



PERGAMON

AVAILABLE AT
www.ComputerScienceWeb.com

POWERED BY SCIENCE @ DIRECT®

Neural Networks 16 (2003) 939–945

Neural
Networks

www.elsevier.com/locate/neunet

2003 Special issue

Neural models of motion integration and segmentation

Ennio Mingolla

Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215, USA

Abstract

A neural model is developed of how motion integration and segmentation processes compute global motion percepts. Figure-ground properties, such as occlusion, influence which motion signals determine the percept. For visible apertures, a line's extrinsic terminators do not specify true line motion. For invisible apertures, a line's intrinsic terminators create veridical feature tracking signals, which are amplified before they propagate across space and are integrated with ambiguous motion signals within line interiors. This integration process is the result of several processing stages: directional transient cells respond to image transients and input to a directional short-range filter that selectively boosts feature tracking signals. Competitive interactions further boost feature tracking signals and create speed-selective receptive fields. A long-range filter gives rise to true directional cells by pooling signals over multiple orientations and opposite contrast polarities. A distributed population code of speed tuning realizes a size–speed correlation, whereby activations of multiple spatially short-range filters of different sizes are transformed into speed-tuned cell responses. These mechanisms use transient cell responses, output thresholds that covary with filter size, and competition. The model reproduces empirically derived speed discrimination curves and simulates data showing how visual speed perception and discrimination are affected by stimulus contrast.

© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: Neural model; Contrast-induced speed perception; Motion integration

1. Motion integration and segmentation

Perception of moving objects in cluttered scenes requires coordinated processes to solve the complementary problems of motion integration and motion segmentation (Braddick, 1993). The present treatment follows that of Grossberg, Mingolla, and Viswanathan (2001), who describe how the integration process joins disparate motion signals that belong to the same object, while the segmentation process keeps separate nearby signals that belong to different objects.

Wallach (1935) (translated by Wuerger, Shapley, and Rubin) first noted that the motion of a featureless line viewed through a circular aperture is ambiguous, and by default perceived as in a direction normal to the line's orientation. This so-called aperture problem confronts any localized detector, such as a cortical neuron with a localized receptive field (Marr & Ullman, 1981). Only when the contour within an aperture contains distinct features, such as a line end, corner, or a discrete blob or dot can a local detector accurately measure both direction and speed of motion. If, on the other hand, a partially occluded object moves in such a way that each of several measurable local motions is subject to the aperture problem, it may be that a process of integration operating at a spatial scale larger than

that of the receptive fields of individual motion detectors is still able to determine the velocity of coherent object motion from the fragmented information.

To solve the twin problems of motion integration and segmentation, the visual system needs to use the relatively few unambiguous motion signals arising from image features to veto and constrain the more numerous ambiguous signals from contour interiors. In addition, the visual system uses contextual interactions to compute a consistent motion direction and velocity when the scene is devoid of any unambiguous motion signals.

The motion of a grating of parallel lines seen moving behind a circular aperture is ambiguous. However, when two such gratings are superimposed to form a plaid, the perceived motion is not ambiguous. Plaids have therefore been used extensively to study motion perception. Three major mechanisms for the perceived motion of coherent plaids have been presented in the literature. (1) *Vector average*. The vector average solution is one in which the velocity of the plaid appears to be the vector average of the normal components of the plaid's constituent gratings. (2) *Intersection of constraints*. A constraint line, first defined by Adelson and Movshon (1982), is the locus in velocity space of all possible positions of the leading edge of a bar or line after some time interval (Fig. 1). The constraint line for

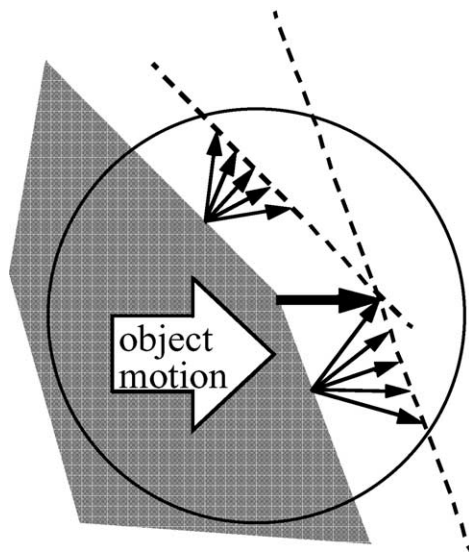


Fig. 1. Viewed through a circular aperture, the two leading edges of a polygon have different normal velocities. Dashed 'constraint lines' indicate the locus of all locations where those leading edges might appear after a small time increment. Their intersection represents the 'intersection of constraints' solution, which of necessity corresponds to the trajectory of the corner feature. Reprinted with permission from [Mingolla et al. \(1992\)](#).

