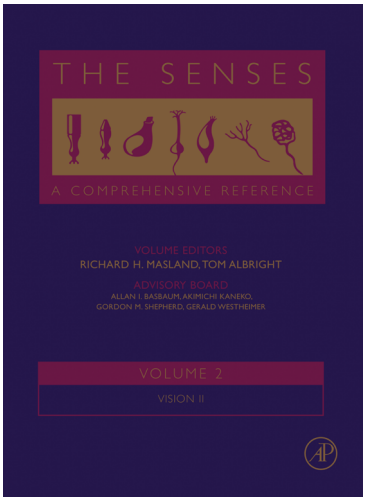

C C Pack and R T Born, Cortical Mechanisms for the Integration of Visual Motion. In: Allan I. Basbaum, Akimichi Kaneko, Gordon M. Shepherd and Gerald Westheimer, editors The Senses: A Comprehensive Reference, Vol 2, Vision II, Thomas D. Albright and Richard Masland. San Diego: Academic Press; 2008. p. 189-218.

2.11 Cortical Mechanisms for the Integration of Visual Motion

C C Pack, McGill University School of Medicine, Montreal, PQ, Canada

R T Born, Harvard Medical School, Boston, MA, USA

Glossary

adaptation Neural process by which the response to a constant stimulus decreases over time.

aperture problem Ambiguity of the true velocity of 1D features (such as lines or edges) sampled through spatially delimited windows. See Section 2.11.4.

barber pole A visual stimulus consisting of a 1D grating moving behind a rectangular aperture. Though the predominant motion energy is in the direction perpendicular to the grating's stripes, the perceived direction of motion is parallel to the long axis of the aperture. This motion illusion was one of the first demonstrations that terminators are powerful motion cues. See Section 2.11.19 and Figure 9(a).

Bayesian model A class of models that use Bayes' theorem for inverting conditional probabilities to incorporate prior knowledge of the world into the intrepretation of sensory data. In a typical Bayesian model of some visual function, one wishes to calculate the probability of occurrence of a particular visual stimulus, *S*, based on some

evidence variable, E, such as a measurement derived from a noisy image or a noisy neural response (written as $p[S/E]$, or the probability of S given E). Typically one has a measure of the inverse probability ($p[E/S]$, the probability of measurement E given stimulus S) and makes use of the prior likelihood of that particular stimulus ($p[S]$) to produce a better estimate, called the posterior probability, of the stimulus that produced the evidence.

bistable Ambiguous visual stimuli that may be seen in either of two mutually exclusive configurations; common non-motion examples are the Necker cube or face-vase illusion.

center-surround opponency: A receptive property in which the neuron's response to a preferred stimulus decreases as the size of the stimulus increases.

coherence/transparency Two mutually exclusive possible perceptual experiences when viewing a moving plaid pattern. Coherence refers to the situation in which the pattern appears to move as a single rigid entity; transparency, to the situation in which the two component gratings appear to slide over one another independently, creating the impression that the nearer surface is partially transparent.

component/pattern prediction Two possible predictions of a neuron's direction tuning curve in response to a 2D plaid stimulus, based on its tuning curve to a 1D grating stimulus; see Sections 2.11.14 and Figure 4.

correspondence problem Problem in determining, for two images that are separated in space and/or time, which image elements belong to the same features. See Section 2.11.2 and Figure 1.

cross-correlation Mathematical technique for determining the degree to which two images are similar; of the two images are snapshots in time, cross-correlation can be used to detect motion; see also correspondence problem; see Section 2.11.6 and Figures 1 and 2.

differentiation Mathematical operation for calculating the rate of change of one variable, x, with respect to another variable, y. The result of the operation is referred to as the **derivative** of x with respect to y. If x represents the position of an object and y is time, differentiation yields the velocity of the object.

end-stopping Receptive property in which responses to extended contours are suppressed by inhibitory zones lying outside the central activating region along the axis of the cell's preferred orientation; see Section 2.11.22 and Figure 10.

extrastriate Regions of visual cortex beyond primary visual cortex (V1).

feature tracking Strategy for solving the aperture problem by locating corresponding 2D features, such as corners of objects or the intersections of plaids, in successive image frames and then calculating their direction of motion; see Section 2.11.19; See also, selectionist model.

frequency domain Shorthand for spatiotemporal frequency domain, a coordinate system for specifying visual motion in terms of two dimensions of spatial frequency (the frequency of a sinusoidal modulation of image intensity of each dimension of 2D space) and one of temporal frequency (the modulation of image intensity over time); the frequency domain representation can be related to the more familiar coordinates of x, y, and t by taking the Fourier transform of the space–time coordinates. Any visual motion sequence (i.e., a movie) can be decomposed into a series of pure sinusoidal gratings (ω_x, ω_y, and ω_t) of differing amplitudes and phases in the same way that complex sounds can be broken down into sums of pure tones (a pure tone being the auditory equivalent of a sinusoidal grating). Because any local sample from a rigidly moving object is constrained to lie on a plane in the spatiotemporal frequency domain, this coordinate system has been used to characterize the receptive fields of neurons that solve the aperture problem. See Figure 6; see also F-plane model.

F-plane model Short for Fourier-plane model. Refers to a class of computational models in which pattern selectivity is created by summing the responses of neurons whose spatiotemporal frequency receptive fields all lie on a single plane in spatiotemporal frequency space; the F-plane model actually represents the power spectrum of the space–time image, as it discards the phase component. See Section 2.11.16 and Figure 6. See also frequency domain.

hypercomplex Term used by Hubel D. H. and Wiesel T. N. (1965) to refer to the receptive fields of neurons that did not respond to extended contours; see also end-stopping and Figure 10.

hypersurface A technical term from differential geometry used to generalize the notion of a curve into multiple dimensions; in the present context it is used to describe the set of points that define the

image intensity as a function of two spatial dimensions (x, y) and one temporal dimension (t).

integrationist model A class of computational model of motion integration that integrates all 1D motion signals to determine the 2D direction of motion; see Section 2.11.15.

intersection of constraints (IOC) Method for computing 2D motion by combining measurements from two or more 1D samples; see Section 2.11.16 and Figure 6.

layers (of cortex) Cut in cross section, the cerebral is a laminar structure, typically consisting of six layers visible with histological stains for cell bodies (such as Nissl stain); neurons whose cell bodies occupy different layers have different patterns of inputs and outputs.

masking Phenomenon in which one visual stimulus impairs the ability to detect another visual stimulus.

motion after-effect (MAE) Visual illusion in which prolonged adaptation to motion in one direction causes a subsequently viewed stationary scene to appear to contain motion in the opposite direction.

motion opponency A receptive field property in which null direction motion causes a reduction in the neuron's response.

MT The middle temporal visual area; also known as V5.

multistable Ambiguous visual stimuli that can be seen in one of several (more than two) mutually exclusive configurations.

nonlinear Any mathematical operation, such as squaring, for which the operation on the sum of two or more inputs is not the same as the sum of the operation's results on the individual inputs; that is, for a linear operation, f, it must be true that $f(a+b) = f(a) + f(b)$. If this relationship does not hold, the operation is nonlinear.

nonlinear systems identification A method for deriving a model of a system based on the relationship between input and output. Often applied to the study of receptive field structure; See Section 2.11.6 and Figure 2.

normalization In the present context it refers to the operation of dividing the output of a given visual filter by the sum of the outputs of all such filters.

plaid Visual motion stimulus created by superimposing two drifting sinusoidal gratings. See Section 2.11.13 and Figure 4.

plaid test Method of classifying visual neurons by comparing responses to both 1D gratings and 2D plaids. See Section 2.11.14 and Figure 4.

preferred/null direction Direction of motion that produces the best/worst response of a given neuron

receptive field The region of visual space in which visual stimuli can alter the firing rate of a neuron; the definition also includes other features of the visual stimulus required, such as the direction of motion, or the temporal pattern of inputs.

retinal disparity Difference in the relative positions produced by a single image on the retinas of the two eyes; such disparity can be used to determine the relative depth (distance from the fixation plane) of the object.

Riemann tensor Mathematical operation used to measure the curvature of a hypersurface.

segmentation The process of determining which image elements belong together.

selectionist model A class of computational model of motion integration that first filters out motion signals emanating from 1D features and thus selectively combines motion signals arising from 2D features; see Section 2.11.25.

sinusoidal grating A visual stimulus in which the intensity of the image is modulated by a sine function along one dimension; if I is a 2D image with the horizontal dimension indexed by x and the vertical dimension by y, the function $I = \sin(\omega x)$ would produce a vertically oriented sinusoidal grating of spatial frequency ω.

T-junction An image feature in the configuration of the letter T generally produced by a near surface occluding a far one.

type I plaid A visual plaid stimulus whose component gratings have directions that lie on opposite sides of the IOC resultant; see Figures 7(a) and 7(b). See also plaid.

type II plaid visual plaid stimuli whose component gratings have directions that lie on the same side of the IOC resultant; See Figure 7(c). See also plaid.

vector sum/average Method for combining two 2D velocity vectors, which have both a direction and a magnitude (speed), to obtain a single velocity vector; geometrically this is done by placing the tail of one vector to the head of the other and drawing a line from the tail of the first to the head of the second, yielding the sum; to obtain the average, the magnitude of the vector sum is scaled by dividing its length by the sum of the lengths of the two

individual vectors. Both methods yield a velocity vector having the same direction. See Figure 7.

V1 Primary visual cortex, the first stage of visual processing in the cerebral cortex and the first stage at which direction selectivity appears; also known as area 17 and striate cortex.

winner-take-all Mathematical operation which, given two or more inputs, produces an output equal to the strongest input, ignoring the weaker ones.

2.11.1 Introduction – Visual Motion

At the level of the retina, visual motion occurs whenever something in the environment moves, or whenever the observer moves or rotates his or her eye. Thus visual motion is a fundamental aspect of our interactions with the world around us. Because retinal image motion has multiple possible causes, it is both computationally challenging and richly informative, serving many functions besides the detection of moving objects. The pattern of image motion created by self-motion, for example, can be used to recover depth and to detect object boundaries as well as to aid in orienting to the environment. It is also useful in guiding eye movements that serve to stabilize images on the retina for high-acuity form vision. However, because movement is a feature that distinguishes many items of great behavioral relevance – potential predators, prey, and mates – the detection of object motion is particularly important.

The latter function will be the focus of this review. In particular, we will address the difficult problem of how information from elementary motion detectors is combined in order to provide accurate representations of the motion of objects. To consider the mechanisms by which brains achieve this integration, we will draw from two sources of data: (1) psychophysical experiments, mainly from humans, and (2) microelectrode recordings from neurons in early cortical motion processing regions of the monkey (and other mammals) – particularly the primary visual cortex (V1) and the middle temporal area (MT). The data will lead us to a discussion of general theoretical approaches as well as specific computational models of motion integration.

2.11.2 The Correspondence Problem

To see why motion integration is necessary, it is first helpful to consider some general problems associated with the basic, low-level detection of motion. Figure 1 shows a pair of successive snapshots of a scene containing the motion of a rigid object. Superimposed on the second panel is the motion that was measured at each point by simply finding pixels in the second image whose luminance corresponded closely to nearby luminances in the first image. There are of course more sophisticated ways to measure motion, and some of them will be discussed in subsequent sections. Although it is apparent that the figure in the pair of images moved from right to left, many of the motion vectors point in other directions. The confusion stems from the fact that many of the pixels in each image have the same luminance, and there is no obvious way to match the individual pixels in the first frame to those in the second frame. This problem is often called the correspondence problem (Ullman, S., 1979), and it renders the computation of velocity difficult for both biological and artificial visual systems.

One might propose that the correspondence problem could be solved both easily and efficiently if one used fewer pixels. That is, if each pixel were the size of the moving figure, then one could imagine a motion algorithm that simply matched the mean luminance of the figure in the first frame to that of the figure in the second frame. The problem with this type of solution is that it requires a visual system that has rather low acuity, which in turn runs the risk of combining motions from separate objects into one measurement. Thus the necessity for high-acuity vision tends to complicate the measurement of quantities that require correspondence between two or more images. Stereo vision and motion are the best-studied examples of such computations. For simplicity we will initially ignore the problem of multiple moving objects and return to this difficulty in later sections.

2.11.3 Motion Noise

Figure 1 also contains a number of motion vectors (purple arrows) that do not correspond to any

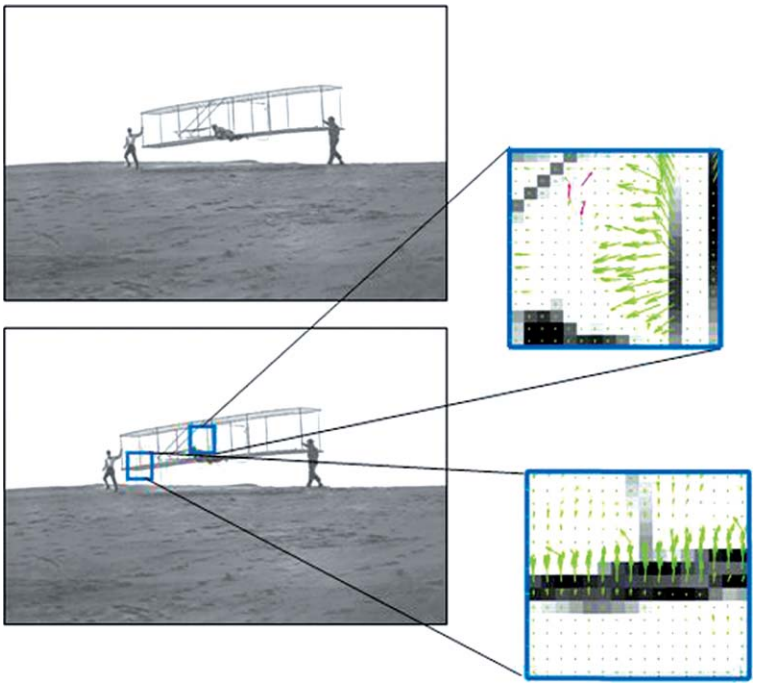

Figure 1 Orville Wright being launched by Dan Tate and his brother Wilbur, in an attempt to fly the Wright Glider in 1902. The green arrows show the average motion vector found between frame 1 (top) and frame 2 (bottom) in a small region of space around each pixel. Purple arrows indicate motion vectors that do not appear to result from the motion of the Glider, which turned out to be unfit for flight. Adapted from a public-domain image on www.first-to-fly.com.

particular object in the scene. These are false matches, in which the luminance of a pixel in the second frame happened to correspond to the luminance of a nearby pixel in the first frame. Such false matches are inevitable when dealing with inputs that are corrupted by noise, jitter, or local image correlations. The visual system possesses at least two mechanisms for reducing such noise in natural images: directional opponency and local pooling. The former has been characterized in early visual areas (Snowden, R. J. *et al.*, 1991; Qian, N. and Andersen, R. A., 1994) and is manifested as a suppression of the response to nearby stimuli moving in opposite directions. Since many sources of motion noise have equal motion energy in all directions, subtracting opposite directions of motion is an efficient way to reduce their contribution. Another way to reduce noise is to pool local measurements over some region of the visual field to produce something like an average of the different local motion vectors (Lisberger, S. G. and Ferrera, V. P., 1997; Recanzone, G. H. *et al.*, 1997; Britten, K. H. and Heuer, H. W., 1999) In fact, direction opponency is a special case of highly localized vector averaging, since the average of two vectors of equal magnitude and opposite direction is zero. It is clear that the visual system possesses mechanisms for pooling across both visual space and velocity space, and these mechanisms pose interesting problems for models of motion integration.

2.11.4 The Aperture Problem

A special case of the correspondence problem occurs when the stimulus contains extended edges or contours, as is almost always the case in natural viewing. As can be seen in Figure 1, when the stimulus consists locally of a single moving edge, the correspondence problem leads to a series of measurements indicating motion vectors perpendicular to the edge's orientation. This is a straightforward consequence of the fact that the parallel component of the actual velocity contains no time-varying information. As a result incorrect measurements will occur whenever an edge that is more than one pixel in length moves in a direction that is not perpendicular to the orientation of the edge. More generally, any motion detector will make similar errors when measuring the motion of a contour that extends beyond its field of view. Consequently, this

problem is often referred to as the aperture problem (Marr, D., 1982), as the limited field of view constitutes a kind of aperture. As it turns out, neurons in the early stages of the primate visual cortex have extremely limited fields of view, so the aperture problem is an important issue in understanding biological vision.

