2

_____

TEMPORAL DYNAMICS OF MOTION INTEGRATION

_____

RICHARD T. BORN[1] JAMES M. G. TSUI[2] & CHRISTOPHER C. PACK[2]

[1]Department of Neurobiology, Harvard Medical School, Boston, MA, USA
[2]Montreal Neurological Institute, School of Medecine, McGill University, Montreal, Quebec, Canada

Abstract

In order to correctly determine the velocity of moving objects, the brain must integrate information derived from a large number of local detectors. The geometry of objects, the presence of occluding surfaces and the restricted receptive fields of early motion detectors conspire to render many of these measurements unreliable or downright misleading. One possible solution to this problem, often referred to as the "aperture problem," involves differential weighting of local cues according to their fidelity: measurements made near 2-dimensional object features called "terminators" are selectively integrated, whereas 1-dimensional motion signals emanating from object contours are given less weight. A large number of experiments have assessed the integration of these different kinds of motion cues using perceptual reports, eye movements and neuronal activity. All of the results show striking qualitative similarities in the temporal sequence of integration: the earliest responses reveal a non-selective integration which becomes progressively selective over time. In this review we propose a simple mechanistic model based on endstopped, direction-selective neurons in V1 of the macaque, and use it to account for the dynamics observed in perception, eye movements, and neural responses in MT.

1. Introduction
Temporal dynamics of perception and the "aperture problem"

Perception is neural computation, and, because neurons are relatively slow computational devices, perception takes time. On the one hand, this sluggish processing is a potential detriment to an animal's survival, and we might expect at least certain perceptual computations to be highly optimized for speed. On the other hand, the relative slowness of some neural systems may be of benefit to the investigator attempting to understand the circuitry responsible for the computation. Indeed, the temporal evolution of perceptual capacities has been exploited by psychophysicists for many years. By measuring reaction times, limiting viewing times, or using clever tricks such as masking to interrupt perceptual processes at different times, they have gained valuable insights into the nature of successive stages of perceptual computations.

One of the general themes that has arisen from this body of work is the idea that, when presented with a novel stimulus, perceptual systems first rapidly compute a relatively rough estimate of the stimulus content and then gradually refine this estimate over time. This is demonstrated, for example, by the fact that human observers require less viewing time to recognize the general category to which an object belongs than to identify the specific object (Rosch, Mervis, Gray, Johnson & Boyes-Braem 1976; Thorpe & Fabre-Thorpe 2001). Similarly, the recovery of stereoscopic depth by comparing images between the two eyes appears to follow a coarse-to-fine progression, with large spatial scales being processed before fine details (Marr & Poggio 1976; Wilson, Blake & Halpern, 1991; Rohaly & Wilson 1993; Rohaly & Wilson 1994). And, as we will describe in some detail below, the visual motion system uses a similar strategy to compute the direction of motion of objects. Such a strategy may reflect the genuine computational needs of sensory systems—such as the use of coarse stereo matches to constrain subsequent fine ones in order to solve the correspondence problem (Marr, Ullman & Poggio, 1979)—as well as selective pressures for animals to be able to rapidly initiate behavioral responses, even in the absence of perfect, or detailed, information.

In this review, we will consider these issues from the perspective of visual motion perception. A solid object can only be moving in one direction at any given time, yet sampling the motion of small regions of the object can result in disparate estimates of this direction. This constraint on the measurement of motion direction is highly relevant to the visual systems of humans and other animals, in which early visual structures have neurons with small receptive fields. A more concrete way of thinking about the limited receptive field size of these visual neurons is as "apertures," depicted as circles in the inset of Fig 1a. These apertures, in conjunction with the geometry of moving objects, create local motion signals that are frequently ambiguous. For example, if a square-shaped object moves upwards and to the right, a neuron with a small receptive field positioned along one of the object's vertical edges can measure only the rightward component of motion. This measurement is ambiguous, because it is consistent with many possible directions of actual object motion. In general a motion measurement made from a one-dimensional (1D) feature will always be ambiguous, because no change can be measured in the direction parallel to the contour. Only neurons whose receptive fields are positioned over a two-dimensional (2D) feature, such as a corner of the square object and often referred to in the literature as a "terminator," can measure the direction of object motion accurately.

Psychophysics of motion integration

A large body of experimental and theoretical work has addressed the question of how various local motion measurements are integrated to produce veridical calculations of object motion. Our purpose here is not to review the entire literature (for this, see Pack & Born, 2007), but rather to focus on one particular aspect of the computation, namely its temporal dynamics, that may be of particular use in elucidating the neural circuitry that carries it out.

The starting point for this project is the observation that observers make systematic perceptual errors when certain stimuli are viewed for a short amount of time (Lorençeau, Shiffrar, Wells & Castet, 1993). That is, the visual system's initial calculations are *not* always veridical. This can be appreciated directly from Movie 1 in which a long, low contrast bar moves obliquely with respect to its long axis. While fixating the red square, most observers see the bar following a curved trajectory, beginning with an upwards component that then bends around to the right. In reality the motion is purely horizontal, so this initial upwards component would seem to be a direct manifestation of the aperture problem: of the many direction-selective neurons whose receptive fields would be confined to the bar's contour, those that should respond maximally are those whose preferred direction is up and to the right; hence the mistaken percept.

This phenomenon was explored by Lorençeau and colleagues (Lorençeau et al. 1993), who asked human observers to report the direction of motion of arrays of moving lines similar to those in Movie 1. The lines were tilted either +20° or -20° from vertical, and they moved along an axis tilted either +20° or -20° from the horizontal. Observers were asked to report whether the vertical component of the motion was upwards or downwards using a 2-alternative forced choice procedure. The key aspects of the experimental design were 1) that neither orientation alone nor a combination of orientation and horizontal direction of motion could be used to solve the task and 2) for a given line orientation, the four possible directions of movement produced two conditions in which motion was perpendicular to the orientation of the lines and two in which it was oblique. Importantly, for the two latter conditions, the tilt of the lines would produce "aperture motion" (that is, local motion measured perpendicular to the contours) whose vertical component was opposite to that of the true direction of line motion. For example, for an array of lines tilted 20° to the left of vertical (counterclockwise), line motion to the right and 20° *downwards* from horizontal would produce aperture motion to the right and 20° *upwards* from horizontal. Thus, for the two test conditions, insofar as the