a featureless bar, or a grating of parallel featureless bars, that is moving behind a circular aperture is parallel to the bar. The authors suggested that the perceived motion of a plaid pattern was defined by the velocity vector of the intersection in velocity space of the constraint lines of the plaid components. They named this the intersection of constraints (IOC) solution to the plaid problem. The IOC solution is the mathematically correct solution to the motion perception problem and, hence, is always veridical. However, as noted below, it does not always predict human motion perception even for coherent plaids. (3) *Feature tracking*. When two one-dimensional gratings are superimposed, they form intersections, which act as features whose motion can be reliably tracked. Other features are line endings and object corners. A third possible solution to the problem of plaid motion perception is that the visual system may be tracking features instead of computing a vector average or an IOC solution. At intersections or object corners, the IOC solution and the trajectory of the feature are always identical. However, in some non-plaid displays described below, the feature tracking solution differs from the IOC solution.

[Ferrera and Wilson \(1990, 1991\)](#) noted that perception of motion for certain plaids, for which the IOC prediction falls outside the arc formed by motion vectors for the two components, is not always in accordance with the IOC prediction. [Mingolla, Todd, and Norman \(1992\)](#) introduced a multi-aperture paradigm in order to combine two component motions without the benefit of trackable intersection features, and found that motion was thereby biased away from the IOC solution and toward the vector

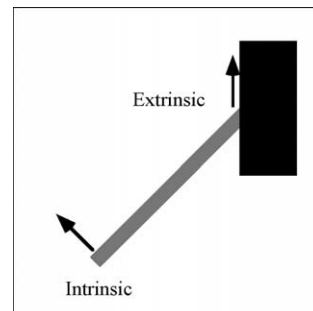


Fig. 2. Extrinsic vs. intrinsic terminators: the boundary that is caused due to the occlusion of the gray line by the black bar is an extrinsic terminator of the line. This boundary belongs to the occluder rather than the occluded object. The unoccluded terminator of the gray line is called an intrinsic terminator because it belongs to the line itself. Reprinted with permission from [Grossberg et al. \(2001\)](#).

average. [Rubin and Hochstein \(1993\)](#) created displays in which moving lines could be seen to move in the direction of a vector average, rather than the IOC direction.

A key insight of recent psychophysical research concerns the significance of line endings in motion displays. The trajectories of line ends are unambiguous, but line endings are not all created equal ([Fig. 2](#)). [Nakayama, Shimojo, and Silverman \(1989\)](#) introduced the terms intrinsic and extrinsic terminators to distinguish ends of lines or edges that 'belong' to an object from ends that are adventitiously formed by an occluding object. While motion signals at extrinsic terminators may over time trace out the shape of a stationary occluding object, they tell us little about the true velocity of the occluded object.

The present model accordingly ascribes a primary role to mechanisms that detect and enhance the trajectories of features such as dots, line endings or corners, by creating 'feature tracking signals.' Note that the determination of whether a part of a scene contains such a trackable feature may be better accomplished by mechanisms of the visual system's 'form' pathway, rather than by the early units of the 'motion' pathway, which may process raw spatio-temporal variations of contrast quickly into 'raw' motion estimates. Accordingly, the inputs to the model described in [Fig. 3](#) are here taken to be the outputs of FACADE mechanisms, as described in a model of Form And Color and DEpth processing first described by [Grossberg \(1994\)](#).

A key property of the FACADE model is its ability to process occlusions at T-junctions in a manner that results in a weakened representation for a boundary at an extrinsic line, compared to what is registered at an intrinsic line end ([Kelly & Grossberg, 2000](#)).