2.11.5 Measurement of Motion in the Primate Brain

The primate visual cortex is an example of a system that has both very high acuity and excellent sensitivity to motion. Both of these properties can be observed at the level of individual neurons. Each neuron in the early stages of the primate visual system responds to stimulation over a very small part of the visual field. In visual neurophysiology, this limited field of view is called the receptive field (RF), and most V1 RFs are smaller than the width of one's thumbnail held at arm's length. It is useful to think of RFs as pinholes (or apertures) through which neurons view the outside world. Their job is to measure certain aspects of the visual world that occur within the pinhole, and they are largely oblivious to events that occur outside this small region. Within any part of the visual cortex, neurons are found whose RFs collectively tile the whole of visual space. In addition to having a spatially delimited RF, each neuron has a limited range of visual stimuli to which it will respond. Some neurons are tuned to the shape of the stimulus, some to the color, and others are tuned to the motion that occurs within their RF. Here we focus on neurons that are selectively responsive to motion, predominantly those in V1 and MT.

2.11.6 Receptive Fields for Measuring Motion

Any system that measures motion must be sensitive to the parts of the visual image that change position over time. That is, any two separate measurements of the position of a moving object will reveal that it has changed position by a certain amount (Δx and Δy) during the interval (Δt) between the two measurements. If the measurements are accurate, these quantities (Δx, Δy, Δt) describe the velocity of the object. Such a measurement is sometimes called a second-order calculation, because it requires a conjunction of two snapshots, separated in space and time.

Second-order measurements are, both intuitively and mathematically (Poggio, T. and Reichardt, W., 1973), the minimal way to compute motion, since a single snapshot cannot meaningfully be said to contain motion. Over the last few decades, researchers have accumulated a great deal of evidence suggesting that this minimal model is actually sufficient to account for the responses of motion-sensitive neurons (Reichardt, W., 1961; Emerson, R. C. *et al.*, 1987; Emerson, R. C. *et al.*, 1992) and certain aspects of human motion perception (Adelson, E. H. and Bergen, J. R., 1985). Such second-order models are equivalent to the better-known motion-energy and Reichardt models (Adelson, E. H. and Bergen, J. R., 1985; Courellis, S. H. and Marmarelis, V. Z., 1992).

A variety of methods have been used to characterize the second-order behavior of V1 neurons, including recently an engineering technique known as nonlinear systems identification (Wiener, N., 1958). The technique is outlined in Figure 2, which shows a sequence of stimuli that were displayed on a computer monitor, which was viewed by an alert macaque monkey. The stimulus was simply a pair of spots, one black and one white, which changed position at random on each refresh of the monitor. The first frame shows the position of the spots at one point in this sequence, and the second frame shows that, a moment later, they changed position, in this case both moving to the right. The dotted squares show the positions occupied by the spots on the previous frame, and the arrows indicate the displacements of the spots from the first to the second frame.

There are four such displacements, corresponding to the four possible ways to match the spots on the first and second frames (white-to-white, white-to-black, black-to-white, black-to-black). Below these two frames is shown the cross-correlation of frame one and frame two. The cross-correlation is a simple statistical way of showing the motion energy that occurs between the frames (van Santen, J. P. and Sperling, G., 1985; Courellis, S. H. and Marmarelis, V. Z., 1992). It shows the amount of motion that occurred at a fixed temporal interval (Δt) in a plane defined by the coordinates (Δx, Δy). In other words, it shows all of the motion vectors that can be found in the two-frame sequence.

The spiking activity of a hypothetical V1 neuron that is sensitive to these motion vectors (e.g., direction-selective) is displayed above the stimulus sequence. To determine which motion vectors the neuron prefers, we average all of the cross-correlations that preceded the spike by a reasonable neuronal latency τ. This procedure allows us to

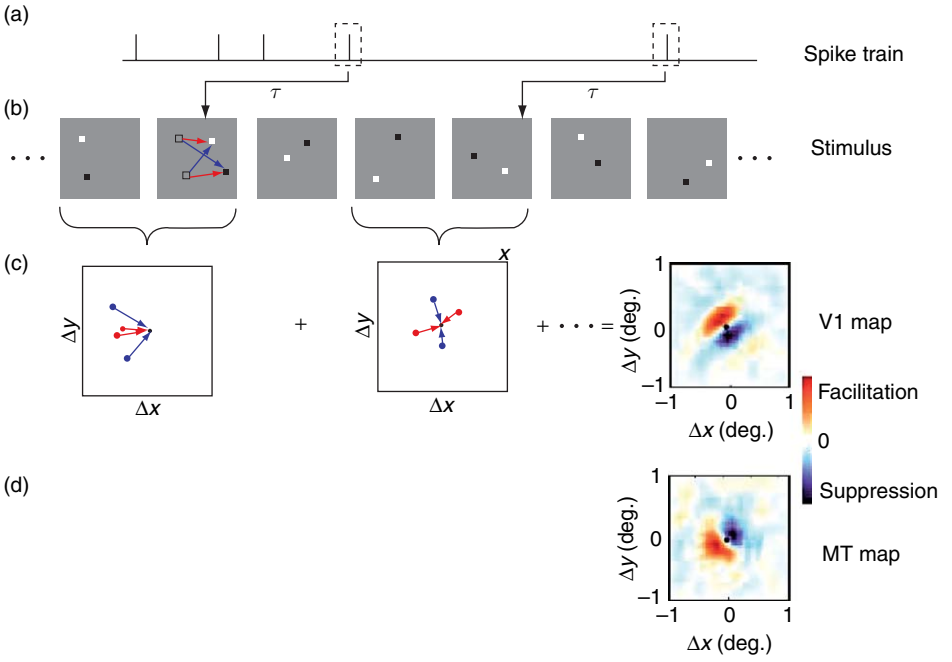

Figure 2 Reverse correlation method. (a) A hypothetical spike train produced by a neuron in response to the mapping stimulus shown in (b). Each time a spike occurs we look back in time by an amount (τ) that corresponds to the latency of the neuron. The stimulus in (b) is a pair of spots that change position randomly on each monitor refresh. Between any two stimulus frames, there are four motion signals, which can be computed by cross-correlating the two images. (c) Displacement maps, corresponding to cross-correlations between pairs of frames, with the origin corresponding to instances in which a spot appeared in the same position twice in a row. The spike-triggered cross-correlations are summed together to produce a map of the average motion vector that led to a spike. A map for an actual V1 cell is shown in the right-most column of (c). (d) A map obtained in the same way for an middle temporal (MT) cell.

characterize the selectivity of the neuron for the velocity parameters Δx, Δy, and Δt.

Such a picture for a V1 neuron is shown on the right side of Figure 2(c). The image shows the response of the neuron as a function of different values of Δx, Δy at a fixed Δt equal to 16 ms, with a neuronal latency of $\tau = 58$ ms. The neuron's response was facilitated for motion sequences down and to the right (bright orange), and suppressed for motion signals up and to the left (dark blue). We will refer such a map as a subunit to indicate that it captures a portion of the neuron's response properties, in this case the second-order selectivity for motion.

2.11.7 A Note on Terminology

The fact that the motion in a stimulus is a second-order property of the input has led to some confusion in the literature. It is common, particularly in theoretical papers, to refer to direction-selective V1 neurons as being linear. Taken literally this would be a contradiction in terms, and it is generally meant as a kind of shorthand to indicate that the neuron's RF acts as a linear spatial filter that works in parallel with a nonlinear mechanism for generating direction selectivity. Some of these models will be discussed below.

Beyond the nonlinear mechanisms that determine direction selectivity, it is generally acknowledged that additional nonlinearities are necessary to account for the behavior of visual neurons. That is, any linear combination of the output of a group of direction-selective neurons will not lead to correct estimates of stimulus velocity. We will discuss the evidence for the possible nonlinearities used by the visual cortex in some detail.

Finally, the use of the term second-order to describe the behavior of direction-selective neurons refers to the statistical order of the computation, and should not be confused with the psychophysical phenomenon called second-order motion, which is a separate concept that will not be discussed in this review.

2.11.8 The Middle Temporal Area of the Visual Cortex

MT has been the subject of intense study since it was first discovered in 1971 (Dubner, R. and Zeki, S. M., 1971). MT neurons receive most of their input from the primary visual cortex (V1), but unlike V1 neurons, they are almost without exception selective for motion direction. Neurons in the MT are also on average more tuned to retinal disparity than V1 neurons, but less tuned to stimulus shape, color, and texture (Born, R. T. and Bradley, D. C., 2005).

Perhaps the most obvious physiological difference between V1 and MT neurons is the RF size. Although the retinotopic arrangement of RFs is similar in the two areas, MT RFs are roughly 10 times the diameter of V1 RFs at any given retinal eccentricity (Albright, T. D. and Desimone, R., 1987). Thus, one might expect that MT neurons could respond to a much larger range of spatial displacements (Δx, Δy) than V1 neurons at a comparable eccentricity. In fact, previous work, using larger stimuli and long display times, has shown that MT neurons do respond better to high speeds than V1 neurons (Mikami, A. *et al.*, 1986; Churchland, M. M. *et al.*, 2005).

Stimuli identical to those used in V1 have been used to compute motion subunits for a large number of MT neurons (Livingstone, M. S. *et al.*, 2001; Pack, C. C. *et al.*, 2003a; Pack, C. C. *et al.*, 2006). An example MT subunit is shown in Figure 2(d). This neuron's response was facilitated for motion sequences up and to the right (bright orange), and suppressed for motion signals down and to the left (blue). However, other than suggesting a different preferred motion direction, the structure of this map is identical to that shown for the V1 neuron in Figure 2(c). That is, the MT neuron responds to roughly the same range of spatial displacements as the V1 neuron.

A systematic study of the subunits in V1 and MT found that the result shown in Figure 2(d) holds for every MT neuron tested: The spatial range over which the MT subunits measure motion is the same as in V1, despite the difference in RF sizes and velocity preferences between the two areas. On average, the optimal spatial displacement in MT is about 0.25°, roughly one-fiftieth the size of the RFs. Nearly identical spatial scales are found in V1 (Pack, C. C. *et al.*, 2003a; Churchland, M. M. *et al.*, 2005; Pack, C. C. *et al.*, 2006), suggesting that the MT subunits represent inputs from V1.

2.11.9 Tiling: The Simplest Model

Given these observations on the MT subunits, a simple model for motion integration by an MT neuron is a tiling model, in which large MT RFs simply sum the outputs of spatially distributed V1 neurons sharing a common preferred direction, with no interactions among the subunits. This model is supported by experiments in which the subunits are mapped at multiple points in an MT RF (Figure 3). The subunits are nearly identical at each point, a finding that held true for all the neurons examined in this way (Pack, C. C. *et al.*, 2006).

Although this is the simplest possible model for how MT RFs may be wired up, the behavior of such a neuron may, in fact, be quite complex. It depends to a large extent on how smart the inputs from V1 are. This will be an important theme that recurs throughout the ensuing discussion: How many of the seemingly more sophisticated motion processing properties of MT neurons are simply inherited from the V1 inputs?

There is a long history in visual electrophysiology of higher-order RF properties first being discovered in an extrastriate visual area, only to be re-discovered in V1 upon closer inspection. There are at least two good reasons why this occurs. First, as we have already seen for MT, extrastriate RFs tend to be much larger than those in V1, and this makes tests with more complex stimuli practically easier to conduct. For example, the miniature eye movements present during fixation are of a similar size to V1 RFs, and this makes many measurements, such as the nonlinear systems identification techniques described above, much more

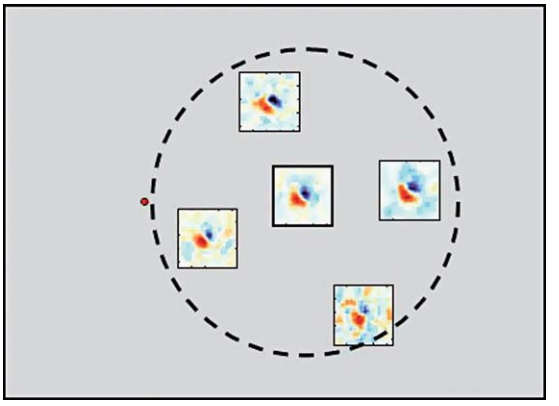

Figure 3 The result of performing the mapping procedure shown in Figure 2 at multiple points within an MT receptive field. The dot shows the fixation point, and the dashed circle shows the estimated extent of the receptive field.

technically challenging. Second, V1 is an incredibly heterogeneous visual area, containing, in essence, all of the basic visual information that it subsequently distributes to a multitude of more specialized extrastriate areas. This means that the neurons that form the principal driving input to a given extrastriate area might constitute only a tiny, and highly homogeneous, fraction of the neurons in V1, thus making comparisons of average properties between V1 and any extrastriate area difficult at best. This is clearly the case for MT, whose V1 inputs are very different from V1 as a whole: they originate mainly from a single sublayer (4B, with a small minority also coming from the so-called solitary cells of Meynert found at the border between layers 5 and 6; Maunsell, J. H. and van Essen, D. C., 1983; Shipp, S. and Zeki, S., 1989), are highly direction selective, respond over a wide range of temporal and spatial frequencies, and are very sensitive to stimulus contrast (Movshon, J. A. and Newsome, W. T., 1996). However, because they constitute such a small fraction of the total V1 population, it has been difficult to characterize very many of them in precise detail and thus to know the true nature of MT's inputs.

Given the above considerations, one way to proceed would be to start with the conceptually simple tiling model and ask what it can explain, and, perhaps more interestingly, what it might explain given certain properties of V1 neurons that have been widely documented and which are quite likely to be properties of the V1 neurons projecting to MT.

2.11.10 Tiling and Motion Noise

As noted above, one benefit of summation across many subunits within a larger RF (i.e., spatial pooling) is an improved signal-to-noise ratio. MT neurons are indeed remarkably good at detecting weak motion signals embedded in noise. In some cases the sensitivity of single MT neurons exceeds the perceptual sensitivity of the animal measured simultaneously (Newsome, W. T. *et al.*, 1989; Britten, K. H. *et al.*, 1992). In order to account for ability of MT neurons to pool many V1 inputs fully, one probably would have to add some kind of divisive normalization (Heuer, H. W. and Britten, K. H., 2002), which appears common to most cortical circuitry (Carandini, M. *et al.*, 1997), in order to compensate for the limited dynamic range of neural spiking. This can also account for the observation that when two small targets move in different directions within a single MT RF (in the absence of attentional influences) the neural response represents the average of the two motion vectors (Lisberger, S. G. and Ferrera, V. P., 1997; Recanzone, G. H. *et al.*, 1997). More on this idea below.

With respect to noise reduction, there appears to be a subtractive interaction between subunits tuned to opposite directions of motion; however, it is also possible that opponent processing is inherited from V1. Andersen and colleagues examined directional interactions in both V1 and MT using two superimposed fields of random dots moving in opposite directions (Snowden, R. J. *et al.*, 1991; Qian, N. and Andersen, R. A., 1994) . In one important study, they compared the amount of suppression produced by the null-direction dot field under conditions in which the two dot fields contained paired dots (i.e., for every dot moving in the preferred direction, there was an immediately adjacent dot moving in the null direction) versus two random dot fields that were spatially unpaired. They found, on average, more suppression to paired dot fields in MT than in V1 and concluded that it was likely that MT performed additional subunit subtraction for purposes of noise reduction (Qian, N. and Andersen, R. A., 1994). However, although the population averages were different, some V1 neurons were as suppressed by paired opponent motion as those in MT (compare figures 4 and 14 of Qian, N. and Andersen, R. A., 1994). Furthermore, the V1 cells most strongly suppressed by paired null motion were also the most strongly direction-selective (see figure 15 of Qian, N. and Andersen, R. A., 1994), a hallmark of the V1 neurons known to project to MT (Movshon, J. A. and Newsome, W. T., 1996). It is thus quite possible that the directional opponency seen in MT neurons is passively inherited from the highly specialized subset of V1 neurons that serve as its inputs. Indeed, it would seem to make sense to perform this very local comparison at a stage where RFs are small. Insofar as this is the case, the tiling model may be sufficient to account for most of the exceptional noise immunity of MT neurons. A certain degree of motion opponency is implicit in the subunits, as seen in the negative response to null-direction motion (Figures 2(c) and 2(d)). However, many models posit an additional opponency stage that subtracts the outputs of individual subunits preferring opposite motion directions (Courellis, S. H. and Marmarelis, V. Z., 1992).