observers' percepts were influenced by the component of motion perpendicular to line orientation, they should tend to report the wrong direction.
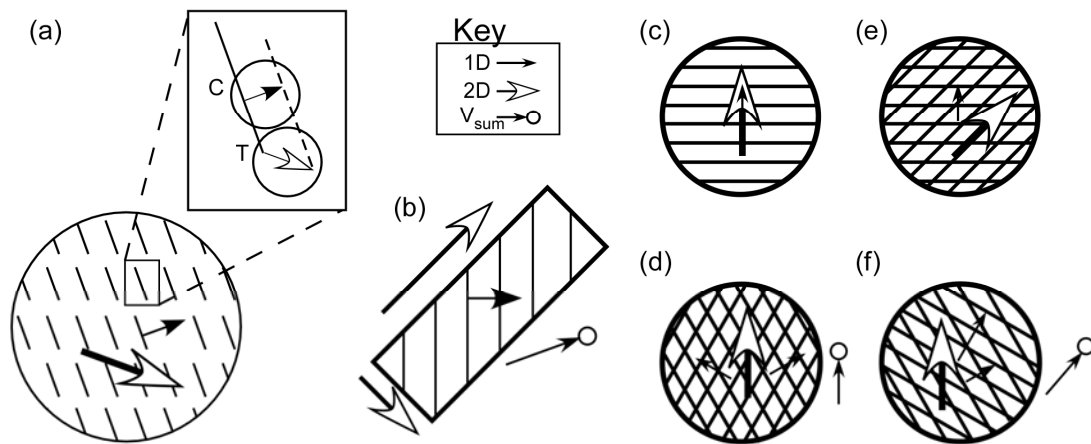

Figure 1: Visual stimuli used to study the dynamics of 1D-to-2D motion. (a) Tilted bar-field used by Lorençeau and colleagues (1993). In this particular example, the 2D direction of motion has a downward component, whereas the 1D direction measured along the contour has an upward component. The inset depicts the situation in greater detail as seen through the apertures of neuronal receptive fields. (b) Barber pole in which the direction of grating motion differs by 45 degrees from that of the perceived direction, which is up and to the right (c) Single grating. (d) Symmetric Type I plaid consisting of two superimposed 1D gratings. (e) Unikinetic plaid. Only the horizontal grating moves (upwards), but the static oblique grating causes the pattern to appear to move up and to the right. (f) Type II plaid in which the perceived direction of the pattern is very different from that of either of the two components or the vector sum. (see also the corresponding movies for each stimulus type)

For the control conditions, the observers' reports were accurate under all stimulus conditions. For the test conditions, however, observers often reported the wrong direction of motion, as if their visual systems had been fooled by the aperture problem. For many conditions, the performance was significantly poorer than chance, indicating that the direction of motion was indeed systematically misperceived and not simply difficult to judge. (If the latter had occurred, performance would have been 50% correct.) The Lorençeau group systematically varied three stimulus parameters—line *length*, line *contrast* and the *duration* of stimulus presentation—in order to probe the conditions under which the visual system was most likely to err. The general result was that for arrays of relatively long lines (~3°) at low contrast (<30%) and presented for short durations (~150 ms), observers never reported the true direction of motion. Conversely, as the lines were made shorter, of higher contrast or were viewed for longer durations, performance improved. Although not all possible

combinations of these three variables were tested, it was clear that they interacted in relatively predictable ways. Thus, for example, even high-contrast (70%) lines of modest lengths (2.5°) were misperceived by many observers when viewing time was limited to 130 ms. Lowering the contrast to 39% greatly reduced performance for all observers, even for relatively long stimulus presentations (up to 0.5 s).

A similar kind of result was obtained by Yo and Wilson (1992) for the perception of "type II" plaids. In their experiments, the task of the observer was to integrate the motion of two superimposed drifting sinusoidal gratings. By definition, each component grating is a 1-dimensional motion stimulus containing directional information only along the axis perpendicular to the orientation of the grating's stripes (Fig. 1c). When two such gratings moving in different directions are superimposed, the resulting direction of the plaid motion can be computed in several different ways yielding different possible directions. For certain combinations—referred to as "type II" plaids—the simple vector sum[1] of the two 1D directions produces one predicted direction, whereas an algorithm sensitive to the direction of motion of the 2D features produced by the gratings' intersections, produces a different predicted direction (Fig. 1f). The main result of the Yo and Wilson study was that observers reported the vector sum direction for brief stimulus presentations (60 ms) but tended to see the feature direction as viewing time was increased, and, as in the tilted bar experiments, the time-course of the transition was prolonged with gratings of lower contrast.

Studies of motion integration using eye movements

For the perceived direction of tilted bars and type II plaids, the effect of viewing time clearly indicated that motion integration is a dynamic process. Early on, the visual system computes an estimate of motion direction, but it is biased by non-veridical measurements along contours. As time goes by, the correct solution emerges. However, psychophysical judgments are by nature discrete: the stimulus is observed for a fixed amount of time and a single response is given. That single response is presumably the outcome of a dynamic computation, during which various possible solutions are represented in the visual cortex. There is therefore no way to determine whether the observer's response reflects the integrated percept, the most recent mental snapshot or some other way of combining percepts over time. In this respect, psychophysics is not ideal for addressing the issue of temporal dynamics.

---

[1] Both the vector sum and the vector average produce resultant vectors with the same direction but different magnitudes. Because we are largely concerned with measurements of direction, we will use the vector sum to refer to both possibilities.

With respect to visual motion, however, one can monitor other outputs that arguably make use of the same motion processing circuitry: eye movements. In the case of certain eye movements, such as ocular following and smooth pursuit, we are afforded a continuous read-out of the process with no requirement for any kind of conscious judgment on the subject's part. This makes smooth eye movements ideal for studying the dynamics of motion integration, not only in humans but in any animal that can move its eyes. An additional benefit for the cortical physiologist is that both these types of eye movement have been tightly linked to neural signals in the middle temporal (MT or V5) and the medial superior temporal (MST) visual areas of macaque monkeys, (Newsome, Wurtz, Dursteler & Mikami, 1985; Groh, Born & Newsome, 1997; Kawano 1999; Born, Groh, Zhao & Lukasewycz, 2000) thus permitting a direct comparison with single unit studies.