2. Neural model

This model employs a cascade of six neural levels, with feedback between the last two. *Level 1*: Input to the model is not a raw spatio-temporal array of changing contrasts, but

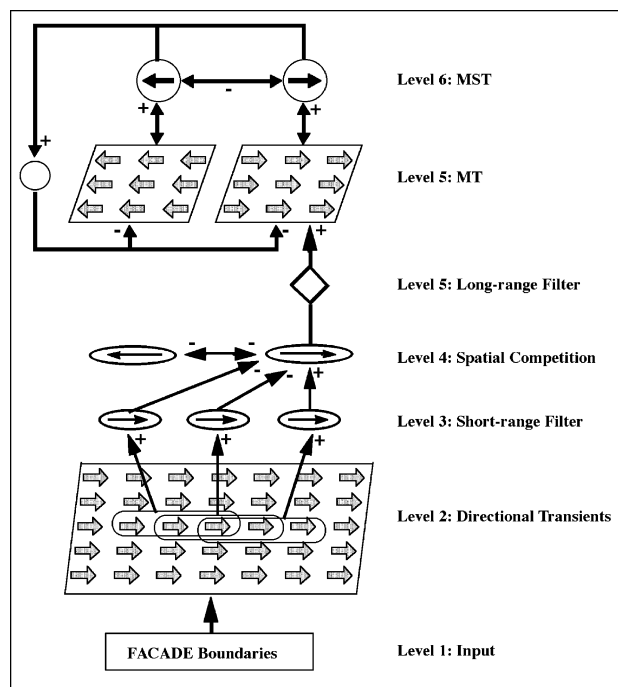


Fig. 3. Model macrocircuit. See text for details. Reprinted with permission from Grossberg et al. (2001).

rather the result of early boundary processing by a network called the FACADE model that is capable of distinguishing extrinsic from intrinsic terminators, based on ‘static’ form cues (Grossberg, 1994; Kelly & Grossberg, 2000). FACADE boundary processing thus models Pathway 1 of Fig. 4, the V1/V2 circuit that is part of the form (or ‘what’ as opposed to ‘where’) pathway. The enhanced response of FACADE boundaries to intrinsic, as compared to extrinsic, terminators yields the seeds of what will become ‘feature tracking signals’ in subsequent levels. *Level 2*: Here directional transient cells begin the process of forming signals sensitive to local motion, via a nulling mechanism such as first suggested by Barlow and Levick (1965) for rabbit retina. *Level 3*: A short-range filter next combines outputs of neighboring directional transient cells that are aligned, thereby gaining selectivity to feature trajectories. *Level 4*: Spatial competition and opponent direction inhibition (Albright, 1984) further overcomes the local aperture problem and enhances directionally unambiguous signals at intrinsic line ends, points, or corners. Signals along line or edge interiors remain relatively weak, because local competition among several competing directions (i.e. the aperture problem) results in normalizing, divisive inhibition. *Levels 5 and 6*: A long-range feedback cooperative/competitive circuit, identified with areas MT and MST, here reorganizes ambiguous directional signals in line or edge interiors based on information from unambiguous trackable features. The nonlinearities of this circuit permit spatially smaller (or, in network terms, numerically fewer) pools of activity that are higher in amplitude to

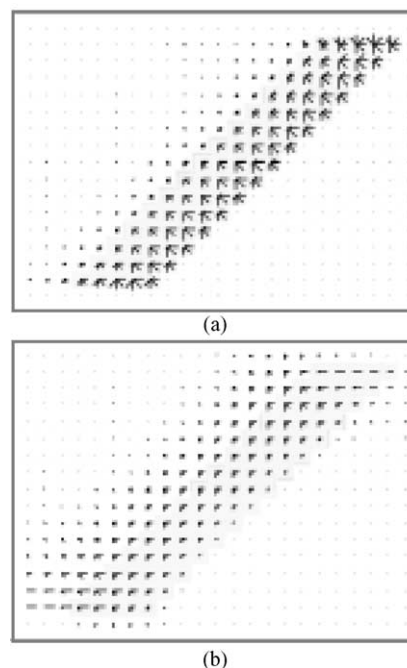


Fig. 4. (a) A diagonal line translating to the right. Output of directional transient cells of Level 2 shows aperture ambiguity at all locations. (b) The Level 4 competition network shows suppression of ambiguous interior signals and directionally unambiguous feature tracking signals at line ends.

overcome spatially larger (and numerically greater) signals along line or edge interiors.