2.11.11 A Problem for the Tiling Model

The tiling model simply sums the outputs of V1 neurons that share a common subunit structure, which in turn predicts the motion vectors to which the neurons will be responsive, and, as we have seen, provides improved motion integration under noisy conditions. The aperture problem, however, guarantees that many of these motion vectors will not just be randomly incorrect, but rather systematically biased (Figure 1(b)). In this case, simply adding them up will not yield the correct stimulus velocity. Rather, the visual system needs to process the motion signals in such a way as to overcome the limitations imposed by the errors that are implicit in the input. This issue lies at the heart of our remaining discussion of motion integration.

2.11.12 Conceptual Approaches to Solving the Aperture Problem

In its most basic form, the problem consists of recovering a two-dimensional (2D) velocity based on measurements that have been rendered one-dimensional (1D) by the existence of edges in the visual world and the small RFs of neurons in the brain. This necessitates some comparison of two or more motion measurements made at edges having different orientations. General approaches to the solution can be divided into two categories that differ primarily according to the stage at which the comparison occurs. In the first category of models, which includes what has become the standard model (Heeger, D. J. *et al*., 1996; Simoncelli, E. P. and Heeger, D. J., 1998), the first stage, assigned to V1 neurons, is relatively unintelligent: it nonselectively extracts only 1D motion signals which are then combined nonlinearly, at the second stage, assigned to MT, that recovers 2D velocity. Because these models combine all local 1D signals to compute the final velocity, we refer them as integrationist models.

In contrast, a second category of models places the comparison between orientations in V1, by way of known mechanisms such as end-stopping that suppress 1D motion signals in favor of motion signals emanating from 2D features. These features are just regions of the image where multiple orientations can be found locally. The second stage then simply averages the stage-one outputs to compute the final 2D velocity, an operation that can be accounted for by the tiling concept. Because the essential strategy of these models is to first select regions of the image where 2D motion measurements are most reliable and then to combine only these in the final computation, we refer to them as selectionist models.

The two categories of models are not mutually exclusive, and in many cases they make similar predictions. In the following paragraphs, we will review the psychophysical and physiological evidence supporting these different approaches, and then consider in more detail some of the computational models that have been put forth to embody them.

2.11.13 Plaids

The initial evidence for the two-stage, integrationist models came from psychophysical and physiological experiments using visual plaids (Adelson, E. H. and Movshon, J. A., 1982; Movshon, J. A. *et al*., 1985). Visual plaids are usually constructed by superimposing two, circularly windowed, sinusoidal (1D) gratings that are rotated with respect to each other (Figure 4). When the gratings are of similar spatial frequency and contrast, the resulting percept is generally of motion in a single direction, a condition referred to as coherence, with the resulting direction being the pattern (2D) direction, as distinguished

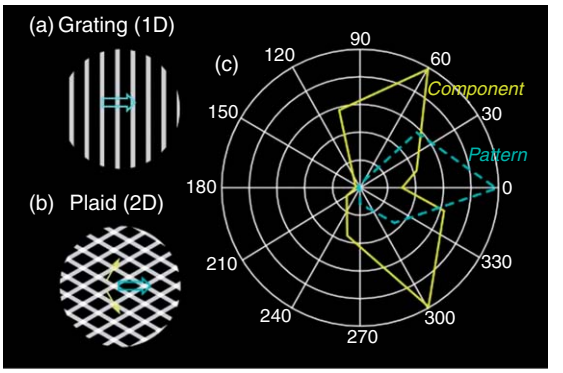

Figure 4 Plaid test for single neurons. The neuron's direction tuning curve to a one-dimensional (1D) grating stimulus (a) is used to make two predictions for the tuning curve to a two-dimensional (2D) plaid (b) in which the two components are rotated ±60° with respect to the original grating. Two predictions for an MT neuron are shown in (c) and consist of the pattern prediction (dashed curve), which is identical to that obtained with the 1D grating, and the component prediction (solid curve), which is obtained by summing two copies of the grating tuning curve that have been rotated ±60°.

from the different component (1D) directions of the two gratings comprising the plaid.[1]

In a series of psychophysical experiments, Movshon and colleagues tested the effects of masking (Adelson, E. H. and Movshon, J. A., 1982) and adaptation (Movshon, J. A. *et al.*, 1985) on human observers' tendencies to perceive coherent motion in plaids. Given the evidence that neurons in striate cortex responded optimally to contours of a particular orientation (Hubel, D. H. and Wiesel, T. N., 1962), they set out to find perceptual evidence for an orientation-selective 1D stage prior to the stage at which coherence was computed. In one experiment, they used 1D dynamic noise (oriented, flickering bars of various widths) to mask the plaids. They reasoned that, if the first-stage filtering was in fact orientation-selective, then the mask should interfere with the perception of coherence when it was parallel to one of the component gratings. In contrast, if the first stage was not orientation-selective, but rather a kind of blob-tracking mechanism that signaled the direction of the plaid's intersections, the effect of the mask would either not depend on its orientation or perhaps would be greatest when it was perpendicular to the direction of pattern motion. They found that the 1D masks were only effective when oriented within 20° of one of the component gratings, in agreement with an orientation-selective first stage.

In a second experiment, they examined the effects of adaptation on the subjects' abilities to both detect the presence of a moving stimulus and to perceive coherence of the plaid pattern. The premise of this experiment was that detection thresholds are a signature of early processing, such as that found in V1 and characterized psychophysically by orientation- and direction-selective adaptation (Sekuler, R. W. and Ganz, L., 1963), whereas the coherence thresholds reflected later stages of motion processing. Any differences in the adaptability of these two measures would be evidence for separate processing stages. Further, by testing for cross-adaptation, that is, effects of adapting with a 1D grating on the perception of a 2D plaid (and vice versa), they might uncover further evidence for an orientation-selective early stage. To see the logic of this, imagine a horizontal grating moving upwards as the adapting stimulus and an upward-moving coherent plaid (whose components move obliquely: up-left and up-right), as the test stimulus. If the earliest stage codes direction of motion, irrespective of orientation, then we might expect to find considerable cross-adaptation for these stimuli, because they both appear to be moving upwards. If, however, the early stage is both orientation- and direction-selective, then we would expect little cross-adaptation since such a stage would see only the components of the plaid, neither of which is horizontal or moving directly upwards. The results revealed little evidence for cross-adaptation of detection thresholds (see figure 7 of Movshon, J. A. *et al.*, 1985), but considerable cross-adaptation for coherence thresholds (see figure 8 of Movshon, J. A. *et al.*, 1985), both consistent with a two-stage model in which the early (detection) stage consists of filters that are tuned to both orientation and direction of motion.

Subsequent psychophysical experiments further supported the two-stage model of Adelson E. H. and Movshon J. A. (1982). In one experiment, Welch L. (1989) measured speed discrimination (Weber fractions) for gratings and plaids. Because the pattern and the components of a plaid move at different speeds (pattern speed is always faster than component speed, according to the cosine rule given below) and because discrimination performance varies with the baseline speed, Welch was able to ask which of the two speeds better predicted discrimination performance for plaids. The answer was clearly that the component speed, not the pattern speed, predicted discrimination performance (Welch, L., 1989). In another experiment, Derrington A. and Suero M. (1991) used the motion after-effect (MAE) to reduce the perceived speed of one of the components of a plaid stimulus. When they did this, they found that the perceived direction of the plaid, after adaptation, deviated in the direction of the nonadapted component. The deviation could be nulled by reducing the speed of the nonadapted component to match the perceived speed of the adapted component (Derrington, A. and Suero, M., 1991). Both results suggested that the visual system first estimated the motion of the plaid's components before combining them to generate the percept of pattern motion.

2.11.14 Plaid Physiology

Based on the psychophysical results with plaids, Movshon and colleagues used these same visual stimuli to characterize direction-selective neurons of the visual cortex. They recorded from both V1 and MT in anesthetized monkeys and from the homologous areas (area 17 and lateral suprasylvian cortex, respectively) in anesthetized cats. This allowed the authors to assess directly whether neurons in a given area tended to respond to the direction of the 1D

components or the 2D patterns, since, for any given stimulus these moved in different directions.

The physiological version of the plaid test was performed as follows (Figure 4). The direction-tuning curve of the neuron was first determined using a 1D sinusoidal grating, and this tuning curve was then used to make two predictions about the shape of the same cell's tuning curve to a plaid stimulus. Insofar as the cell sees only the 1D components of the plaid, its direction tuning curve should be bi-lobed (Figure 4, solid yellow line), with the peak of each lobe corresponding to the direction of plaid motion that places one of the component's direction of motion in the cell's preferred direction. For example, suppose a neuron prefers rightward motion to the grating stimulus. When tested with the plaid stimulus, the neuron's response is plotted as a function of the direction of the pattern motion. For a 120° plaid, when the pattern is moving to the right, neither of the grating components is moving in the preferred direction. However, for plaid pattern directions of either +60° or −60°, one or the other component grating will be moving in the cell's preferred direction. In contrast, if the cell sees pattern motion, the plaid-derived direction tuning curve is predicted to be identical to that obtained with the grating.[2] By correlating the actual tuning curve obtained to a plaid with each of the two predictions, the authors were able to quantify the extent to which a given neuron was responding to the direction of motion of the 1D components or the 2D pattern.

The results, like the plaid psychophysics, supported their two-stage model, and did so more directly. The investigators found that the majority of the neurons in cat area 17 and monkey V1 were of the component type (59 of 69 cells, 85%) with the remaining 10 neurons falling in the unclassed zone, corresponding to neurons for which neither prediction was significantly better than the other. Importantly, none of the neurons were classified as pattern-type. The results from macaque MT neurons were somewhat different: while a majority of the neurons were still classified as either component (40%) or unclassed (35%), a significant minority were characterized as pattern (25%). Thus V1 neurons appeared to see only the 1D motion (as predicted by the aperture problem) while at least some of those in MT were able to combine the 1D measurements into a representation of 2D velocity. This result clearly supported models of the integrationist type.

2.11.15 Integrationist Models

Accepting for the moment the notion that V1 neurons respond only to the 1D motion components, the challenge raised by the experiments of Movshon and colleagues is to explain how an MT neuron can have inputs that are component-selective and an output that is pattern-selective. An extremely simple explanation is depicted in Figure 5. Here three model MT

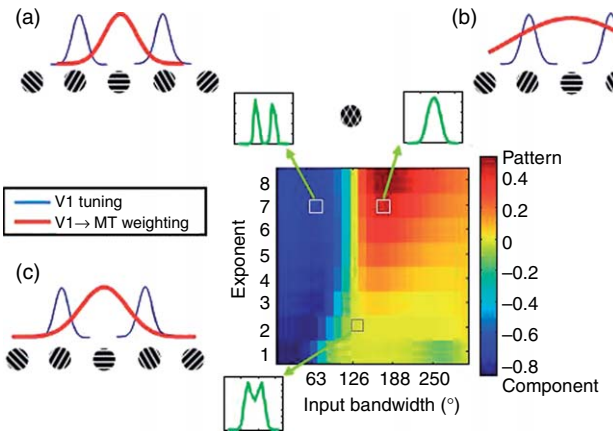

Figure 5 Three examples of hypothetical primary visual cortex (V1) tuning curves and middle temporal (MT) weighting functions. The blue tuning curves show the responses of V1 neurons to the component gratings in a plaid moving upward, which is the preferred direction of the simulated MT neuron. The red curves show the function that weights the output of the V1 neurons in the projection to MT. The central panel shows how tuning for pattern motion is affected by different bandwidths of the weighting function and different output nonlinearities. The green tuning curves show specific examples of simulated neuronal tuning to plaid patterns. The simulated plaid contained components moving in direction separated by 120°. Simulated V1 tuning bandwidth was 27° (Albright, T. D. 1984). The range of exponents was taken from DeAngelis, G. C. *et al.* (1993).

neurons receive inputs from a group of component-selective V1 neurons. For each neuron each V1 input is weighted by a Gaussian function (red curve) centered on the MT neuron's preferred direction, which in each case is upward. In Figure 5(a) the Gaussian function is sufficiently narrow that the V1 responses to the component gratings (blue curves) receive little weight when the plaid moves in the MT neuron's preferred direction. The result is a bi-lobed tuning curve (green) typical of a neuron that is classified as a component-selective. In contrast the tuning width of the MT neuron in Figure 5(b) is sufficiently large that the neuron responds to both of the plaid gratings simultaneously when the pattern direction is the same as the preferred direction of the MT neuron. The summed response to the two components is then greater than when either of the component gratings is moving in the preferred direction. In essence the MT neuron is blurring or averaging the motion inputs, so that the component directions no longer appear to be visible to it. As mentioned above, this kind of blurring is thought to have a role in eliminating motion noise, which is likely to be randomly distributed over space (Qian, N. and Andersen, R. A. 1994).

Note that the hypothetical neuron shown in Figure 5(b) would not yet meet the definition of pattern selectivity given by Movshon J. A. *et al.* (1985). The reason is that such neurons are defined as exhibiting responses that cannot be predicted from a linear combination of the responses to the components. For the neuron shown in Figure 5(b), the response to the plaid is still a linear sum of the response to the component gratings, and thus would not be considered a true pattern-selective neuron.

Nevertheless, the existence of a response peak in the pattern direction suggests that the tuning-width explanation may be a reasonable place to start. To determine if this type of explanation can be brought into line with the results of the plaid experiments, we can equip the hypothetical MT neuron with some standard nonlinearities that are known to influence the responses of all neurons. For instance, neurons have a threshold for firing spikes, and so input causing depolarization of the postsynaptic membrane that does not reach this threshold cannot be observed in the spiking activity. Furthermore, the spiking response of a neuron is generally not a linear function of its input, but rather an expansive nonlinear function. These are two uncontroversial nonlinearities that have been observed many times (Carandini, M. and Ferster, D., 2000). We will therefore consider how they might influence the responses of MT neurons to plaid stimuli.

We simulated a population of MT neurons with different bandwidths, thresholds, and expansive nonlinearities.[3] Each neuron was then classified according to a pattern index (Stoner, G. R. and Albright, T. D., 1992; Pack, C. C. *et al.*, 2001) that captured how well it conformed to the pattern prediction, taking into account the overlap between the component and pattern predictions. The central panel in Figure 5 shows the results, with red pixels corresponding to neurons that would statistically be considered pattern-selective, and blue corresponding to component-selective. As expected, increasing the bandwidth was necessary but not sufficient for generating pattern selectivity (Figure 5, bottom rows). However, when a large bandwidth was combined with either a high firing threshold (not shown) or a sharply accelerating nonlinearity, pattern selectivity emerged. Indeed using a range of parameters similar to that observed in V1 studies (e.g., DeAngelis, G. C. *et al.*, 1993) produces a distribution of component and pattern selectivities that is not unlike the distribution seen in anesthetized macaque MT. In this model, pattern and component neurons are identical, except that the former have broader tuning bandwidths than the latter (Figure 5, top rows). This can be seen clearly by comparing Figures 5(a) and 5(b), which have exactly the same nonlinearities, but differ radically in their component/pattern classification.