A number of labs have availed themselves of eye movement recordings elicited by a variety of visual stimuli that contain 1D and 2D features moving in different directions to study the dynamics of motion integration. As many of these experiments will be described in greater detail elsewhere in this volume (Chapter 12 by Masson, Montagnini & Ilg), we will focus our discussion here to the essential common features and to the results pertaining to manipulations of contrast and contour length that parallel the psychophysical results described above.

Two different yet closely related smooth eye movements have been used to probe the 1D-to-2D progression. First, *ocular following* (OF) is a very short-latency, automatic eye movement in response to motion of all, or a large part of, the visual field (Miles & Kawano 1986; Kawano & Miles 1986; Miles, Kawano & Optican, 1986). The second type of eye movement is *smooth pursuit*, in which the subject actively tracks the motion of a single target, usually a small spot. Because of the different nature of the visual stimuli used to evoke these two eye movements—large textures for OF versus single objects for smooth pursuit—they have been suited to probing slightly different features of the computation of motion direction. In particular, OF responses have been evoked by barber poles (Masson, Rybarczyk, Castet & Mestre, 2000) and plaids (Masson & Castet 2002), both essentially textured stimuli that readily lend themselves to manipulations of stimulus contrast. Smooth pursuit has proven useful for studying the effects of contour length (Born, Pack, Ponce & Yi, 2006), using single tilted bars (Pack & Born 2001; Born et al. 2006) or objects composed of oblique lines, such as rhombi (Masson & Stone 2002). In the middle ground, where textured patches are reduced in extent and objects are made larger, the distinction between the two types of eye movement

becomes blurry and may either cease to exist or reflect a transition over time from OF to
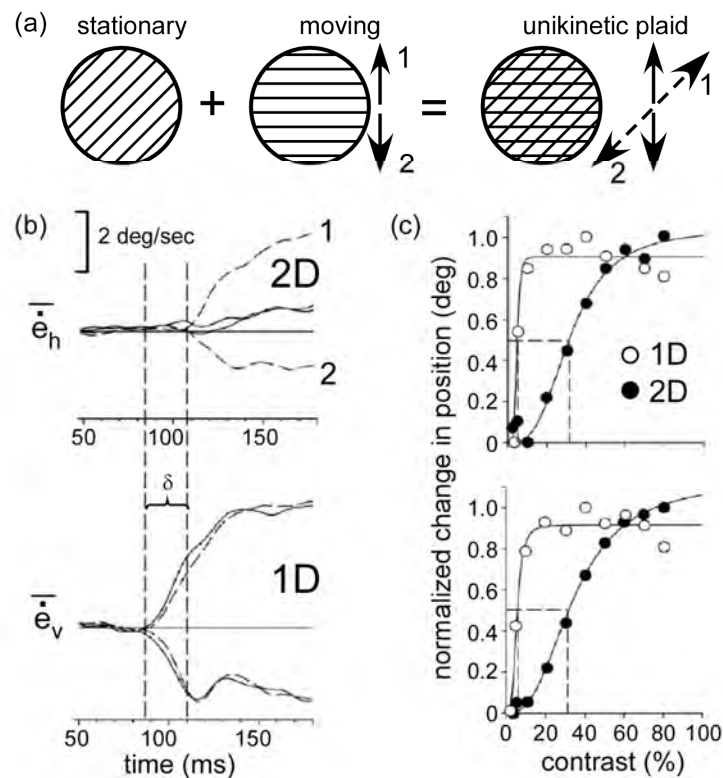pursuit (Masson & Stone 2002).

*Figure 2. Ocular following elicited by unikinetic plaids from a study by Masson & Castet (2002). (a)
The unikinetic plaid is composed of an oblique stationary grating superimposed upon a horizontal
grating which can move either up (1) or down (2). The resulting motion appears to move along an
oblique axis. (b) Eye movement responses elicited by either a single horizontal grating (solid lines) or
the unikinetic plaid (dashed lines). In both cases, the 1D motion perpendicular to the horizontal
grating is purely vertical and so is manifest as the vertical component of the eye velocity ($\dot{e}_v$). For eye
movements evoked by the unikinetic plaid (dashed lines), the 1D component seen in the vertical eye
velocity traces are identical to those evoked by the grating alone. The 2D component, seen in the
horizontal eye velocity traces ($\dot{e}_h$), has a longer latency ($\delta \approx 20$ ms). (c) Contrast sensitivity of the
early (1D) and late (2D) components of ocular following elicited by unikinetic plaids in two different
human observers. The 1D response shows a very steep dependency on contrast and early saturation,
similar to that for magnocellular neurons, whereas the 2D response is shallower and saturates at
higher contrasts, more characteristic of parvocellular neurons.*

In any case, the main results from all of the experiments mentioned above have an essential feature in common: the initial eye movement is in the direction of the 1D component of stimulus motion and only some tens of ms later does it reflect the 2D direction. This temporal evolution appears to reflect, at least in part, a difference in the time required to process the two different motion cues. For both types of eye movement, the latency of the 1D response leads that of the 2D response—compare the OF data in Fig. 2b with the pursuit data in Fig. 3a. Though differing in absolute latency, both show a common relative latency difference between the 1D and 2D responses of approximately 20 ms.

One potential caveat in interpreting the results as a true difference in processing speeds is that, in both cases, the purely 2D component of the eye movement is smaller—the direction difference between the 1D and 2D components of stimulus motion is only 45°, so even the 2D component, when broken down into eye movement coordinates that are either parallel or perpendicular to the 1D direction, has half of its amplitude represented on the 1D axis—and thus may only appear delayed because it takes longer to rise above the baseline plus noise. This explanation is made unlikely, however, by a number of stimulus manipulations in which the relative strengths of the 1D *vs.* 2D components have been varied. For example, Masson et al. (2000) either masked off the center of the barber pole (a relative decrease of the 1D motion strength) or blurred the edges of the barber pole aperture (a relative decrease of the 2D strength) and found that the respective amplitudes of the two components of the eye movement responses changed in the expected ways, but the latencies of the two remained the same and showed the characteristic 20 ms gap between the 1D and 2D responses (δ in Fig. 2b).

