
[82] von Grünau M, Faubert J. Intraattribute and interattribute motion induction. Perception 1994;23:913–28.
[83] Bradley DC, Qian N, Andersen RA. Integration of motion and stereopsis in middle temporal cortical area of macaques. Nature 1995;373:609–11.
[84] Carman GJ, Albright TD. Representation of depth from motion and stereo cues in macaque cortical visual area MT. Soc Neurosci Abstr 1993;19:1283.