The nonlinearities used in the simple model described above are routinely observed in real neurons, and hence have become standard features of neural models. For example the model of Simoncelli E. P. and Heeger D. J. (1998), which will be described in more detail below, uses similar nonlinearities in its implementation of an intersection of constraints (IOC) rule to compute pattern motion. However, as we have shown here the IOC part of the model is not strictly necessary to produce pattern selectivity, although it presumably plays an important role in modeling other types of data. The strong prediction of the hypothesis shown in Figure 5 is that the pattern index would correlate on a cell-by-cell basis with each neuron's tuning bandwidth. This appears to be the case, since pattern neurons were observed to have substantially broader tuning than component neurons (Albright, T. D., 1984).[4]

A recent model has built upon a similar approach to model MT neurons, and reached a similar conclusion. Rust N. C. *et al.* (2006) developed a model in which component and pattern cells differed in tuning

bandwidth, but had similar nonlinearities. However, to model pattern cells realistically, it was necessary to add two additional features. The first was an inhibitory influence of V1 neurons on MT neurons with different preferred directions. This allowed the model to use lower (and probably more realistic) exponents than those used in the model shown in Figure 5. The second addition was strong surround suppression at the level of V1. This feature of the model is similar to the central mechanism of many selectionist models, which will be described in a subsequent section of this chapter.

2.11.16 The Intersection of Constraints, or Fourier-Plane, Model

As for the tuning-width model above, in the model proposed by Simoncelli E. P. and Heeger D. J. (1998) each V1 subunit sees only the small slice of space-time orthogonal to its preferred orientation (Figure 6), that is, it is tuned to a relatively narrow range of spatial and temporal offsets for contours of a given orientation, and it cannot discern exactly how a particular $\Delta x/\Delta t$ combination arose. Thus the model's V1 outputs suffer from the aperture problem.

The aperture problem is solved at the second stage, where MT cells collect inputs from many V1 component cells each of whose orthogonal, 1D speed measurement is consistent with a given 2D velocity. Conceptually, one can think of each 1D measurement as being consistent with a number of possible 2D velocities, all of which must fall on a line in velocity space (Figure 7, dashed lines). The intersection of any two of these constraint lines (provided the 1D measurements were made at edges having different orientations and belonging to the same object) provides the solution. As a result, this type of model is often referred to as an intersection of constraints, or IOC, model.

To see better how the IOC calculation might be carried out by a neuron, we consider an example MT cell that is to be tuned for true 2D velocity upwards at $10^\circ\ s^{-1}$ (Figure 6(a)). Such a cell would receive excitatory inputs from a family of horizontally oriented V1 cells with upwards direction preferences but with different combinations of preferred spatial and temporal offsets ranging from spatially fine subunits, say an optimal Δx of 0.1° and a Δt of 10 ms, to spatially coarser subunits, such as one with optimal values of a Δx of 1° and a Δt of 100 ms – all consistent with a speed of $10^\circ\ s^{-1}$ perpendicular to their horizontal orientation. In addition, this same MT cell would receive excitatory drive from additional $\Delta x/\Delta t$ families of V1 cells tuned to different orientations/directions but corresponding to slower preferred orthogonal speeds, according to a cosine relationship: $S_{V1} = S_{MT} * \cos(\theta_{V1} - \theta_{MT})$, where S_{MT} and θ_{MT} are the desired preferred speed and direction of the MT neuron, and S_{V1} and θ_{V1} are the corresponding directional preferences of the oriented V1 inputs. Thus, the entire range of possible orientation–velocity relationships is described by a circle in velocity space (Figure 6(b)). For our example, MT neuron, the optimal Δx's and Δt's of the family of inputs preferring right oblique orientations and upright motion would all have preferred orthogonal speeds of $7^\circ\ s^{-1}$, and, similarly, the preferred orthogonal speed of the

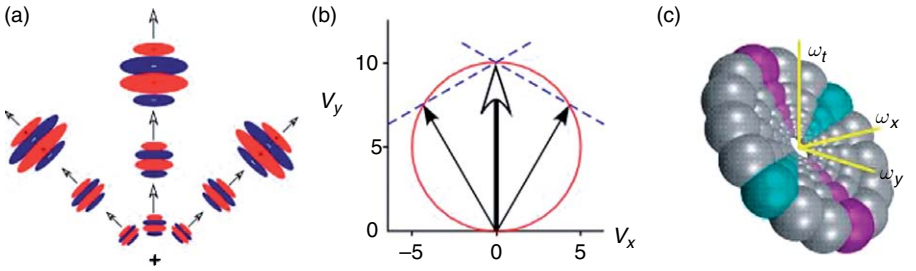

Figure 6 Integrationist model of pattern direction selectivity in middle temporal (MT) proposed by Simoncelli E. P. and Heeger D. J. (1998). (a) Partial collection of primary visual cortex (V1) inputs to a pattern cell tuned to a velocity of $10^\circ\ s^{-1}$ upwards. Each subunit corresponds to the receptive field of a V1 complex cell. (b) Velocity–space representation of the intersection of constraints (IOC) calculation. Any one-dimensional (1D) velocity measurement (solid arrows) is consistent with a range of two-dimensional (2D) velocities falling along a constraint line (dashed lines) perpendicular to its motion vector. For two such measurements made at different orientations, the 2D velocity is given by the point of intersection of the two constraint lines. Conversely, all 1D velocities consistent with a given 2D velocity (hollow arrow) fall on a circle in velocity space. (c) The frequency–space representation of the model depicted in (a). See text for additional details. (a,c) Adapted from figures 2D and 3B of Simoncelli, E. P. and Heeger, D. J. 1998. A model of neuronal responses in visual area MT. Vision Res. 38, 743–761.

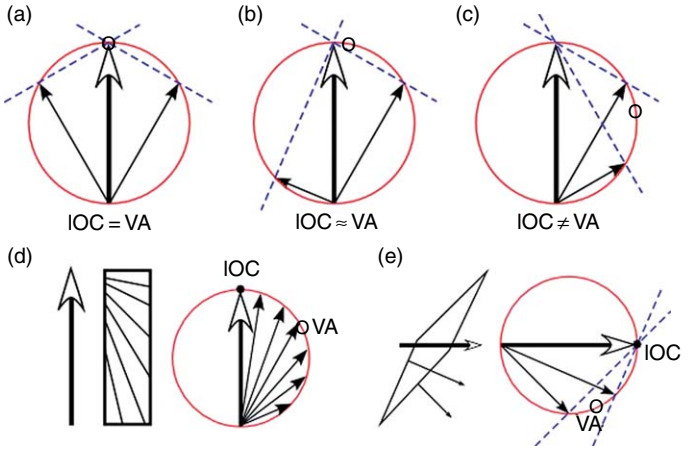

Figure 7 Velocity–space representations of different, multicomponent motion stimuli. (a) Symmetric plaids of the kind used by Adelson E. H. and Movshon J. A. (1982) in which the intersection of constraints (IOC) and vector average (VA) produce identical directions. (b) Asymmetric type I stimuli in which the two components move at different speeds. For most such stimuli, the IOC and VA yield approximately the same direction of motion, though the VA is not always perfectly accurate. (c) Type II stimuli in which the one-dimensional (1D) velocities of the two components both lie on the same side of the resultant. For these stimuli, the IOC and VA directions are quite different. (d) Type II barber poles used by Rubin N. and Hochstein S. (1993) along with the velocity–space representation of the component motion vectors. (e) Thin rhombus from Weiss, Y. *et al*. (2002), which, at low contrasts, appears to move in the VA direction.

vertically oriented V1 inputs to this cell would be $0° \text{ s}^{-1}$. The range of Δx's and Δt's within each family of V1 inputs at a given orientation confers upon the model MT cell true speed tuning (i.e., independent of the spatial composition of the stimulus) and the different families at different orientations/directions allow MT cells to solve the aperture problem and respond to the pattern direction of a plaid.

While the velocity–space construction used above is conceptually useful, in practice the model is implemented in the spatiotemporal frequency domain (Figure 6(c)). In the first stage, the Fourier-transformed visual stimulus is multiplied by the frequency-space representations of oriented Gabor filters whose output is half-squared (i.e., the negative values are clipped off and the result is squared) and normalized to produce a measure of motion energy (Adelson, E. H. and Bergen, J. R., 1985) orthogonal to the orientation of the Gabor. When plotted in the three-dimensional space comprised of two dimensions of spatial frequency (ωx and ωy) and one of temporal frequency (ωt), the selectivity of a given model V1 neuron appears as a pair of localized blobs positioned symmetrically about the origin, and the different spatial scales of the filters (corresponding to the $\Delta x/\Delta t$ families described above) fall along a line passing through the origin. In this frequency space the locus of points corresponding to a unique 2D velocity describes a plane, so a given MT cell just sums up all of the spatiotemporal blobs within the plane consistent with its particular preferred direction and speed. The plane of blobs in Figure 6(c) thus comprises the responses of V1 neurons representing eight different orientations/directions at five different spatial scales. Because the model MT neurons in stage two can be thought of as planar templates in frequency space, the IOC model has also been referred to as the Fourier-plane, or F-plane, model (Born, R. T. and Bradley, D. C., 2005).

This model has been highly successful in accounting for, not only the original plaid data that motivated it, but a number of other known MT properties, such as responses to random dots embedded in noise (Newsome, W. T. *et al*., 1989; Britten, K. H. *et al*., 1992). Subsequently, further support for this type of model came from Okamoto H. *et al*. (1999), who showed that component MT neurons have bimodal direction tuning for dots moving at high speeds, while pattern cells have bimodal tuning for bars moving at slow speeds. Both of these results are direct predictions of the F-plane model (Simoncelli, E. P. *et al*., 1996), although they are also consistent with other types of integrationist models (Kawakami, S. and Okamoto, H., 1996; Albright, T. D., 1984). The F-plane model also predicts bimodal responses in MT pattern cells for stimuli composed of overlapping dot fields, a prediction that was recently confirmed (Bradley, D. C. *et al*., 2005). Finally a direct test of the F-plane model in MT

found a population of cells that demonstrated the spatiotemporal selectivity predicted by the model (Perrone, S. A. and Thiele, A., 2001). However, these cells turned out not to be pattern cells when tested with plaids (Priebe, N. J. *et al.* 2003), as described below (see Section 2.11.18 Challenges to Integrationist Models).

2.11.17 Other Integrationist Models

There are a number of other models whose basic structure is very similar to that of the IOC model described above in that they all first indiscriminately calculate local, 1D measures of motion and then combine them nonlinearly to yield the IOC solution at a subsequent stage of the model. One source of relatively minor differences is the nature of the elementary motion detectors used at the first stage. Perhaps the most popular method is the motion energy model of Adelson E. H. and Bergen J. R. (1985), which was used by Heeger and colleagues in the model detailed above (Heeger, D. J., 1987) as well as by Grzywacz, N. M. and Yuille, A. L. (1990). Other methods include the gradient constraint method of Limb J. O. and Murphy J. A. (1975), used in the model of Fennema C. and Thompson W. (1979), and those based on scalar motion sensors described by Watson, A. B. and Ahumada, A. J. (1985) and used in the motion integration model of Ogata M. and Sato T. (1991). Similarly, the IOC calculation at the second stage can be realized in a variety of ways. We have already seen two of them in the form of the velocity–space construction, also used by Sereno M. E (1993) and Albright T. D. (1984) and the F-plane template implementation by Heeger D. J. (1987). Another way of calculating the intersection of constraint lines is to use the inverse Hough transform, and this has been featured in several other integrationist models (Fennema, C. and Thompson, W., 1979; Ogata, M. and Sato, T., 1991; Kawakami, S. and Okamoto, H., 1996). The details of these models are important – they affect the models' performance in comparison to human observers and how well it describes the response properties of neurons in V1 and MT, and they also are critical for evaluating the model's biological plausibility in terms of computations that can be carried out by neural circuits – but they are beyond the scope of this review. The issue of greater concern is to see to what extent they are consistent with the existing psychophysical and physiological data, and to compare them with models of a fundamentally different kind.

2.11.18 Challenges to Integrationist Models

As mentioned above, indirect evidence for the F-plane model comes from experiments that have confirmed the model's predictions on bimodal direction tuning. However, these same findings are predicted by other models, so they do not constitute a direct test of any particular hypothesis. A more decisive test of the F-plane model would involve stimulating MT neurons with gratings of different spatial and temporal frequencies, and measuring the extent to which neuronal responses display velocity invariance. That is, according to the F-plane model pattern-selective neurons should be selective for velocity in a manner that is largely independent of the spatiotemporal composition of the stimulus. This experiment has been done by Priebe N. J. *et al.* (2003), who found that true velocity tuning is rare in MT and, more importantly, it does not correlate in any obvious way with the pattern-component categorization (Priebe, N. J. *et al.*, 2003). In fact, velocity selectivity as hypothesized by the F-plane model is no more common in MT than in V1 (Priebe, N. J. *et al.*, 2006), suggesting that it is not related to the computation of pattern motion. Similar findings on velocity tuning were found using bars (Mikami, A. *et al.*, 1986) and small spots (Pack, C. C. *et al.*, 2006). The F-plane model thus does not appear to be consistent with the finding that the vast majority of MT neurons are capable of accurately encoding motion direction despite large changes in the spatiotemporal components that comprise the stimulus (Pack, C. C. and Born, R. T., 2001).

To the extent that integrationist models do not account for the MT data, it is useful to examine some of the models' underlying assumptions. Perhaps the strongest of these hypotheses is the notion of a purely linear first stage that measures only 1D spatial frequency components. This concept has often proven useful in vision modeling, but it is important to keep in mind its status as an approximation to the real behavior of V1 neurons. Real neurons at all stages of visual processing have a variety of nonlinear responses, some of which may be important to the integration of motion signals. A second category of models, which we call selectionist models, examines the extent to which these early nonlinearities might contribute to motion integration. Before considering

the models themselves, we will review some of the issues that motivated their conception.

2.11.19 Intersection of Constraints, Vector Average, or Feature Tracking?

The early experiments on the perception of plaid motion used component gratings whose orthogonal 1D velocities were symmetrically placed on either side of the resultant 2D pattern velocity, producing so-called symmetric type I plaids (Figure 7(a)). For such symmetrical stimuli, a simple average (or sum) of the two 1D vectors will produce a resultant vector with the same direction (though a different speed) as that produced by the IOC computation. For simplicity, we will refer to this type of computation as the vector average (VA). The experiments of Adelson E. H. and Movshon J. A. (1982) could not distinguish between these two computations. Furthermore, their stimuli contained potentially trackable 2D features, in the form of the bright blobs formed at the intersections of the two gratings. Such features could be detected by a simple operation that detects luminance maxima (Bowns, L., 1996). And since these 2D features always move in the pattern direction, this strategy cannot be easily distinguished from an IOC. While the adaptation and masking experiments (Adelson, E. H. and Movshon, J. A., 1982) argued against a feature-tracking strategy for plaids composed of sinusoidal gratings, we will see below that other, more salient, features can dramatically affect perception.

While symmetric plaids cannot distinguish between an IOC (or feature-based) computation and a VA, other stimuli can be constructed in which the component velocities either straddle the resultant asymmetrically (asymmetric type I; Figure 7(b)) or both lie to the same side of the resultant (type II; Figure 7(c)). In terms of probing the nature of the second-stage computation, type II stimuli are particularly interesting, because the perceived direction predicted by an IOC can be very different from that predicted by a VA (Figure 7(c)). In this case, the IOC computation produces the veridical direction of pattern (or object) motion, whereas the VA is inaccurate. As such it would seem to make sense for the visual system to use the IOC. Surprisingly, however, human observers misperceive the direction of motion of such stimuli under certain conditions, and they do so in the direction predicted by the VA. This was found, for example, for type II plaids of low contrast or those viewed for brief durations (Yo, C. and Wilson, H. R., 1992), for modified, type II, barber pole stimuli (Rubin, N. and Hochstein, S., 1993; Figure 7(d)), for multiple line segments each presented within a separate aperture (Mingolla, E. *et al.*, 1992), and for a thin, rhombus at low contrast (Weiss, Y. *et al.*, 2002; Figure 7(e)). For all of the above stimuli, there was a consistent direction of rigid translational motion that an IOC computation recovers. The fact that this direction was not perceived argues strongly that the visual system does not always use an IOC calculation.