Similar latency differences were also found for unikinetic plaids (Fig. 1e), thus allowing Masson and Castet (2002) to independently measure the effects of varying contrast on the 1D (early) and 2D (late) eye movement responses. The fact that the two responses showed markedly different contrast response functions (Fig. 2c) can be taken as further evidence that the neural circuitry underlying the computation of 1D vs. 2D motion is at least partially distinct—either because completely different pathways are involved (Masson & Castet 2002) or, as we will argue below, because the feedback mechanisms responsible for surround suppression or "endstopping" have a lower contrast sensitivity than do the feedforward center mechanisms. This evidence also argues against the notion that a single mechanism tuned to 1D motion can account for the temporal dynamics of 1D-to-2D motion responses based on longer latencies for frequency components having lower amplitudes (Majaj et al. 2002) .

The other stimulus variable of interest with respect to the psychophysical results is that of contour length. This was explored in some detail by Born et al. (2006), whose results for smooth pursuit agreed, at least qualitatively, with those of Lorençeau and colleagues for perception. Specifically, when monkeys were required to track tilted bars of varying length, the 1D response—measured as the deviation in the direction of the eye movement from the true, 2D direction of bar motion—was both larger in amplitude and more prolonged in duration for longer bars (Fig. 3b-c). The angular deviation of pursuit over time was, in general, well described by a single exponential function. In addition to providing a single parameter, τ, that may be useful for comparisons with past and future experiments in perception and physiology, the single-exponential nature of the process may be a clue to the underlying mechanism. That is, the temporal dynamics observed in these experiments may reflect the gradual accumulation of evidence for a feature that is measured by a single pathway, rather than the integration of two independent pathways.
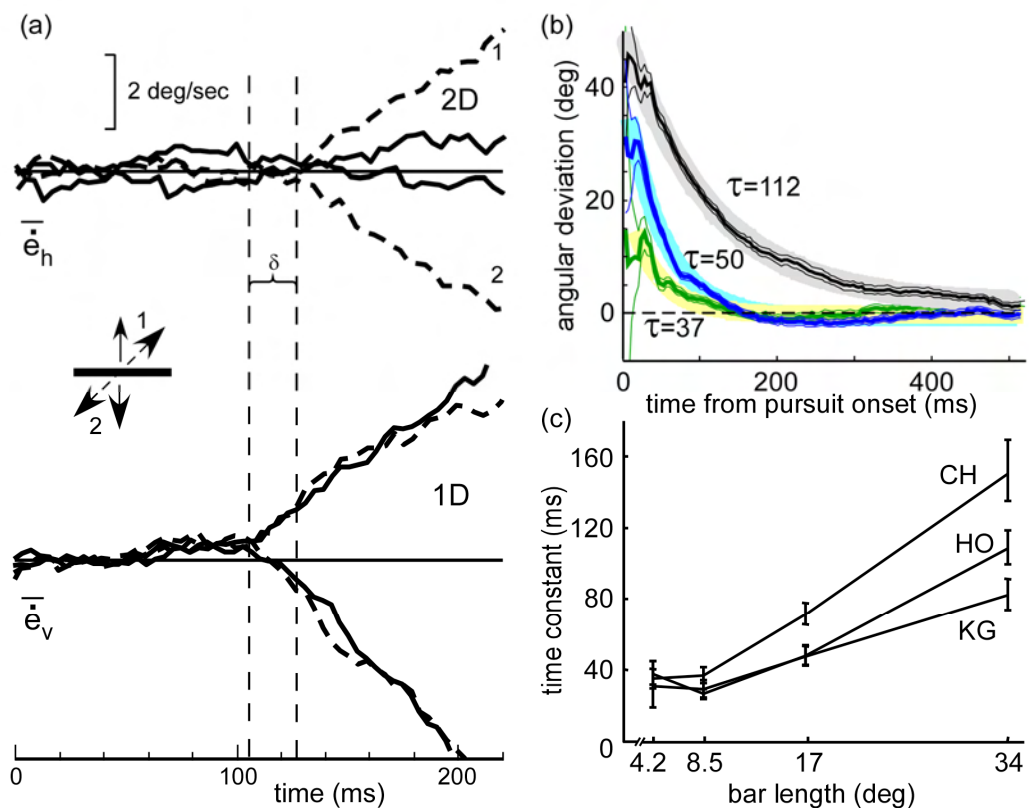

Figure 3. Bar pursuit data from our own lab (Born et al., 2006) displayed so as to be directly comparable to the ocular following experiments of Masson & Castet (2002). (a) Monkeys actively pursued a horizontal bar that moved either vertically (solid lines), up (1) or down (2), or obliquely (dashed lines), up and to the right (1) or down and to the left (2). In both cases, the eye velocity response to the 1D component is reflected in the vertical component of the eye velocity ($\dot{e}_v$). For the

obliquely moving bar, the responses to the 1D component seen in the vertical eye velocity traces are identical to those evoked by the vertically moving bar (equated for speed). The 2D responses, seen in the horizontal eye velocity traces ($\dot{e}_h$), show a longer latency ($\delta \approx 20$ ms). This data can be directly compared to that shown in Fig. 2b. (b) Time course of the angular deviation for tilted bars of different lengths (34°, 17° and 4°) along with the best-fitting single exponential decay function for monkey HO. Values for τ are in msec. (c) Time constant, τ, as a function of bar length for each of three different monkeys. Panels (b) and (c) were modified from Fig. 4 of Born et al. (2006).

To summarize up to this point, nearly a dozen different studies using both perceptual reports and smooth eye movements in response to a broad range of visual stimuli reveal a dynamic motion integration process. And, while stimulus differences preclude direct quantitative comparisons, the qualitative similarities—particularly with respect to the effects of stimulus contrast and contour length—are striking. On the whole they strongly indicate that all are tapping into common underlying neural circuits. Insofar as this is so, the preceding sections have provided a number of valuable clues as to the nature of this circuitry and provided important constraints for the computational models necessary to account for the relationship between neurons and behavior.