An insight into key model mechanisms is best gotten by considering a simple example of early model stage outputs. Fig. 4a shows that the directional transient cells of Level 2 code an ambiguous array of local motion directions, with the direction normal to the moving line being preferred. The short-range filter and spatial competition stages suppress regions of homogenous, ambiguous direction signals, however, while favoring those signals at the depicted line’s end that code veridically for horizontal rightward motion.

This model has successfully simulated data on the barberpole illusion and its variants involving motion capture, such as spotted barberpole patterns (Shiffrar, Li, & Lorenceau, 1995); line capture (Ramachandran & Inada, 1985); the triple barberpole illusion (Shimojo, Silverman, & Nakayama, 1989); and the chopsticks illusion (Anstis, 1990). Moreover, Pack and Born (2001) have found direct evidence that cells of MT dynamically change their directional preference to the motion of lines moving obliquely to their orientation, as predicted by the terminator-to-interior propagation of signals in the model, see Figs. 4–6.

3. Contrast and speed

The model is sensitive to speed through the use of multi-scale filters at Level 3, followed by intra-scale and

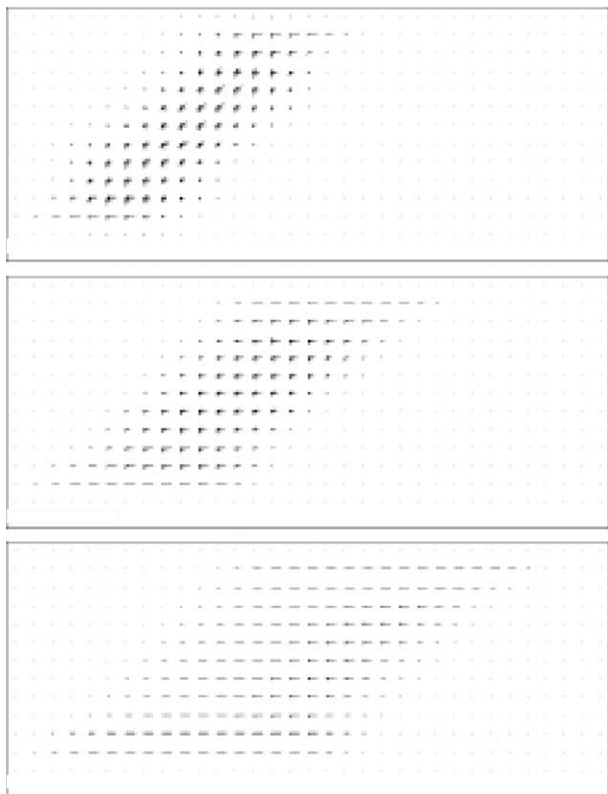


Fig. 5. Activities of long-range filters evolve over time in response to a rightward-moving diagonal line. The early distribution (top) contains ambiguous signals in the interior regions. Successive time slices (middle and bottom) show that the distribution of interior signals normal to the diagonal line is replaced by veridical horizontal signals. Reprinted with permission from Chey, Grossberg, and Mingolla (1997).

inter-scale competition, as described in Figs. 7 and 8. The present description follows that of Chey, Grossberg, and Mingolla (1998).

In the present model speed is represented through the distributed activity of speed-tuned units, or cell populations.

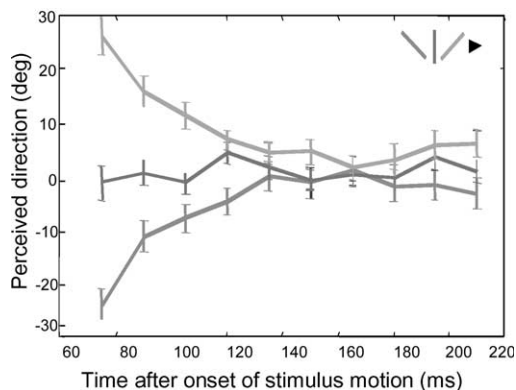


Fig. 6. The preferred directional response of MT cells of macaque changes over time. Initial responses are biased by the orientation of line segments (top right) translating through the receptive field. Over time, cell responses reflect the true horizontal motion of translating stimuli. Reprinted with permission from Pack and Born (2001).