Further evidence against a strict IOC computation came from experiments in which type I and type II plaids were compared with respect to their ability to interfere with the perception of a test pattern moving in the IOC resultant direction (Ferrera, V. P. and Wilson, H. R., 1987) and with respect to direction discrimination thresholds (Ferrera, V. P. and Wilson, H. R., 1990). In both sets of experiments, type II plaids behaved very differently from their type I counterparts. In particular, the type II plaids were much less effective at masking, being no more effective than their 1D component that was nearest to the resultant direction (Ferrera, V. P. and Wilson, H. R., 1987), and they yielded much higher direction discrimination thresholds that were also biased towards the direction of motion of their components (Ferrera, V. P. and Wilson, H. R., 1990). To explain their psychophysical results, Wilson and colleagues (1992) suggested a type of integrationist model, in which a VA of Fourier and nonFourier motion components is computed.

Some of these same experiments also revealed that 2D features, such as dots or line endings, had a powerful effect on the perceived motion of visual patterns. In the experiments of Rubin N. and Hochstein S. (1993), using dashed, instead of solid, lines to construct the modified barber pole stimulus dramatically changed the perceived direction of motion from the VA to veridical. In another series of experiments, they added variable numbers of randomly placed dots to the regions between the solid lines and found that even a single dot was sufficient for observers to report the true direction of pattern motion. Mingolla E. *et al.* (1992) found a similar transition from a VA of the isolated 1D elements to the true direction of pattern motion when they added 2D features such as small rectangles defining the line segment's endpoints. Finally, the classic barber pole illusion, originally described by Wallach (Wallach, H., 1935; Wuerger, S. *et al.*, 1996), is a powerful demonstration of the ability of the motion of 2D features to influence perception. In this case, the

features are the angular endings of the windowed grating, often referred to as terminators. Despite the fact that the 1D component contributes much more motion energy, the prevailing direction of 2D terminator motion along the long axis of the aperture dominates the percept.

The barber pole illusion might seem to indicate that the 1D motion signals are completely ignored by the visual system, but this is clearly not the case. The influence of 1D motion can be seen in the phenomenon of capture (Castet, E. *et al.*, 1999), which refers to the greater probability of perceiving motion in the terminator direction as the direction of the 1D signals approaches it more nearly. Other evidence for an active competition between 1D and 2D motion signals is the multistability of barber pole stimuli, which refers to the possibility of seeing one of three modal directions (long axis, short axis, or perpendicular to contour) during single brief observations (Castet, E. *et al.*, 1999) and to the fact that the perceived direction alternates randomly during prolong viewing (Wallach, H., 1935). The multistability is particularly apparent when the 2D signals are weakened by rendering them extrinsic[5] (Castet, E. *et al.*, 1999), suggesting mechanisms for computing the salience of 2D features (Shimojo, S. *et al.*, 1989).

2.11.20 Dynamics of 1D and 2D Computations

In the context of the competition between 1D contours and 2D features, perhaps the simplest way to probe motion integration psychophysically is to construct a stimulus that consists entirely of line segments moving in a direction that is not perpendicular to their orientation. As for barber poles, the contour motion is in a different direction than the motion of the line-segments' ends. This kind of stimulus was first used in psychophysical experiments by Lorençeau J. *et al.* (1993), who demonstrated that humans perceive the motion of such stimuli inaccurately when viewed for brief durations or when the lines were of low contrast. Importantly, these inaccurate percepts were not randomly wrong, but rather were systematically biased in the direction perpendicular to the edges of the line segments, as predicted by the aperture problem. This result was thus similar to those obtained with type II plaids (Yo, C. and Wilson, H. R., 1992), and strongly suggested an interesting temporal dynamic from an early VA of 1D signals to a more veridical percept at later times, produced by either an IOC or feature-based computation.

2.11.21 Bar-Field Physiology

The stimuli used by Lorençeau J. *et al.* (1993) can be readily adapted to neurophysiological experiments, and they can test many of the same hypotheses that have been addressed with plaid stimuli. Indeed from the point of view of models with a purely linear first stage, the tilted bar stimulus is simply a plaid with multiple component gratings. A crucial difference is that the tilted bar stimulus contains spatial frequency components that move in the direction of the pattern as a whole. These components always have substantially lower amplitude than the components that move perpendicular to the orientation, so under this hypothesis any neuron that accurately measures their motion must have nonlinear response properties.[6]

The response of MT neurons to tilted bar stimuli has been measured using an experimental design that dissociated stimulus orientation from direction of motion. The bars moved in one of eight directions, and on different trials they were tilted at angles of 45°, 90°, or 135° with respect to the motion direction. The 90° tilt condition served as a measure of the baseline direction tuning of each neuron, since in this case the local measurements perpendicular to the orientation were by definition correct. For the 45° and 135° conditions, local measurements would be expected to be inaccurate, and so these stimuli provided a different way to probe the ability of MT neurons to overcome the aperture problem.

Figure 8 shows the experimental results. The early responses in MT showed substantial biases for motion perpendicular to edge orientation, as would be predicted from the simple tiling model described previously. In contrast, the later responses were almost completely independent of stimulus orientation for nearly every neuron in the population.

This result differs from what one would expect based on the plaid experiments (Movshon, J. A. *et al.*, 1985). In particular, the existence of component neurons has been interpreted to mean that a substantial fraction of MT neurons have responses that can be modeled based solely on linear spatial frequency channels. This is clearly not the case for the vast majority of neurons in the Pack C. C. and Born R. T. (2001) study[7] and indeed other studies have found that all MT neurons exhibit nonlinear behavior in response to even modest stimulus manipulations (Stoner, G. R.

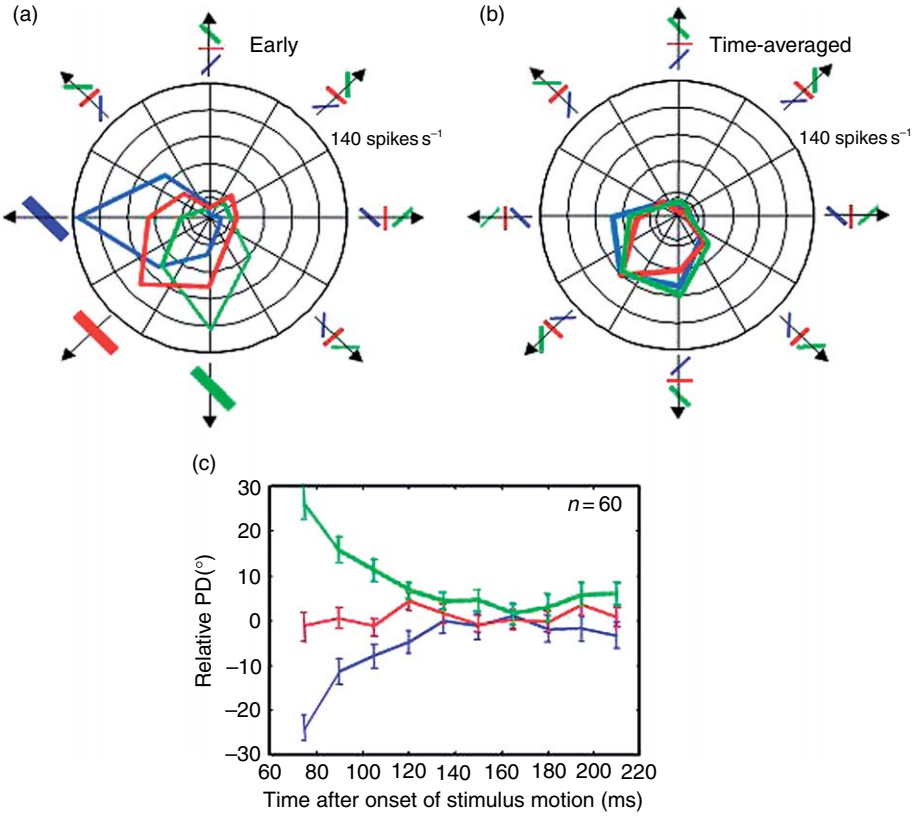

Figure 8 Response of middle temporal (MT) neurons to the bar field stimulus. (a) For a single MT neuron, the early part of the response depends on both the orientation and direction of the bar field. This neuron responds best whenever the bar has a left-oblique orientation and a leftward or downward motion component, indicating that it sees only the component of motion perpendicular to the bars. (b) The later part of the response depends only on the motion direction. (c) The transition from orientation-dependent responses to purely motion-dependent responses is evident in the population of 60 MT neurons. PD, preferred direction.

and Albright, T. D., 1992; Pack, C. C. *et al.*, 2004; Krekelberg, B. and Albright, T. D., 2005). This is not to say that linear models are incorrect – the hypothesis of linearity is after all just a way of representing the stimulus, and the responses are clearly related to the stimulus. Rather the results indicate that linear models are not sufficient to account for the behavior of MT cells, even if one grants that they may provide an accurate description of V1 outputs.

The results with bar fields did not distinguish between integrationist and selectionist models, since both would be expected to measure motion direction accurately. However, they did provide an impetus for looking for such differences in V1. This proves critical for distinguishing between the two types of models, because selectionist models would predict that selectivity for the motion of 2D features should be found in V1, while integrationist models predict that only 1D motion is represented at the first stage. Further motivation to look for a representation of 2D feature motion in V1 came from psychophysical experiments, which strongly suggested that the motion of terminators is calculated at a very fine spatial scale (Power, R. P. and Moulden, B., 1992; Kooi, F. L., 1993). A particularly dramatic illustration of this is that the barber pole illusion is completely abolished by cutting small notches in the aperture so that the direction of local terminator motion becomes the same as that of the 1D contours (Figure 9(a)).

A similar effect was seen for MT neurons in conscious monkeys (Pack, C. C. *et al.*, 2004). These investigators demonstrated that MT neurons reveal an effect similar to that of the barber pole illusion: their directional responses were dominated by the motion of the 2D terminators and not by that of the 1D contours. In fact, by testing neurons with barber poles of different aspect ratios, they were able to show that, as a population, MT neurons compute the VA of

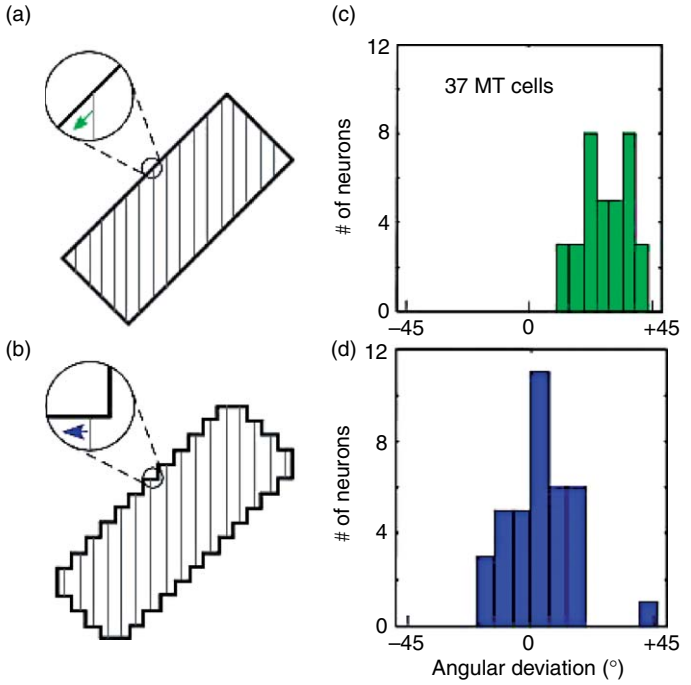

Figure 9 Effect of aperture shape on the responses of middle temporal (MT) cells. In the standard barber pole illusion (a), the dominant perceived direction of motion is parallel to the long edge of the aperture – in this case, down and to the left. Cutting notches in the aperture (b) abolishes the illusion, such that the perceived direction is now to the left, in the same direction as the one-dimensional grating. The notches have the same effect on MT cells (c) and (d). For normal barber poles (c), the direction vectors of the MT population are deviated from the grating direction toward the direction of motion of the terminators on the long axis of the aperture. (d) Cutting notches in the aperture eliminates this deviation, and the MT population now shifts back to the direction of grating motion. Adapted from figure 8 of Pack, C.C., Gartland, A.J., and Born, R.T. 2004. Integration of contour and terminator signals in visual area MT of alert macaque. J. Neurosci. 24, 3268–3280.

the two directions of terminators – those along the short and long edges of the aperture – weighted according to their relative frequencies, with little influence of the 1D signals.[8] This was measured as a deviation in the neuron's preferred direction away from the 1D and towards the 2D direction of motion (Figure 9(a)).When notches were cut in the barber poles (emulating the psychophysical experiment of Kooi F. L. (1993)) the MT population response collapsed back to the 1D direction, just as for the percept (Figure 9(b)). The most striking aspect of this result was the relative dimensions of the features: the barber poles were scaled to nearly fill the center of the MT RFs, making them, on average, 6–10° along the long axis of the aperture, while the indentations that abolished the barber pole effect were only 0.4° in length, less than one-tenth the size of the MT RF's linear dimensions. In fact, at the eccentricities tested, the indentation's dimensions were much more closely matched to RF sizes in V1 (Van Essen, D. C. *et al.*, 1984).

Similar results were also obtained with plaid stimuli. In this experiment, Majaj N. J. *et al.* (2007) first characterized MT cells using the plaid test, but with versions of plaids that were a fraction of the size of the MT RF. They next tested the same neurons with pseudoplaids that had been, in effect, pulled apart so that the two component gratings were now side-by-side instead of overlapping, yet both still well within the RF center. The effect of this manipulation was always to make the cell's direction tuning curve less patternlike, suggesting again that the 2D computation is performed locally at a spatial scale smaller than that of MT RFs.

2.11.22 Physiological Evidence for Early 2D Motion Signals

The first physiological evidence for 2D motion signals early in the cortical motion pathways had already been provided by Hubel D. H. and Wiesel T. N. (1965) who described neurons that were both direction-selective and end-stopped (or hypercomplex in their original nomenclature; Figure 10). The

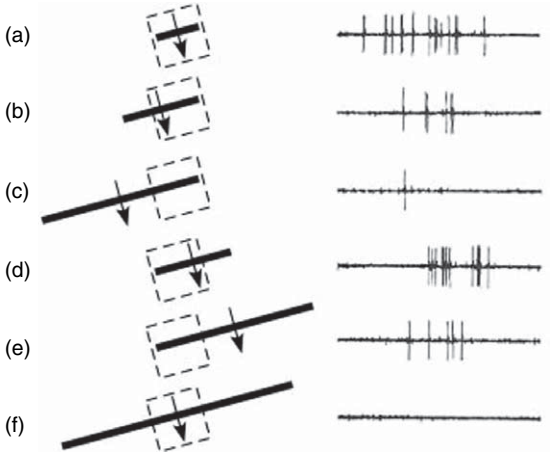

Figure 10 An end-stopped, direction-selective cell recorded by Hubel D. H. and Wiesel T. N. (1965). The neuron responds to a short bar, just covering its activating region (a), and to longer bars whose endpoints are centered over the activating region (b)–(e), but gives no response to a long bar centered over its activating region (f). Adapted from figure 17 of Hubel, D. H. and Wiezel, T. N. 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the Cat. J. Neurophysiol. 28, 229–289.

cell responds preferentially to motion downwards (and slightly to the right), giving no response to the opposite direction of movement. Moreover, the cell is completely silent to any bar longer than the cell's tiny activating region. Such a neuron would appear to be a perfect candidate for providing 2D motion signals.[9] However, as Hubel D. H. and Wiesel T. N. only tested motion orthogonal to the orientation of the (very short) bar, they were not able to rigorously test the neuron's relative immunity to the aperture problem, nor did they demonstrate true 2D direction selectivity for line endings.

Both of these properties were subsequently demonstrated for neurons in striate cortex of alert monkeys by Pack C. C *et al.* (2003b). Figure 11a shows the RF of an end-stopped, direction-selective V1 neuron, as mapped with small spots identical to those shown in Figure 2. In this case the analysis reveals the parts of visual space to which the neuron responds (i.e., the RF). Figure 11(d) shows the same neuron's second-order subunit.[10] calculated using exactly the same method as that shown in Figure 2. Neither map shows any characteristic that would distinguish the neuron from the standard notion of a direction-selective complex cell.