Neural correlates of motion integration

Having observed perceptual and behavioral signatures of a dynamic integration process in vision, one might wonder how these processes are represented in the brain. Although there are many brain areas devoted to visual motion processing, most studies to date have focused on the middle temporal area (MT) of the macaque monkey cortex. Neurons in MT are particularly sensitive to the velocity of visual stimuli and relatively insensitive to other stimulus attributes, such as color and texture. Furthermore, individual MT neurons integrate motion signals over larger regions of visual space than neurons in the primary visual cortex (V1), which provides the bulk of the input to MT. The question of whether the spatial integration observed at the level of single MT neurons displays the sort of temporal dynamics observed perceptually has been the subject of several neurophysiological studies in the last few years.

Pack and Born (Pack & Born 2001) recorded the responses of single MT neurons to stimuli that were similar to those used by Lorençeau et al. (1993) to study human perception. The main differences were that the bars were positioned to fill the receptive field of the neuron under study (Fig. 4a), and that they moved somewhat faster than those used in the psychophysical studies. Of course, the output of single neurons can be evaluated on a much finer time-scale than that of human perceptual reports, and the MT results revealed a dynamic

integration process that evolved over approximately 60 ms. In agreement with both behavioral and perceptual studies, the earliest responses reflected a non-specific integration that was heavily biased toward the component of motion perpendicular to the orientation of the bars (Figure 4b). In other words, the neurons were initially fooled by the aperture problem, and their responses did not begin to approximate the correct direction of motion until some time later (Fig. 4c,d; see also movies: *cu085c90.avi*, *cu085a45.avi*, *cu085b135.avi*). Interestingly, after the dynamics had stabilized, the MT population maintained an average error of roughly 5 degrees, indicating that the motion integration process is not quite perfect, even for stimulus durations of 2000 ms. A similar residual bias was observed psychophysically (for plaid stimuli) by Yo & Wilson (1992).
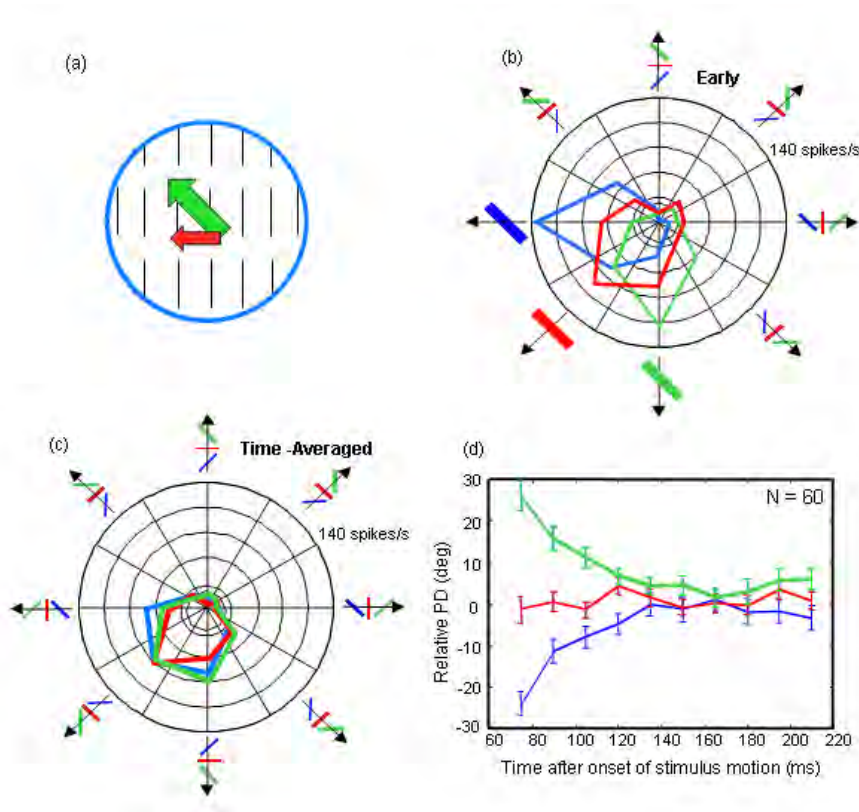
Figure 4. Response of MT neurons to the bar field stimulus. (a) The bar field stimulus consisted of rows of bars with the same orientation, filling the receptive field (blue circle). When the bars moved obliquely with respect to their orientation (green arrow), the component of motion perpendicular to the bar motion (red arrow) differed from the true direction by 45°. (b) 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. (c) The later part of the response depends only on the motion direction. (d) The transition from orientation-dependent responses to purely motion-dependent responses is evident in the population of 60 MT neurons. See also movies: cu085c90.avi, cu085a45.avi, cu085b135.avi, for a single-cell example of the temporal dynamics.

The temporal dynamics observed in MT bear at least a superficial similarity to integration processes observed in other aspects of vision. Various studies have found that the earliest responses of visual neurons encode a coarse description of the stimulus features, including orientation (Ringach, Hawken & Shapley, 1997), stereoscopic depth (Menz & Freeman 2004), faces (Sugase, Yamane, Ueno & Kawano, 1999; Tsao, Freiwald, Tootell & Livingstone, 2006), and complex shapes (Hegde & Van Essen 2004; Brincat & Connor 2006). In all cases later responses were linked to more specific details of the stimulus, as evidenced by a narrowing of tuning curves or a decorrelation of population activity in areas such as V1, V2, and IT.

The results on motion processing can be viewed in similar terms, since any resolution of the aperture problem requires the visual system to discern a specific velocity from a large number of candidate velocities. The range of possible velocities would be represented by neurons tuned to the various possibilities, and the neuronal signature of the solution would be a reduction of activity in most of these neurons (which would also entail a decorrelation of population activity). Although this refinement of the velocity representation was implicit in the results of Pack and Born (2001), their result did not elucidate either the underlying mechanism or precisely where in the brain the computation was taking place. That is, it remained possible that the solution to the aperture problem was occurring in some other part of the brain, and that MT was merely reporting the result. An obvious possibility was area V1, since it is the primary source of input to MT.

At a first glance, one might be inclined to rule out the possibility of a resolution of the aperture problem in area V1 a priori, based on the small sizes of the receptive fields found there. Indeed these receptive fields are in effect tiny apertures, and the geometry depicted in Figure 1 guarantees that V1 neurons will be incapable of signaling velocity accurately. Although this assumption is generally valid, there are important exceptions which turn out to be important for understanding the way the brain measures motion.