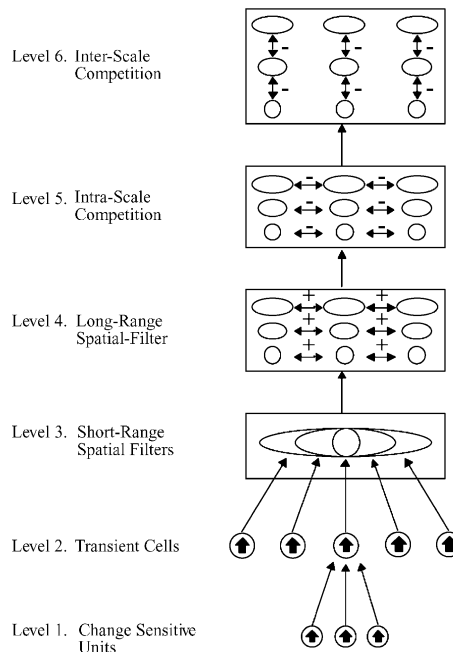


Fig. 7. See text for details. Reprinted with permission from Chey et al. (1998).

We define speed-tuned cells as those that respond preferentially to a limited, continuous range of speeds, as opposed to speed-sensitive cells, but that do not exhibit a preference for a particular speed. The speed tunings of model cells arise primarily from their different spatial scales, which determine the size of their input fields. A key model hypothesis is that larger scales respond preferentially to faster stimuli; we call this covariation the size–speed correlation.

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. How can the units of such a scheme be properly tuned? Each scale is turned on whenever a contrast passes through the region corresponding to the filter’s spatial extent. 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 the same retinal location. 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?

Our work suggests that two simple mechanisms suffice. The first is a scale-proportionate threshold, which requires units of larger scales to have larger absolute activity (or, equivalently, similar proportions of their maximum possible

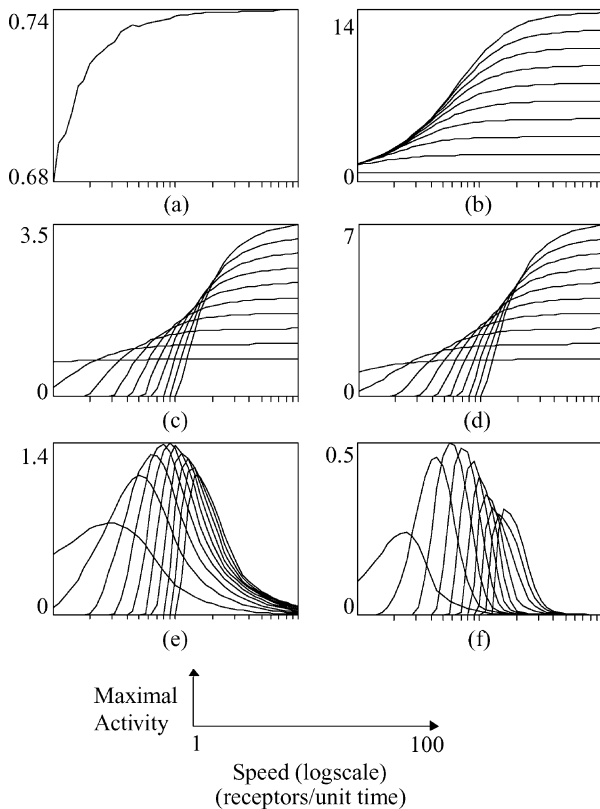


Fig. 8. Maximal responses of cells in the network to a variety of simulated speeds are plotted. From the top left, across then down, plots show (a) transient cells, (b) short-range spatial filters, (c) thresholded short-range filters, (d) long-range filters, (e) intra-scale competition, and (f) inter-scale competition. For levels where there are multiple spatial scales at a single location, 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, as is particularly evident at fast speeds. Vertical axis scales vary and are indicated next to each plot. Reprinted with permission from [Chey et al. \(1998\)](#).

activity) to transmit a signal. The second is competition, both among units of similar scales and across units of differing scales.

[Fig. 7](#) presents a schematic outline of network 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 Level 3 short-range filters are available 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 1 transient cell, scale 2 from 3 and scale 3 from 5. The transient cells that each filter draws from overlap both between scales at a single location and between locations at a single scale. Thus, the largest scale depicted in the outline draws input from a superset of the transient cells that the smaller scales draw from. A long-range spatial filter at Level 4 spatially blurs the thresholded outputs of the short-range

filters. The output of the long-range filters forms input to Level 5 intra-scale competition across space. This competition is enacted through spatial antagonistic center-surround mechanisms. At Level 6, inter-scale competition then takes place between all scales at each spatial location, again enacted through a center-surround mechanism. The results of this final competitive stage form the output of the network and can be interpreted as a pre-speed estimate.