Based on these maps, one can generate predictions of the same neuron's response to a second stimulus, in which the small spots were replaced with two long bars, one white and one black. In this experiment, the bars matched the neuron's preferred orientation, and were substantially longer than the RF shown in

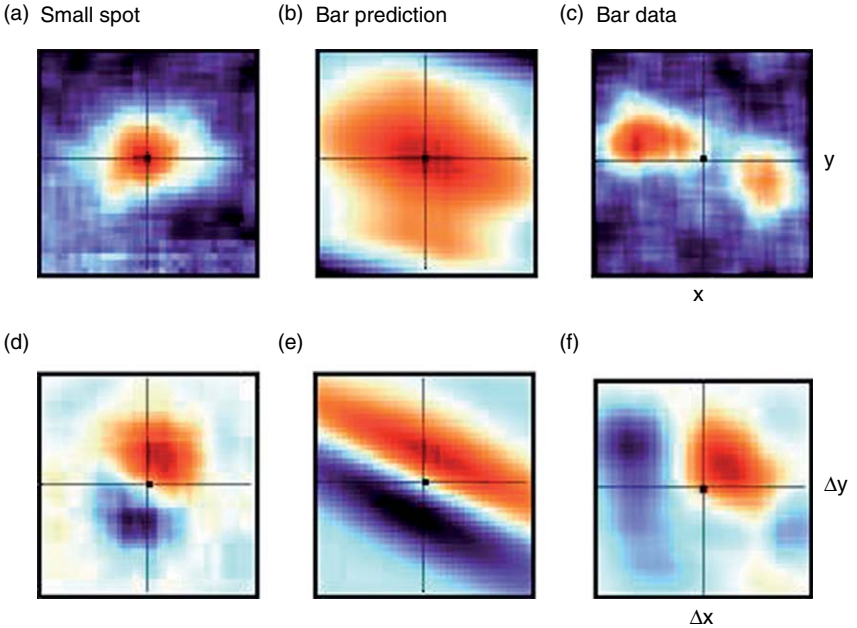

Figure 11 Subunit structure for a direction-selective, end-stopped primary visual cortex (V1) neuron. (a) The map of the receptive field obtained with small spots. (b) The predicted response to long bars flashed at different positions near the receptive field. (c) The actual map obtained with long bars. (d)–(f) As in (a)–(c), but the maps and predictions are for two-frame displacements of the long bars.

Figure 11(a). Otherwise the experiment was identical to the first one, with the bars changing position at random on each frame. The predictions, computed using convolution integrals, are shown in Figures 11(b) and 11(e). Not surprisingly, the long bar has the effect of stretching both maps along the bar's axis of orientation. This is what would be expected if the neuron viewed the long bar as simply a collection of spots like those used to generate the original maps.

The actual data, obtained using a variant of the analysis shown in Figure 2, is shown in Figures 11(c) and 11(f). The map of responses to individual bar positions, shown in Figure 11(c), reveals a characteristic signature of end-stopping: two hot spots with an intervening cold zone, which gives the maps the appearance of a dumb-bell (Figure 11(c)). One can view this map as depicting the parts of the bar to which the neuron responds best, and the two hot spots as indicating that the neuron only responses to the endpoints of the bar. The response to the center of the bar is suppressed by the mechanism that generates end-stopping, with the result that the neuron is essentially a detector of 2D features. This confirms that end-stopped neurons respond to the endpoints of long bars as was previously shown by Hubel D. H. and Wiesel T. N. (1965). In the study of Pack C. C. *et al.* (2003), the presence of end-stopping was also verified by the standard method of sweeping bars of various lengths across the RF and measuring the response of the neuron.

The data for the second-order subunit also showed a striking departure from the linear prediction shown in Figure 11(e). Rather than being smeared along the axis of bar orientation, the map shows a discrete region of activation which suggests a preference for roughly the same range of velocities as when the cell was stimulated with small spots. The cell was likely to fire an action potential in response to any two-bar sequence proceeding from up-right to down-left, corresponding to a preferred direction of 225°. This closely matched the preferred direction in the map obtained with spots, as well as the preferred direction obtained by a conventional direction tuning curve using swept bars or drifting random dots. The similarity of all three measures presumably reflects the neuron's selectivity to the bar's endpoints, because the motion of these 2D features are not influenced by the aperture problem. That this representation was truly independent of the 1D motion signals generated by the bar was shown by generating additional second-order maps using long bars that were rotated ±45° from the cell's preferred orientation. In all cases, the preferred direction indicated by the second-order map changed little, if at all.

It has not been shown that the type of end-stopped, direction-selective V1 neuron described above actually projects to MT, so it is possible that MT neurons receive only 1D motion signals from nonend-stopped cells and must use an IOC or VA to recover the 2D motion *de novo*. This seems unlikely, however, based on the known properties of neurons in 4B, which, in terms of numbers of neurons, provide over 90% of the V1 input to MT (Maunsell, J. H. and van Essen, D. C., 1983; Shipp, S. and Zeki, S., 1989). It is well established that layer 4B has a great proportion of highly direction-selective neurons (Dow, B. M., 1974; Blasdel, G. G. and Fitzpatrick, D., 1984; Livingstone, M. S. and Hubel, D. H., 1984; Hawken, M. J. *et al.*, 1988), and that most of these neurons also exhibit strong suppressive surrounds, including end-stopping (Sceniak, M. P. *et al.*, 2001). While these previous studies did not identify any of the recorded neurons as projecting to MT, it highly likely that many of them did, as the MT-projecting neurons are the largest neurons in layer 4B (Sincich, L. C. and Horton, J. C., 2003) and thus more likely to be sampled by a microelectrode than their smaller neighbors (Towe, A. L. and Harding, G. W., 1970; Humphrey, D. R. and Corrie, W. S., 1978; Lemon, R., 1984). In sum, it is highly likely that the bulk of the V1 inputs to MT originating from layer 4B are both direction-selective and strongly surround suppressed.

The finding of 2D motion selectivity in end-stopped cells suggests a reasonable explanation for the responses to bar fields in MT, but it remains unclear to what extent end-stopping can account for the observed responses to plaids. Theoretically the possibility makes sense, since most macaque V1 neurons are strongly end-stopped, and such neurons respond poorly to stimuli that contain only one orientation (Jones, H. E. *et al.*, 2001; Sceniak, M. P. *et al.*, 2001). For these neurons the component gratings in a plaid stimulus would not elicit strong responses, but the points near the intersections of the two gratings contain multiple orientations and thus might elicit stronger responses. These points move in the direction of the plaid pattern, so a simple explanation for the responses of pattern-selective MT neurons is that they receive input from V1 neurons that are themselves biased toward pattern selectivity.

Unfortunately, the end-stopping hypothesis is somewhat difficult to test with plaids, because pattern selectivity is typically defined with respect to predictions based on the responses to individual gratings.

Because a grating is defined as having only one orientation, a neuron that (by whatever mechanism) responded only to multiple orientations would be untestable by this method. In other words, a neuron that was pattern selective to the exclusion of component response would be unclassifiable. Consequently there is a bias inherent in all plaid studies towards overestimating the percentage of neurons that respond only to motion components.

Evidence against the idea of a pattern-selective projection from V1 to MT comes from Movshon J. A. and Newsome W. T. (1996), who found that V1 neurons that were identified as projecting to MT were all classified as component-type. However, the data consisted of only 12 neurons, half of which were from layer 6 where neurons appear specialized to provide 1D motion signals (Gilbert, C. D., 1977; Sceniak, M. P. *et al.*, 1999) but which provides a tiny fraction of MT's V1 input (<10%), and the degree of surround suppression present in these neurons was not reported.

The situation is further complicated by recent studies that have reported the existence of pattern-selective neurons in V1 (Tinsley, C. J. *et al.*, 2003; Guo, K. *et al.*, 2004). One group has found that, in macaque monkeys, general anesthesia appears to abolish these pattern responses (Guo, K. *et al.*, 2004). This is interesting because other groups have found that nonlinearities responsible for contextual surround effects are reduced by anesthesia (Lamme, V. A. *et al.*, 1998) and for stimuli of low contrast (Levitt, J. B. and Lund, J. S., 1997; Polat, U. *et al.*, 1998; Kapadia, M. K. *et al.*, 1999; Sceniak, M. P. *et al.*, 1999; Anderson, J. S. *et al.*, 2001; Cavanaugh, J. R. *et al.*, 2002), both of which characterized the physiological recordings of Movshon J. A. *et al.* (2003). This finding is also consistent with reports that, in MT, general anesthesia has the effect of reducing the sensitivity to pattern motion for both bar fields and certain types of plaids (Pack, C. C. *et al.*, 2001; but see Movshon, J. A. *et al.* 2003).

2.11.23 Selective Motion Integration

Integrationist models are useful for combining visual motion signals across 2D space, but a significant difficulty arises when multiple objects are present in 3D space. In this case it does not make sense to integrate across them, as they may be moving in different directions. A particularly elegant example is the case of a moving plaid stimulus, in which the two component gratings are located at different depths from the observer. Psychophysical work has shown that observers naturally segmented the two gratings to form a noncoherent percept when the depth ordering is defined by transparency (Stoner, G. R. and Albright, T. D., 1990), binocular disparity (Stoner, G. R. and Albright, T. D., 1998), surface segmentation (Trueswell, J. C. and Hayhoe, M. M. 1993; Stoner, G. R. and Albright, T. D., 1996), or occlusion (Lorençeau, J. and Shiffrar, M., 1992; McDermott, J. and Adelson, E. H. 2004). Correlates of selective motion integration based on depth segmentation were subsequently found in MT (Stoner, G. R. and Albright, T. D., 1992; Duncan, R. O. *et al.*, 2000; Thiele, A. and Stoner, G. 2003; Pack, C. C. *et al.*, 2004) and in area posteromedial lateral suprasylvian cortex (PMLS) of the cat (Castelo-Branco, M. *et al.*, 2000). These observations suggest that motion signals are identified early in visual processing and selected for integration based in part on their relative depths.

2.11.24 Theoretical Considerations: Redundancy Reduction

Another important source of motivation for selectionist models was a line of thinking, dating back to Attneave F. (1954) and Barlow H. (1961), which considered the goal of early sensory processing to be a reduction in redundancy of the primary sensory information, using principles developed by Claude Shannon in his seminal work on information theory (Shannon, C. E., 1948; Shannon, C. E. and Weaver, W., 1963). One of the important ideas that emerged from this field was the notion that regions of the image that were uniform in space and time were highly redundant and should be ignored in favor of regions of change. From this perspective, features of retinal RFs can be thought of as performing the mathematical operation of differentiation of the image luminance function with respect to various parameters, including: (1) visual space, observed as the RF property known as center-surround opponency (Kuffler, S. W., 1953; Hartline, H. K. and Ratliff, F., 1957), (2) time, manifest as adaptation (Hartline, H. K., 1941), (3) space–time, seen as sensitivity to movement (Hassenstein, B. and Reichardt, W., 1956; Reichardt, W., 1961; Barlow, H. B. and Levick, W. R., 1965), and iv) chromaticity, as evidenced by color opponency (Svaetichin, G. and MacNichol, E. F., Jr., 1959). Likewise, once a representation of the local orientation of contours has been performed by the cortex (Hubel, D. H. and Wiesel, T. N., 1962), one can construct a derivative for

orientation with respect to visual space, an operation which would identify regions of high curvature or other 2D features, which, as it turns out, are highly informative about shape (Attneave, F., 1954; Hubel, D. H. and Livingstone, M. S., 1987; Dobbins, A. *et al.*, 1989). In addition, as described above, such a representation can be further elaborated upon to compute the motion of 2D features and thus solve the aperture problem (Pack, C. C. *et al.*, 2003b). This is the basic approach of many selectionist models.

2.11.25 Selectionist Models

The theoretical considerations of forming efficient representations, combined with the psychophysical and physiological observations demonstrating the potency of the motion of terminators, have yielded a variety of models that we term selectionist. The central idea behind these models is that not all local motion vectors are treated equally. Rather, regions of redundant (and ambiguous) 1D motion vectors, such as those emanating from edges, are filtered out or, put another way, highly informative regions, such as those emanating from corners or line-endings, are selected prior to integration.

A key distinguishing feature of these models is how they have implemented the selection process. One approach, championed by Barth and colleagues (Zetzsche, C. and Barth, E., 1990; Barth, E., 2000) uses a geometric formulation of the luminance function as a hypersurface, and the operation of selecting regions of curvature (as measured by the Riemann tensor), to effectively model the kind of end-stopped, direction selective cells shown in Figures 10 and 11 and thus directly represent 2D velocity. The important point about these models is that they have spatial filters that respond selectively to image regions that contain multiple orientations. There are many ways to design such an operator, but as pointed out by Barth and colleagues all of them must involve nonlinear operations, the simplest nonlinearity being multiplication (Zetzsche, C. and Barth, E., 1990).

Along these lines, Skottun B. C. (1998; 1999) developed neuronal models that simply multiply the responses of orientation-tuned linear filters to arrive at selectivity for 2D features. This operation is sufficient to reproduce many of the results on both end-stopping (Skottun, B. C., 1998) and plaid tuning (Skottun, B. C., 1999). The multiplication operation is used for mathematical convenience, but in principle it could be approximated by any number of the standard nonlinearities known to affect neuronal responses (Tal, D. and Schwartz, E. L., 1997).

Another important model was put forward by Nowlan S. J. and Sejnowski T. J. (1995), and it is selectionist in spirit, even though the selective calculation is nominally performed at the MT stage of the model. In their case, the nonlinear step selects regions of change in local motion vectors by simulating MT neurons with suppressive surrounds (Allman, J. *et al.*, 1985). Such neurons are inhibited when motion is in the same direction in both center and surround, so they respond poorly, if at all, to the motion of 1D contours. This selectivity network then retinotopically gates a second, parallel integration network that pools the motion signals from the selected regions (Figure 12). Thus, for example, for a moving square, MT neurons in the selection network having RF centers overlying the contours respond weakly because there are similar directions of motion in both their RF centers and surrounds.

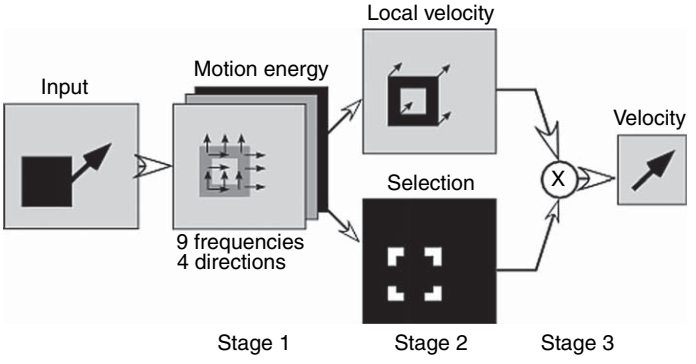

Figure 12 Selectionist model proposed by Nowlan S. J. and Sejnowski T. J. (1995). See text for details. Adapted from figure 2 of Nowlan, S. J. and Sejnowski, T. J. 1995. A selection model for motion processing in area MT of primates. J. Neurosci. 15, 1195–1214.

Surround-inhibited neurons with RF centers positioned at the corners are not effectively suppressed, resulting in a selection of these regions for pooling by the integration network, effectively producing feature tracking. The same logic allows the model to effectively track the motion of 2D intersections in plaid stimuli.