Consider the situation depicted in Fig. 1a. The aperture problem applies to all motion measurements made along the length of each line, but near the ends of the line, velocity can be measured accurately even by neurons with small receptive fields. The reason is that the line-endings or terminators are two-dimensional, and hence permit the measurement of both the perpendicular and parallel components of the bar's velocity. Of course the terminators comprise only a small fraction of each bar's area, and so the existence of these signals does not by itself lead to a solution to the aperture problem. Rather what is needed is a selective

integration process that ignores or suppresses the velocities measured along the length of each bar.

A hint at the mechanism by which this selective integration is accomplished was present in the pioneering work of (Hubel & Wiesel 1965). In V1 of the anesthetized cat, they discovered neurons that failed to respond to oriented stimuli that extended beyond their central activating regions. Hubel and Wiesel called these neurons hypercomplex, but this term was eventually replaced by a more descriptive term: "endstopped neurons". Many endstopped neurons responded well to terminators, suggesting that in combination with direction selectivity these cells could provide the kind of processing necessary to overcome the aperture problem (see movie: *HWendstop-short.avi*). This possibility was subsequently confirmed by Pack, Livingstone, Duffy and Born (2003), who showed that direction-selective endstopped neurons in the alert macaque were capable of encoding motion direction in a manner that was independent of stimulus orientation. Moreover, this invariance of direction selectivity emerged only after a brief delay – the initial response was influenced by the aperture problem in precisely the way that one would expect based on the geometrical analysis described above (see movie *endstop.avi*). Thus there is good evidence that the dynamic process of motion integration that is observed in MT begins and is partially completed at the level of V1. This possibility is further supported by the observation that the V1 layer that provides the bulk of the projection to MT is also the layer that has the highest prevalence of end-stopped neurons (Sceniak, Hawken & Shapley, 2001).
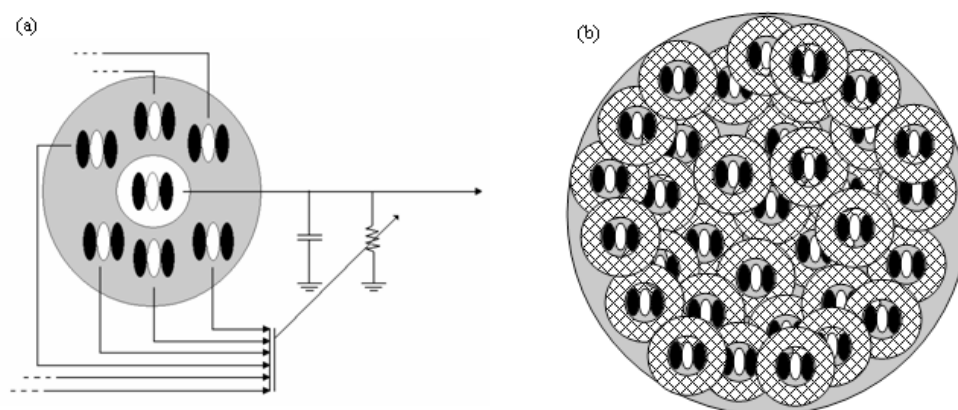

Figure 5. A simple model of motion integration in MT. (a) Circuit for a model endstopped cell, after Carandini et al. (1997). The model is composed of series of identical subunits, each of which is a motion energy detector. The output of the central subunit is modulated by the outputs of several surrounding subunits, and their interaction is characterized by an RC circuit with variable resistance.

(b) The model MT cell simply sums the outputs of endstopped cells like those in (a), but with receptive fields at different spatial positions.

A computational model of motion integration

End-stopping is a specific instance of a more general phenomenon known as surround suppression, which occurs whenever an increase in stimulus size leads to a reduction in neuronal response. Such an inhibitory influence may reflect neuronal mechanisms of normalization, which serve to calibrate the sensitivity of individual neurons to the overall amount of visual stimulation reaching a given visual area. Normalization models have been proposed in various contexts to account for certain nonlinear behaviors that have been observed in V1 neurons (e.g., Heeger 1992). While these models have been successful in explaining many of the V1 results, there has not to our knowledge been a computational account of the interaction of end-stopping and direction selectivity in V1.

A candidate for such a model is depicted in Fig. 5a. In this conception the direction selectivity of each cell is based on an implementation of the motion energy model (Adelson & Bergen 1985), with the parameters fully constrained by measurements from macaque V1 neurons (Pack, Conway, Born & Livingstone, 2006). The end-stopping in each neuron is due to inhibitory input from other V1 neurons with nearby receptive fields, based on the circuit model proposed by(Carandini et al. 1997). The model of Carandini et al. implements normalization by dividing the activity of each cell by the summed activity of its neighbors. When the receptive fields of the neighboring cells occupy different spatial positions, normalization translates into surround suppression, since large stimuli activate more of the surrounding cells than do small stimuli. Our model extends the proposal of Carandini and colleagues by incorporating a limited spatial range for the normalization pool and a realistic time constant for the surround interactions. Both of these extensions turn out to be important for the ability of the model to account for the dynamic aspects of motion processing.
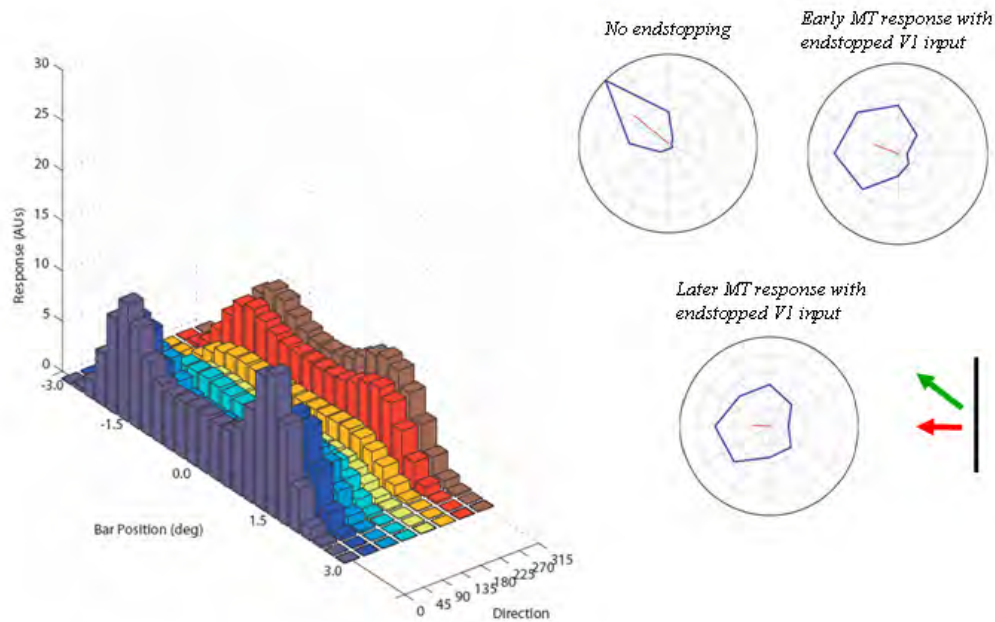
Figure 6. Model output. (a). Response of a model V1 cell to a bar moving through different positions and at different orientations. The model responds best to the endpoints of the bar, irrespective of bar orientation, which modulates the overall response level. (b). Output of the model MT cell in response to bars oriented 45° with respect to their direction of motion, when V1 endstopping is disabled. The cell normally prefers leftward motion, but in response to the tilted bar stimulus its tuning curve rotates by nearly 45°. (c) Early response of the MT neuron with endstopped V1 input. The tuning curve is rotated by more than 20° from its actual preferred direction. (d) Later response of the same model MT cell. The tuning curve is centered on leftward motion, with a small residual error of roughly 5°.