This speed model has been used to simulate a wide variety of data, as described by [Chey et al. \(1998\)](#). To illustrate how model mechanisms interact with key psychophysical findings, one particular simulation study is reviewed here.

4. Contrast effects on speed coding

In addition to affecting discrimination performance, several studies have reported that contrast affects the perceived speed of moving objects. [Thompson \(1982\)](#) reported that low contrast gratings were perceived to move more slowly than high contrast gratings. [Ferrera and Wilson \(1990\)](#) found that contrast influenced the perceived speed of coherent plaid patterns formed from superimposed gratings. [Castet, Lorenceau, Shiffrar and Bonnet \(1993\)](#) found a contrast-induced reduction in perceived speed of translating lines. [Anstis \(2001\)](#) has provided some particularly compelling demonstrations of the importance of contrast in speed perception.

The data simulated here are from a study by [Stone and Thompson \(1992\)](#) 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 percentage of test gratings perceived as moving faster were recorded as a function of the test speed for each contrast level.

To show how the model simulates the change in relative speed judgments due to contrast variations, we used the previously defined speed measure, and calculated speed judgments on the basis of differences between speed measures obtained from two inputs which were simulated separately. We passed the difference between the two speed measures through an ‘error function’ to obtain a simulated probability of an observer judging one speed as faster than the other. The results of this process are shown in [Fig. 9b](#). For inputs of the same contrast, we obtain a standard error function. Changing the simulated contrast or receptor activity amplitude causes a curve shift similar to that observed in the data. In both the data and the simulation, there is little change in the curve shapes under this shift. In addition, both data and simulation show that increasing input magnitude results in a greater shift than decreasing it, at least under this contrast range.

The effects of increasing input magnitude are ultimately bounded by the shunting properties of the transient cells. Thus, beyond a certain contrast level, the model predicts no

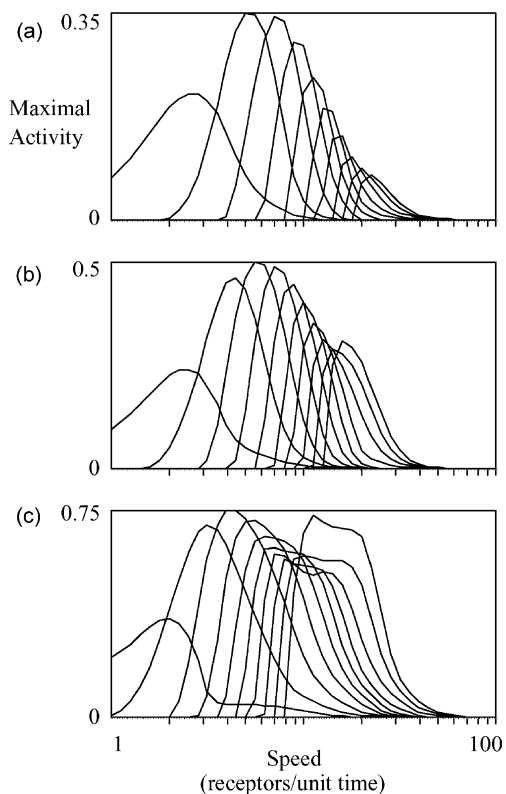


Fig. 9. Maximal network activities over speed using different receptor response amplitudes, hypothesized to correspond to different stimulus contrasts. Plot (a) shows low amplitude receptor activity (0.75), (b) shows intermediate amplitude (1) and (c) shows high amplitude (2). Increasing receptor activity amplitude causes large scales to respond more vigorously and at lower speeds. This biases the network to provide higher speed estimates. Reprinted with permission from [Chey et al. \(1998\)](#).

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 is decreasing in the model due to decreased receptor amplitude, the total activity in the final network level is also decreasing. 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, i.e. 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. Failure to use stimuli in this range may result in failure to observe contrast-induced speed perception.

The model accounts for contrast-induced 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.