2.11.26 Hybrid Models

Thus the selectionist category of models can be thought of as instantiating a kind of feature tracking, whereas the integrationist models implement an IOC computation. The perceptual data, however, suggest that a single computational strategy will not suffice: under some circumstances, an IOC or feature-tracking rule seems to be in effect, whereas for others, a VA best describes the integration process. This lack of parsimony has motivated computational modelers to seek more general descriptions that can better accommodate the variety of behaviors. A notable recent effort by Weiss Y. *et al.* (2002) essentially embeds many of the features of the IOC calculation in a Bayesian model. The Bayesian calculation involves weighting the first-stage inputs according to the uncertainty associated with the local motion measurement, with the result that 2D features strongly influence the velocity–space calculation in the second stage. Because the velocity–space constraints are subject to noise, the constraint lines are more or less fuzzy depending on the contrast of the stimulus. The stimulus-based evidence is multiplied by a second probability distribution that a priori favors slow speeds (a symmetrical 2D Gaussian centered about the origin in velocity space) in order to obtain the posterior probability in favor of a given 2D object velocity.

These properties endow the model with a rather intuitively appealing behavior that explains much of the existing psychophysical data (Weiss, Y. and Adelson, E. H., 1998). The two key observations are, (1) that the VA percept tends to dominate in conditions where the stimulus-based evidence is relatively weak (low contrast, short duration, or a narrow range of component orientations), and (2) the velocity predicted by the VA is always slower than that of the IOC. The model's behavior can be understood in terms of the relative influence of the prior favoring slow speeds. When the stimulus-based evidence is strong, the prior has little effect and the result is essentially an IOC. When the stimulus-based evidence is weak, the likelihood functions are dim and fuzzy, and the prior has a relatively large effect, pushing the result towards a slower speed, which approaches the VA.

In a related study, Stocker A. A. and Simoncelli E. P. (2006) used a speed discrimination task to estimate the slowness prior and the likelihood functions at different speeds and contrasts for individual observers. Their measurements revealed significant differences between the measured functions and those assumed by the model of Weiss Y. *et al.* (2002). For example, the slowness prior was not well described by a Gaussian, but rather flattened out at higher speeds, and the inverse relationship between the width of the likelihood function and stimulus contrast was different for different observers. It will be important to test whether such differences can be used to better account for individual performance on other motion integration tasks, such as those used by Weiss Y. *et al.* (2002).

Weiss Y. *et al.* (2002) did not hypothesize a physiological basis for the model's computations, but many of the phenomena reported are also consistent with physiological properties of end-stopped V1 cells. End-stopping (and nonlinear contextual effects in general) is weak or nonexistent for stimuli of low contrast (Levitt, J. B. and Lund, J. S., 1997; Polat, U. *et al.*, 1998; Sceniak, M. P. *et al.*, 1999) and takes time to manifest itself in the neuronal response (Pack, C. C. *et al.*, 2003b). Under such circumstances, the local velocity measurements made in V1 are essentially 1D, as postulated by integrationist models, although the perceived direction is more similar to a VA than an IOC. For high-contrast stimuli end-stopping is effective, allowing 1D motion signals to be filtered out, thus emphasizing the veridical velocity of 2D features. Both of these observations are at least qualitatively consistent with the Bayesian computation described by Weiss Y. *et al.* (2002).

2.11.27 Future Challenges

While there has been much speculation about the role of various direction-selective neurons in motion integration and the solution to the aperture problem (Albright, T. D., 1984; Movshon, J. A. *et al.*, 1985; Pack, C. C. *et al.*, 2003b), there has been a complete lack of experiments in which these neural signals are directly compared to perceptual reports. For example, while Pack C. C. *et al.* (2003b) demonstrated that end-stopped, direction-selective neurons in V1

could, in principle, provide faithful 2D velocity measurements of features, there is no direct evidence that such neurons are actually used by perceptual systems (or even that these neurons project to MT, though, statistically, many are likely to do so). Similarly, the evidence that V1 neurons projecting to MT provide only 1D motion information (Movshon, J. A. and Newsome, W. T., 1996) cannot be considered definitive given the small sample size, the considerable oversampling of layer 6 neurons, which are known to prefer long bars (Gilbert, C. D., 1977; Sceniak, M. P. *et al.*, 2001), and the aforementioned shortcomings of the plaid test for character-izing end-stopped neurons. What is needed is a demonstration of which neurons are actively contributing to a given percept at a given time. Such evidence would need to consist of trial-by-trial correlations between the activity of these neurons and the choices of the monkey on a suitable perceptual task (Parker, A. J. and Newsome, W. T., 1998). For example, if pattern cells are important for the perception of coherent motion, then, for ambiguous plaid stimuli for which the animal may report either coherence or transparency, one would expect pattern cells to fire at higher rates during trials on which the animal reports coherence. An inverse relationship would be predicted for component cells. Such measures of choice probability (Britten, K. H. *et al.*, 1996) would strongly support the proposed roles of these neurons. Furthermore, insofar as there is spatial clustering of pattern or component neurons in MT or some other visual structure, microstimulation experiments could be used to attempt to bias an animal's reports in predictable ways, as has been done for perceived direction of motion in random dot displays (Salzman, C. D. *et al.*, 1990; 1992).

Both a challenge and a potential boon to studies linking neurophysiology to percepts of motion integration is the fact that many of the stimuli used to study it are perceptually bi- or even multistable. This is true of both plaids (Stoner, G. R. and Albright, T. D., 1996; Hupé, J. M. and Rubin, N., 2003) and barber poles (Rubin, N. and Hochstein, S., 1993; Castet, E. *et al.*, 1999). For example, prolonged viewing of plaid stimuli produces spontaneous perceptual transitions between coherence and transparency (Hupé, J. M. and Rubin, N., 2003), and this is true even for plaids whose intersection luminance values are not consistent with transparency (see Relevant Websites section). Such phenomena argue for an active, ongoing competition amongst local motion cues and will present future models with challenges similar to those encountered by models of binocular rivalry (Leopold, D. A. and Logothetis, N. K., 1999). Furthermore, the plaid multistability (Hupé, J. M. and Rubin, N., 2003) and the profound influence of shape and occlusion cues on the integration of motion (Lorençeau, J. and Shiffrar, M., 1992; Lorençeau, J. and Alais, D., 2001), underline the fact that integration and segmentation are tightly linked, even at early stages of visual processing.

In fact, relatively few motion models address the complications that arise when the visual input contains multiple objects. In such cases motion integration appears to be influenced by a mechanism that performs segmentation, presumably to avoid integrating over multiple objects or surfaces (Shimojo, S. *et al.*, 1989). The segmentation system has proven difficult to model, in part because it is influenced by nonmotion cues, especially perceived depth as defined by retinal disparity, T-junctions, or transparency. Most segmentation models involve the explicit detection of motion boundaries based on depth cues, along with the smoothing of motion signals across individual objects or surfaces (Nowlan, S. J. and Sejnowski, T. J., 1995; Koechlin, E. *et al.*, 1999; Lidén, L. and Pack, C., 1999; Bulthoff, H. *et al.*, 1989; Grossberg, S. *et al.*, 2001; Bayerl, P. and Neumann, H., 2004).

The smoothing in these models is often accomplished by a winner-take-all interaction of motion signals across networks of neurons with small RFs. This process is motivated by the perceptual observation that objects appear to move rigidly, despite the presumed diversity in local motion measurements. That is, each part of an object appears to move in the same direction, even when small RFs are reporting different velocities. However, evidence for this kind of smoothing process in V1 has not been found (Pack, C. C. *et al.*, 2004), even for stimuli that should engage the perceptual winner-take-all mechanism. In fact for barber-pole stimuli the responses in MT can be estimated with reasonable precision based on a local end-stopping procedure observable at the level of individual V1 neurons (Pack, C. C. *et al.*, 2004). However, neither the V1 nor the MT responses appear to be entirely consistent with perceptual observations, suggesting that it may make sense to start the search for neural correlates of percep-tual smoothing in regions of the cortex beyond V1 and MT.

2.11.28 Final Thoughts

As we have tried to emphasize, the existing evidence suggests that under some circumstances observers

perceive motion that is consistent with the standard integrationist model (Simoncelli, E. P. and Heeger, D. J., 1998). In contrast, there are many exceptions in the psychophysical literature, and the results of physiological tests of the model have been equivocal at best.

Evidence for 2D feature selectivity in V1 has been found in a few studies, although it is not clear how this selectivity influences MT responses. In this regard a significant problem is the lack of a standard selectionist model that can be meaningfully compared to the standard integrationist model. While many selectionist models exist, they tend to be substantially more complicated than the integrationist models discussed here, and as a result the predictions are less clear, rendering a detailed comparison problematic. Thus an important avenue for future research will be the development of a standard selectionist model. There is no reason why this cannot be done, since as we have pointed out many of the core computations are the same in both types of models.

Endnotes

[1] (Under certain conditions, such as when the components are configured to simulate transparent occlusion, even plaids with identical components can fail to cohere (Stoner *et al.*, 1990).)

[2] Even though the pattern direction would be the same as the grating direction under the plaid prediction, there is no good reason why the tuning curves should be identical. They are, after all, quite different stimuli with different perceptual properties. Nevertheless, it is an intuitively simple way to set up the test.

[3] The model's response to a component grating moving in direction d is given by

$$\left[\left[\left[\sum_i g(p_i - d)h(P - p_i)\right] - \theta\right]^+\right]^n$$

where g is a Gaussian function that describes the ith V1 neuron's direction tuning as a function of preferred direction p_i, h is another Gaussian that weights V1 inputs according to the MT neuron's preferred direction P, and θ is a threshold.

[4] The MT neurons described in Albright T. D. (1984) were categorized as belonging to type I or type II, the latter having broader tuning than the former. It was subsequently shown that these two categories corresponded to component and pattern neurons, respectively (Rodman, H. R. and Albright, T. D., 1989). The same explanation, based on differences in tuning bandwidths, also applies to the observed differences in orientation tuning preferences between pattern and component neurons (Rodman, H. R. and Albright, T. D., 1989).

[5] Extrinsic terminators are created by using binocular disparity or other occlusion cues to indicate that the grating endings are accidents of occlusion (Shimojo, S. *et al.*, 1989) and therefore not likely to belong to the moving object. Two distinguish the two types of terminators (those belonging to an object versus those that are created by occlusion) psychophysicists refer to intrinsic and extrinsic terminators, respectively. This classification, though useful, is undoubtedly too simple. Experiments by Castet E. *et al*. (1999) indicate that there may be different degrees of extrinsicness.

[6] An alternative possibility is that a linear neuron could be selective for the high spatial frequencies found near the endpoints of the moving edge. However, this does not appear to be the case for MT neurons, which generally prefer low spatial frequencies.

[7] It has often been argued that the time course observed in this experiment constitutes evidence for linear responses in MT, since the motion components moving in the global stimulus direction are low in amplitude, and hence might be expected to have longer latencies. While it is entirely possible that the 2D features are processed more slowly than 1D features, the stronger hypothesis of linearity merely begs the question of why the low-amplitude component dominates the neuronal response.

[8] Since the direction of the 1D grating motion was always 45° from the direction of either set of terminators, these experiments did not test for the influence of grating direction (capture) found psychophysically by Castet E. *et al*. (1999).

[9] Another compelling example of such a neuron is shown in one of Hubel and Wiesel's now-classic movies (see Relevant Website section).

[10] Technically, both types of maps are second-order subunits, since the response to the individual spots is independent of contrast. To avoid confusion we continue to refer only to the motion maps (Figures 11(c) and 11(f)) as second-order subunits.

References

Adelson, E. H. and Bergen, J. R. 1985. Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A 2, 284–299.

Adelson, E. H. and Movshon, J. A. 1982. Phenomenal coherence of moving visual patterns. Nature 300, 523–525.

Albright, T. D. 1984. Direction and orientation selectivity of neurons in visual area MT of the macaque. J. Neurophysiol. 52, 1106–1130.

Albright, T. D. and Desimone, R. 1987. Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. Exp. Brain Res. 65, 582–592.

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

Anderson, J. S., Lampl, I., Gillespie, D. C., and Ferster, D. 2001. Membrane potential and conductance changes underlying length tuning of cells in cat primary visual cortex. J. Neurosci. 21, 2104–2112.

Attneave, F. 1954. Some informational aspects of visual perception. Psychol. Rev. 61, 183–193.

Barlow, H. 1961. Possible Principles Underlying the Transformation of Sensory Messages. In: Sensory Communication (*ed*. W. A. Rosenblith), pp. 217–234. MIT Press.

Barlow, H. B. and Levick, W. R. 1965. The mechanism of directionally selective units in rabbit's retina. J. Physiol. (Lond.) 178, 477–504.

Barth, E. 2000. A geometric view on early and middle level visual coding. Spat. Vis. 13, 193–199.

Bayerl, P. and Neumann, H. 2004. Disambiguating visual motion through contextual feedback modulation. Neural Comput. 16, 2041–2066.

Blasdel, G. G. and Fitzpatrick, D. 1984. Physiological organization of layer 4 in macaque striate cortex. J. Neurosci. 4, 880–895.

Born, R. T. and Bradley, D. C. 2005. Structure and function of visual area MT. Annu. Rev. Neurosci. 28, 157–189.

Bowns, L. 1996. Evidence for a feature tracking explanation of why type II plaids move in the vector sum direction at short durations. Vision Res. 36, 3685–3694.

Bradley, D. C., Goyal, M. S., and Scott, B. B. 2005. Pattern Velocity Computation by Primate MT Neurons. Program No. 136.9. 2005, Abstract Viewer/Itinerary Planner, Washington, DC: Society for Neuroscience.

Britten, K. H. and Heuer, H. W. 1999. Spatial summation in the receptive fields of MT neurons. J. Neurosci. 19, 5074–5084.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., and Movshon, J. A. 1996. A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis. Neurosci. 13, 87–100.

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

Bulthoff, H., Little, J., and Poggio, T. 1989. A parallel algorithm for real-time computation of optical flow. Nature 337, 549–555.

Carandini, M. and Ferster, D. 2000. Membrane potential and firing rate in cat primary visual cortex. J. Neurosci. 20, 470–484.

Carandini, M., Heeger, D. J., and Movshon, J. A. 1997. Linearity and normalization in simple cells of the macaque primary visual cortex. J. Neurosci. 17, 8621–8644.

Castelo-Branco, M., Goebel, R., Neuenschwander, S., and Singer, W. 2000. Neural synchrony correlates with surface segregation rules. Nature 405(6787), 685–689.

Castet, E., Charton, V., and Dufour, A. 1999. The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli. Vision Res. 39, 915–932.

Cavanaugh, J. R., Bair, W., and Movshon, J. A. 2002. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. J. Neurophysiol. 88, 2530–2546.

Churchland, M. M., Priebe, N. J., and Lisberger, S. G. 2005. Comparison of the spatial limits on direction selectivity in visual areas MT and V1. J. Neurophysiol. 93, 1235–1245.

Courellis, S. H. and Marmarelis, V. Z. 1992. Nonlinear Functional Representations for Motion Detection and Speed Estimation Schemes. In: Nonlinear Vision (*eds*. B. Nabet and R. B. Pinter), pp. 91–108. CRC Press.

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. 1993. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. J. Neurophysiol. 69, 1118–1135.

Derrington, A. and Suero, M. 1991. Motion of complex patterns is computed from the perceived motions of their components. Vision Res. 31, 139–149.

Dobbins, A., Zucker, S. W., and Cynader, M. S. 1989. Endstopping and curvature. Vision Res. 29, 1371–1387.

Dow, B. M. 1974. Functional classes of cells and their laminar distribution in monkey visual cortex. J. Neurophysiol. 37, 927–946.

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

Duncan, R. O., Albright, T. D., and Stoner, G. R. 2000. Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. J. Neurosci. 20, 5885–5897.

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

Emerson, R. C., Citron, M. C., Vaughn, W. J., and Klein, S. A. 1987. Nonlinear directionally selective subunits in complex cells of cat striate cortex. J. Neurophysiol. 58, 33–65.

Fennema, C. and Thompson, W. 1979. Velocity determination in scenes containing several moving objects. Comp. Graph. Image Process. 9, 301–315.

Ferrera, V. P. and Wilson, H. R. 1987. Direction specific masking and the analysis of motion in two dimensions. Vision Res. 27, 1783–1796.

Ferrera, V. P. and Wilson, H. R. 1990. Perceived direction of moving two-dimensional patterns. Vision Res. 30, 273–287.