Figure 6a shows the response of the model to bars at different positions, and it is clear that the neuron responds primarily to the endpoints. This property is invariant with bar orientation, as was observed for real endstopped neurons (Orban, Kato & Bishop, 1979; Pack et al. 2003). We simulated a population of these endstopped neurons with receptive fields positioned at different points in visual space. The model MT neuron simply integrated the activity of the population of identical end-stopped V1 neurons (Fig. 5b). Fig. 6c,d shows the output of the MT neuron for the bar-field stimuli used in the Pack and Born (2001) experiment. In this simulation, the bars were oriented at an angle of 45° with respect to the direction of motion, and the model cell preferred leftward motion. The salient points of the MT data – the accurate measurement of motion and the associated temporal dynamics – are both captured in the model response (Fig. 6c,d). The residual bias due to the perpendicular component of the bar velocity is roughly 5°, which is consistent with that observed in MT of alert monkeys (Pack & Born, 2001). For comparison, Fig. 6b shows the large errors in the model output when end-stopping in V1 is disabled.

A recently published model (Rust, Mante, Simoncelli & Movshon, 2006) uses a similar normalization mechanism to account for the responses of MT neurons to plaid stimuli. In this model normalization plays a role similar to that played by endstopping in the model described above, namely to eliminate spurious motion signals that result from the aperture problem. However, this model lacks temporal dynamics, which would be necessary to provide a full account of the dynamic integration of plaid stimuli seen in MT (Pack, Berzovskii & Born, 2001; Smith, Majaj, Bair & Movshon, 2005). Furthermore, the normalization mechanism in the Rust et al. (2006) model lacks spatial structure. Indeed the model operates entirely in the velocity domain, so it cannot generate selective responses to endpoints or any other spatial feature of the stimulus.

The Rust et al. (2006) model is primarily an account of the continuum from "component" to "pattern" cells in MT, the latter being generally assumed to be the cells that solve the aperture problem. However, the model shown in Fig. 5 is capable of measuring the direction of the bar field stimulus, despite the fact that it would be classified as a "component" cell when tested with plaids. That is, it would respond to the motions of the gratings that make up the plaid, rather than the motion of the entire pattern (simulation results confirmed but not shown). To obtain "pattern" selectivity for plaids, Rust et al. (2006) required a second mechanism, namely a specific pattern of excitatory and inhibitory weights in the feedforward input from V1 to MT. The weighting for "pattern" cells in their model was broader in direction space than that for "component" cells, and it also had inhibitory lobes for directions far from the preferred direction.

Although we have not yet tested this idea, we suspect that the model of Rust et al. (2006) could account for the bulk of the existing data if it used an end-stopping mechanism similar to that shown in Fig. 5. This mechanism would provide the temporal dynamics observed in MT with bar fields (Pack & Born 2001), plaids (Pack et al. 2001; Smith et al. 2005), and barber poles (Pack, Gartland & Born, 2004). In combination with the feedforward weighting proposed by Rust et al. (2006), this mechanism would render most cells capable of measuring direction accurately for stimuli that contained terminators, but would limit the proportion of "pattern" cells to those that had broad direction tuning. This might also explain why the proportion of "pattern-like" cells in MT tends to decrease under general anesthesia (Pack et al. 2001), since the tuning bandwidth is also seen to be narrower in anesthetized animals (Pack, Berezovskii, and Born, unpublished observations).

Conclusions

In summary the existing physiological data demonstrate that MT neurons integrate the motion of complex stimuli gradually over a period of roughly 60 ms after the onset of stimulus motion. When the stimulus contains motion that can be measured accurately by end-stopped V1 neurons, the integration is nearly perfect for all MT neurons. For certain kinds of plaids, the distinction between "component" and "pattern" neurons reflects primarily the variability in tuning bandwidths instantiated by the projection from V1 to MT, with a possible contribution for inhibition from neurons with non-preferred directions (Rust et al. 2006). However, the distinction between "component" and "pattern" cells does not generalize well, as most "component" cells are perfectly capable of integrating pattern motion for other kinds of stimuli. Models that incorporate realistic estimates of the nonlinearities present in V1 are likely to provide a satisfactory account of these data, though a full model with all of these characteristic has yet to be implemented. In addition, a weighting of V1 inputs according to spatial and temporal frequency preferences may be helpful in measuring speed (Simoncelli & Heeger 1998). Such a mechanism could easily be incorporated into the framework outlined here.

Acknowledgements

Supplementary materials (CD-ROM)

Movie 1 *tilted.swf*

Movie 2 There are six shock-wave files that depict the stimuli diagrammed in figure 1. Their names are: *Figure1a.swf*, *Figure1b.swf*, *Figure1c.swf*, *Figure1d.swf*, *Figure1e.swf* and *Figure1f.swf*.