Stimulus contrast changes translate into changes in receptor activity amplitude in the model, so that high contrast stimuli generate larger receptor activity amplitudes.

Since the model is based on spatio-temporal summation, one might expect that increasing input amplitude (by raising contrast) might result in a failure of speed estimates, by causing large-scale cells to respond at very slow speeds where they would normally be inactive. Several 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. [Fig. 9](#) shows how contrast affects the model's activity distribution across scales.

5. Previous models

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 modelled by the elaborated Reichardt detector ([van Santen & Sperling, 1985](#)), which differs from the original Reichardt formulation in that preliminary spatial filtering is performed by the receptors before their outputs are multiplied. A similar model (shown to be formally equivalent by van Santen and Sperling) is the motion energy model ([Adelson & Bergen, 1985](#)) in which temporal filtering takes the place of the delay. An alternative model was suggested by [Watson and Ahumada \(1985\)](#). In their model, the temporal response patterns of directionally selective sensors were used to derive speed estimates.

The present model differs from these formulations in that it starts with simple transient cell responses and builds directionality and speed sensitivity from these. The 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. Support for this hypothesis is found in the fact that the emergent tuning varies with input parameters in a manner that matches changes in perceived speed in response to stimulus parameter manipulations in psychophysical experiments.

6. Conclusion

The circuits that underlie motion speed and direction processing are well described by a multi-scale, competitive

network with feedback, designed to emulate the process of the V1-V2-MT-MST circuit of the primate visual system.

Acknowledgements

Supported in part by N00014-01-1-0624 (ONR).

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2(2), 284–299.
- Adelson, E. H., & Movshon, 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.
- Anstis, S. M. (1990). Imperceptible intersections: The chopstick illusion. In A. Blake, & T. Troscianko (Eds.), *AI and the Eye*, Chap. 5 (pp. 105–117). New York: Wiley.
- Anstis, S. M. (2001). Footsteps and inchworms: illusions show that contrast affects apparent speed. *Perception*, 30(7), 785–794.
- Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in the rabbit's retina. *Journal of Physiology*, 178, 477–504.
- Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neurosciences*, 16(7), 263–268.
- Castet, E., Lorenceau, J., Shiffrar, M., & Bonnet, C. (1993). Perceived speed of moving lines depends on orientation, length, speed and luminance. *Vision Research*, 33(14), 1921–1936.
- Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion grouping: from aperture ambiguity to object speed and direction. *Journal of the Optical Society of America A*, 14(10), 2570–2594.
- Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38, 2769–2786.
- Ferrera, V. P., & Wilson, H. R. (1990). Perceived direction of moving two-dimensional patterns. *Vision Research*, 30, 273–287.
- Ferrera, V. P., & Wilson, H. R. (1991). Perceived speed of moving two-dimensional patterns. *Vision Research*, 31, 877–893.
- Grossberg, S. (1994). 3D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, 55, 48–120.
- Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, 41(19), 2521–2553.
- Kelly, F., & Grossberg, S. (2000). Neural dynamics of 3D surface perception: figure-ground separation and lightness perception. *Perception & Psychophysics*, 62(8), 1596–1618.
- Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, 211, 151–180.
- Mingolla, E., Todd, J. T., & Norman, J. F. (1992). The perception of globally coherent motion. *Vision Research*, 32, 1015–1031.
- Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: its relation to image segmentation grouping and the recognition of occluded objects. *Perception*, 18, 55–68.
- 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.
- Ramachandran, V. S., & Inada, V. (1985). Spatial phase and frequency in motion capture of random-dot patterns. *Spatial Vision*, 1(1), 57–67.
- Rubin, N., & Hochstein, S. (1993). Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. *Vision Research*, 33, 1385–1396.
- Shiffrar, M., Li, X., & Lorenceau, J. (1995). Motion integration across differing image features. *Vision Research*, 35(15), 2137–2146.
- Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29(5), 619–626.
- Stone, L. S., & Thompson, P. (1992). Human speed perception is contrast dependent. *Vision Research*, 32(8), 1535–1549.
- Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, 22, 377–380.
- van Santen, J. P. H., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America*, 2(2), 300–321.
- Wallach, H. (1935). On the visually perceived direction of motion. *Psychologische Forschung*, 20, 325–380.
- Watson, B., & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America*, 2(2), 322–342.