Gilbert, C. D. 1977. Laminar differences in receptive field properties of cells in cat primary visual cortex. J. Physiol. 268, 391–421.

Grossberg, S., Mingolla, E., and Viswanathan, L. 2001. Neural dynamics of motion integration and segmentation within and across apertures. Vision Res. 41, 2521–2553.

Grzywacz, N. M. and Yuille, A. L. 1990. A model for the estimate of local image velocity by cells in the visual cortex. Proc R Soc Lond B Biol Sci. 239(1295), 129–161.

Guo, K., Benson, P. J., and Blakemore, C. 2004. Pattern motion is present in V1 of awake but not anaesthetized monkeys. Eur. J. Neurosci. 19, 1055–1066.

Hartline, H. K. 1941. The neural mechanisms of vision. Harvey Lectures Ser. 37, 39.

Hartline, H. K. and Ratliff, F. 1957. Inhibitory interaction of receptor units in the eye of Limulus. J. Gen. Physiol. 40, 357–376.

Hassenstein, B. and Reichardt, W. 1956. Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers, Chlorophanus. Z. Naturforsch. Teil. B 11, 513–524.

Hawken, M. J., Parker, A. J., and Lund, J. S. 1988. Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the Old World monkey. J. Neurosci. 8, 3541–3548.

Heeger, D. J. 1987. Model for the extraction of image flow. J. Opt. Soc. Am. A 4, 1455–1471.

Heeger, D. J., Simoncelli, E. P., and Movshon, J. A. 1996. Computational models of cortical visual processing. Proc. Natl. Acad. Sci. U. S. A. 93, 623–627.

Heuer, H. W. and Britten, K. H. 2002. Contrast dependence of response normalization in area MT of the rhesus macaque. J. Neurophysiol. 88, 3398–3408.

Hubel, D. H. and Livingstone, M. S. 1987. Segregation of form, color, and stereopsis in primate area 18. J. Neurosci. 7, 3378–3415.

Hubel, D. H. and Wiesel, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160, 106–154.

Hubel, D. H. and Wiezel, T. N. 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J. Neurophysiol. 28, 229–289.

Humphrey, D. R. and Corrie, W. S. 1978. Properties of pyramidal tract neuron system within a functionally defined subregion of primate motor cortex. J. Neurophysiol. 41, 216–243.

Hupé, J. M. and Rubin, N. 2003. The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. Vision Res. 43, 531–548.

Jones, H. E., Grieve, K. L., Wang, W., and Sillito, A. M. 2001. Surround suppression in primate V1. J. Neurophysiol. 86, 2011–2028.

Kapadia, M. K., Westheimer, G., and Gilbert, C. D. 1999. Dynamics of spatial summation in primary visual cortex of alert monkeys. Proc. Natl. Acad. Sci. U. S. A. 96, 12073–12078.

Kawakami, S. and Okamoto, H. 1996. A cell model for the detection of local image motion on the magnocellular pathway of the visual cortex. Vision Res. 36, 117–147.

Koechlin, E., Anton, J. L., and Burnod, Y. 1999. Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area MT. Biol. Cybern. 80, 25–44.

Kooi, F. L. 1993. Local direction of edge motion causes and abolishes the barberpole illusion. Vision Res. 33, 2347–2351.

Krekelberg, B. and Albright, T. D. 2005. Motion mechanisms in macaque MT. J. Neurophysiol. 93, 2908–2921.

Kuffler, S. W. 1953. Discharge patterns and functional organization of mammalian retina. J. Neurophysiol. 16, 37–68.

Lamme, V. A., Zipser, K., and Spekreijse, H. 1998. Figure-ground activity in primary visual cortex is suppressed by anesthesia. Proc. Natl. Acad. Sci. U. S. A. 95, 3263–3268.

Lemon, R. 1984. Methods for Neuronal Recording in Conscious Animals, Wiley.

Leopold, D. A. and Logothetis, N. K. 1999. Multistable phenomena: changing views in perception. Trends Cogn. Sci. 3, 254–264.

Levitt, J. B. and Lund, J. S. 1997. Contrast dependence of contextual effects in primate visual cortex. Nature 387, 73–76.

Lidén, L. and Pack, C. 1999. The role of terminators and occlusion cues in motion integration and segmentation: a neural network model. Vision Res. 39, 3301–3320.

Limb, J. O. and Murphy, J. A. 1975. Estimating the velocity of moving images in television signals. Comp. Graph. Image Process. 4, 311–327.

Lisberger, S. G. and Ferrera, V. P. 1997. Vector averaging for smooth pursuit eye movements initiated by two moving targets in monkeys. J. Neurosci. 17, 7490–7502.

Livingstone, M. S. and Hubel, D. H. 1984. Specificity of intrinsic connections in primate primary visual cortex. J. Neurosci. 4, 2830–2835.

Livingstone, M. S., Pack, C. C., and Born, R. T. 2001. Two-dimensional substructure of MT receptive fields. Neuron 30, 781–793.

Lorençeau, J. and Alais, D. 2001. Form constraints in motion binding. Nat. Neurosci. 4, 745–751.

Lorençeau, J. and Shiffrar, M. 1992. The influence of terminators on motion integration across space. Vision Res. 32, 263–273.

Lorençeau, J., Shiffrar, M., Wells, N., and Castet, E. 1993. Different motion sensitive units are involved in recovering the direction of moving lines. Vision Res. 33(9), 1207–1217.

Majaj, N. J., Carandini, M., and Movshon, J. A. 2007. Motion integration by neurons in macaque MT is local, not global. J. Neurosci. 27, 366–370.

Marr, D. 1982. Vision. W.H. Freeman & Co.

Maunsell, J. H. and van Essen, D. C. 1983. The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. J. Neurosci. 3, 2563–2586.

McDermott, J. and Adelson, E. H. 2004. The geometry of the occluding contour and its effect on motion interpretation. J. Vis. 4(10), 944–954.

Mikami, A., Newsome, W. T., and Wurtz, R. H. 1986. Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. J. Neurophysiol. 55, 1328–1339.

Mingolla, E., Todd, J. T., and Norman, J. F. 1992. The perception of globally coherent motion. Vision Res. 32, 1015–1031.

Movshon, J. A. and Newsome, W. T. 1996. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J. Neurosci. 16, 7733–7741.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., and Newsome, W. T. 1985. The Analysis of Moving Visual Patterns. In: Pattern Recognition Mechanisms (*eds*. C. Chagas, R. Gattass, and C. Gross), pp. 117–151. Vatican Press.

Movshon, J. A., Albright, T. D., Stoner, G. R., Majaj, N. J., and Smith, M. A. 2003. Cortical responses to visual motion in alert and anesthetized monkeys. Nat. Neurosci. 6, 3 (reply by Pack, C. C., Berezovskii, V. K. and Born, R. T., pp. 3–4).

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

Nowlan, S. J. and Sejnowski, T. J. 1995. A selection model for motion processing in area MT of primates. J. Neurosci. 15, 1195–1214.

Ogata, M. and Sato, T. 1991. Motion perception model with interaction between spatial frequency channels. Syst. Comput. Jap. 22, 30–39.

Okamoto, H., Kawakami, S., Saito, H., Hida, E., Odajima, K., Tamanoi, D., and Ohno, H. 1999. MT neurons in the macaque exhibited two types of bimodal direction tuning as predicted by a model for visual motion detection. Vision Res. 39(20), 3465–3479.

Pack, C. C. and Born, R. T. 2001. Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature 409, 1040–1042.

Pack, C. C., Berezovskii, V. K., and Born, R. T. 2001. Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. Nature 414, 905–908.

Pack, C. C., Born, R. T., and Livingstone, M. S. 2003a. Two-dimensional substructure of motion and stereo interactions in primary visual cortex of alert macaque. Neuron 37, 525–535.

Pack, C. C., Conway, B. R., Born, R. T., and Livingstone, M. S. 2006. Spatiotemporal structure of nonlinear subunits in macaque visual cortex. J. Neurosci. 26, 893–907.

Pack, C. C., Gartland, A. J., and Born, R. T. 2004. Integration of contour and terminator signals in visual area MT of alert macaque. J. Neurosci. 24, 3268–3280.

Pack, C. C., Livingstone, M., Duffy, K., and Born, R. T. 2003b. End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. Neuron 39, 671–680.

Parker, A. J. and Newsome, W. T. 1998. Sense and the single neuron: probing the physiology of perception. Annu. Rev. Neurosci. 21, 227–277.

Perrone, J. A. and Thiele, A. 2001. Speed skills: measuring the visual speed analyzing properties of primate MT neurons. Nat. Neurosci. 4(5), 526–532.

Poggio, T. and Reichardt, W. 1973. Considerations on models of movement detection. Kybernetik 13, 223–227.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., and Norcia, A. M. 1998. Collinear stimuli regulate visual responses depending on cell's contrast threshold. Nature 391, 580–584.

Power, R. P. and Moulden, B. 1992. Spatial gating effects on judged motion of gratings in apertures. Perception 21, 449–463.

Priebe, N. J., Cassanello, C. R., and Lisberger, S. G. 2003. The neural representation of speed in macaque area MT/V5. J. Neurosci. 23, 5650–5661.

Priebe, N. J., Lisberger, S. G., and Movshon, J. A. 2006. Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. J. Neurosci. 26, 2941–2950.

Qian, N. and Andersen, R. A. 1994. Transparent motion perception as detection of unbalanced motion signals. II. Physiology. J. Neurosci. 14, 7367–7380.

Recanzone, G. H., Wurtz, R. H., and Schwarz, U. 1997. Responses of MT and MST neurons to one and two moving objects in the receptive field. J. Neurophysiol. 78, 2904–2915.

Reichardt, W. 1961. Autocorrelation, a Principle for the Evaluation of Sensory Information by the Central Nervous System. In: Sensory Communications (*ed*. W. A. Rosenblith), pp. 303–318. Wiley.
Rodman, H. R. and Albright, T. D. 1989. Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). Exp. Brain Res. 75(1), 53–64.
Rubin, N. and Hochstein, S. 1993. Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. Vision Res. 33, 1385–1396.
Rust, N. C., Mante, V., Simoncelli, E. P., and Movshon, J. A. 2006. How MT cells analyze the motion of visual patterns. Nat. Neurosci. 9, 1421–1431.
Salzman, C. D., Britten, K. H., and Newsome, W. T. 1990. Cortical microstimulation influences perceptual judgements of motion direction. Nature 346, 174–177.
Salzman, C. D., Murasugi, C. M., Britten, K. H., and Newsome, W. T. 1992. Microstimulation in visual area MT: effects on direction discrimination performance. J. Neurosci. 12, 2331–2355.
Sceniak, M. P., Hawken, M. J., and Shapley, R. 2001. Visual spatial characterization of macaque V1 neurons. J. Neurophysiol. 85, 1873–1887.
Sceniak, M. P., Ringach, D. L., Hawken, M. J., and Shapley, R. 1999. Contrast's effect on spatial summation by macaque V1 neurons. Nat. Neurosci. 2, 733–739.
Sekuler, R. W. and Ganz, L. 1963. Aftereffect of seen motion with a stabilized retinal image. Science 139, 419–420.
Sereno, M. E. 1993. Neural Computation of Pattern Motion. MIT Press.
Shannon, C. E. 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423.
Shannon, C. E. and Weaver, W. 1963. The Mathematical Theory of Communication. University of Illinois Press.
Shimojo, S., Silverman, G. H., and Nakayama, K. 1989. Occlusion and the solution to the aperture problem for motion. Vision Res. 29, 619–626.
Shipp, S. and Zeki, S. 1989. The organization of connections between areas V5 and V1 in macaque monkey visual cortex. Eur. J. Neurosci. 1, 309–332.
Simoncelli, E. P. and Heeger, D. J. 1998. A model of neuronal responses in visual area MT. Vision Res. 38, 743–761.
Simoncelli, E. P., Bair, W. D., Cavanaugh, J. R., and Movshon, J. A. 1996. Testing and refining a computational model of neuronal responses in area MT. Investigative Ophthalmology and Visual Science 37, 916.
Sincich, L. C. and Horton, J. C. 2003. Independent projection streams from macaque striate cortex to the second visual area and middle temporal area. J. Neurosci. 23, 5684–5692.
Skottun, B. C. 1998. A model for end-stopping in the visual cortex. Vision Res. 38, 2023–2035.
Skottun, B. C. 1999. Neuronal responses to plaids. Vision Res. 39, 2151–2156.
Snowden, R. J., Treue, S., Erickson, R. G., and Andersen, R. A. 1991. The response of area MT and V1 neurons to transparent motion. J. Neurosci. 11, 2768–2785.
Stocker, A. A. and Simoncelli, E. P. 2006. Noise characteristics and prior expectations in human visual speed perception. Nat. Neurosci. 9, 578–585.
Stoner, G. R. and Albright, T. D. 1992. Neural correlates of perceptual motion coherence. Nature 358, 412–414.
Stoner, G. R. and Albright, T. D. 1996. The interpretation of visual motion: evidence for surface segmentation mechanisms. Vision Res. 36(9), 1291–1310.
Stoner, G. R. and Albright, T. D. 1998. Luminance contrast affects motion coherency in plaid patterns by acting as a depth-from-occlusion cue. Vision Res. 38(3), 387–401.
Svaetichin, G. and MacNichol, E. F., Jr. 1959. Retinal mechanisms for chromatic and achromatic vision. Ann. N. Y. Acad. Sci. 74, 385–404.
Tal, D. and Schwartz, E. L. 1997. Computing with the leaky integrate-and-fire neuron: logarithmic computation and multiplication. Neural Comput. 9, 305–318.
Thiele, A. and Stoner, G. 2003. Neuronal synchrony does not correlate with motion coherence in cortical area MT. Nature 421(6921), 366–370.
Tinsley, C. J., Webb, B. S., Barraclough, N. E., Vincent, C. J., Parker, A., and Derrington, A. M. 2003. The nature of V1 neural responses to 2D moving patterns depends on receptive-field structure in the marmoset monkey. J. Neurophysiol. 90, 930–937.
Towe, A. L. and Harding, G. W. 1970. Extracellular microelectrode sampling bias. Exp. Neurol. 29, 366–381.
Trueswell, J. C. and Hayhoe, M. M. 1993. Surface segmentation mechanisms and motion perception. Vision Res. 33(3), 313–328.
Ullman, S. 1979. The Interpretation of Visual Motion, MIT Press.
Van Essen, D. C., Newsome, W. T., and Maunsell, J. H. 1984. The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. Vision Res. 24, 429–448.
van Santen, J. P. and Sperling, G. 1985. Elaborated Reichardt detectors. J. Opt. Soc. Am. A 2, 300–321.
Wallach, H. 1935. Uber visuell wahrgenommene Bewegungsrichtung. Psychol. Forsch. 20, 325–380.
Watson, A. B. and Ahumada, A. J. 1985. Model of human visual-motion sensing. J Opt Soc Am A. 2(2), 322–341.
Weiss, Y. and Adelson, E. H. 1998. Slow and smooth: a Bayesian theory for the combination of local motion signals in human vision. AI Memo 1624.
Weiss, Y., Simoncelli, E. P., and Adelson, E. H. 2002. Motion illusions as optimal percepts. Nat. Neurosci. 5, 598–604.
Welch, L. 1989. The perception of moving plaids reveals two motion-processing stages. Nature 337, 734–736.
Wiener, N. 1958. Nonlinear Problems in Random Theory. MIT Press.
Wuerger, S., Shapley, R., and Rubin, N. 1996. ''On the visually perceived direction of motion'' by Hans Wallach: 60 years later. Perception 25, 1317–1367.
Yo, C. and Wilson, H. R. 1992. Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. Vision Res. 32, 135–147.
Zetzsche, C. and Barth, E. 1990. Fundamental limits of linear filters in the visual processing of two-dimensional signals. Vision Res. 30, 1111–1117.

Further Reading

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

Relevant Websites

http://www.cns.nyu.edu/~hupe/plaid_demo/demo_plaids.html

http://viperlib.york.ac.uk/scripts/Portweb.dll?field=filename&op=matches&value=HypercomplexCortCell250.mpg&template=main_record&catalog=protol