Movie 3-5 *cu085c90.avi, cu085a45.avi, cu085b135.avi*

Movie 4: *HWendstop-short.avi*

Movie 7: *endstop.avi*

References

Adelson E H & Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A* 2: 284-299.

Born RT, Groh JM, Zhao R & Lukasewycz SJ (2000) Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. *Neuron* 26:

725-734.

Born RT, Pack CC, Ponce CR & Yi S (2006) Temporal evolution of 2-dimensional direction signals used to guide eye movements. *Journal of Neurophysiology,* 95: 284-300.

Brincat SL & Connor CE (2006) Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron,* 49: 17-24.

Carandini M & Heeger DJ (1994) Summation and division by neurons in primate visual cortex. *Science,* 264: 1333-1336.

Groh JM, Born RT & Newsome WT (1997) How is a sensory map read Out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *The Journal of Neuroscience* 17: 4312-4330.

Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Visual Neuroscience 9:* 181-197.

Hegde J & Van Essen DC (2004) Temporal dynamics of shape analysis in macaque visual area V2. *Journal of Neurophysiology* 92: 3030-3042.

Hubel DH & Wiesel, TN (1965) Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology* 28:229-289.

Kawano K (1999) Ocular tracking: behavior and neurophysiology. *Current Opinion in Neurobiology* 9: 467-473.

Kawano K & Miles FA (1986) Short-latency ocular following responses of monkey. II. Dependence on a prior saccadic eye movement. *Journal of Neurophysiology* 56: 1355-1380.

Lorençeau J, Shiffrar M, Wells N & Castet E (1993) Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research* 33: 1207-1217.

Majaj N, Smith MA, Bair W & Movshon JA (2001) Temporal dynamics of neural responses to gratings and plaids in macaque area MT. *Perception* **30**: ECVP Abstract Suppl.

Majaj, N., Smith, M. A., Kohn, A., Bair, W., & Movshon, J. A. (2002). A role for terminators in motion processing by macaque MT neurons? [Abstract]. Journal of Vision, 2(7):415, 415a, http://journalofvision.org/2/7/415/, doi:10.1167/2.7.415.

Marr D & Poggio T (1976) Cooperative computation of stereo disparity. *Science,* 194: 283-287.

Marr D, Ullman S & Poggio T (1979) Bandpass channels, zero-crossings, and early visual information processing. *Journal of the Optical Society of America,* 69: 914-916.

Masson GS & Castet E (2002) Parallel motion processing for the initiation of short-latency ocular following in humans. *Journal of Neuroscience,* 22: 5149-5163.

Masson GS, Rybarczyk Y, Castet E & Mestre DR (2000) Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. *Visual*

Neuroscience, 17: 753-767.

Masson GS & Stone LS (2002) From following edges to pursuing objects. *Journal of Neurophysiology,* 88: 2869-2873.

Menz MD & Freeman RD (2003) Stereoscopic depth processing in the visual cortex: a coarse-to-fine mechanism. *Nature Neuroscience,* 6: 59-65.

Miles FA & Kawano K (1986) Short-latency ocular following responses of monkey. III. Plasticity. *Journal of Neurophysiology,* 56: 1381-1396.

Miles FA, Kawano K & Optican LM (1986) Short-latency ocular following responses of monkey. I. Dependence on temporospatial properties of visual input. *Journal of Neurophysiology,* 56: 1321-1354.

Newsome WT, Wurtz RH, Dursteler MR & Mikami A (1985) Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *The Journal of Neuroscience,* 5: 825-840.

Orban GA, Kato H & Bishop PO (1979) Dimensions and properties of end-zone inhibitory areas in receptive fields of hypercomplex cells in cat striate cortex. *Journal of Neurophysiology,* 42: 833-849.

Pack CC, Berezovskii VK & Born RT (2001) Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature,* 414: 905-908.

Pack CC & Born RT (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature,* 409: 1040-1042.

Pack CC & Born RT (2007) Cortical mechanisms for the integration of visual motion. in: *The Senses: A Comprehensive Reference*. Eds.Albright TD & Masland RH. Vol. 2. Amsterdam: Elsevier. pp. 192-218.

Pack CC, Conway BR, Born RT & Livingstone MS (2006) Spatiotemporal structure of nonlinear subunits in macaque visual cortex. *The Journal of Neuroscience,* 26: 893-907.

Pack CC, Gartland AJ & Born RT (2004) Integration of Contour and Terminator Signals in Visual Area MT of Alert Macaque. *The Journal of Neuroscience,* 24: 3268-3280.

Pack CC, Livingstone MS, Duffy KR & Born RT (2003) End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. *Neuron,* 39: 671-680.

Ringach, DL, Hawken MJ & Shapley R (1997) Dynamics of orientation tuning in macaque primary visual cortex. *Nature,* 387: 281-284.

Rohaly AM & Wilson HR (1993) Nature of coarse-to-fine constraints on binocular fusion. *Journal of the Optical Society of America A,* 10: 2433-2441.

Rohaly AM & Wilson HR (1994) Disparity averaging across spatial scales. *Vision Res*, 34: 1315-1325.

Rosch E, Mervis CB, Gray WD, Johnson DM & Boyes-Braem P (1976) Basic objects in

natural categories. *Cognitive Psychology,* 8: 382-439.

Rust NC, Mante V, Simoncelli EP & Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nature Neuroscience,* 9: 1421-1431.

Sceniak MP, Hawken MJ & Shapley R (2001) Visual spatial characterization of macaque V1 neurons. *Journal of Neurophysiology,* 85: 1873-1887.

Simoncelli EP & Heeger DJ (1998) A model of neuronal responses in visual area MT. *Vision Research,* 38: 743-761.

Smith MA, Majaj N & Movshon, JA (2005) Dynamics of motion signaling by neurons in macaque area MT. *Nat. Neurosci.* 8: 220-228.

Sugase Y, Yamane S, Ueno S & Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature,* 400: 869-873.

Thorpe SJ, & Fabre-Thorpe M (2001) Seeking Categories in the Brain. *Science,* 291: 260-263.

Tsao DY, Freiwal, WA, Tootell RB &Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science,* 311: 670-674.

Wilson HR, Blake R & Halpern DL (1991) Coarse spatial scales constrain the range of binocular fusion on fine scales. *Journal of the Optical Society of America A* 8: 229-236.

Yo C & Wilson HR (1992) Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research* 32: 135-147.