Mechanisms of stereoscopic vision: the disparity energy model

Izumi Ohzawa

The past year has seen significant advances in our understanding of the role played by the primary visual cortex (V1) in stereoscopic vision. Recently, the mechanism by which complex cells in V1 respond to random-dot stereograms has been characterized; it appears that their response properties greatly reduce the complexity of one of the critical links for stereopsis, the correspondence problem.

Addresses
School of Optometry, University of California at Berkeley, Berkeley, California 94720-2020, USA; e-mail: izumi@pinoko.berkeley.edu

Current Opinion in Neurobiology 1998, **8**:509–515

http://biomednet.com/elecref/0959438800800509

© Current Biology Publications ISSN 0959-4388

Abbreviations

RDS	random-dot stereogram
RF	receptive field
sd	standard deviation
V1	primary visual cortex
V2	secondary visual cortex

Introduction

The neural mechanism of stereoscopic vision has been one of the key areas of research in visual neuroscience since the pioneering study by Barlow *et al.* [1], who discovered neurons in cat striate cortex that are tuned to a wide variety of stereoscopic depths. They found that some of these neurons responded to a stimulus that is positioned further away from the plane of fixation, whereas the other neurons responded to a stimulus positioned on the fixation plane or closer. This result was later confirmed in the visual cortex of alert behaving monkey [2].

In a striking extension of this work, Poggio *et al.* [3] demonstrated that complex cells in monkey primary and secondary visual cortex (V1 and V2, respectively) respond to stereoscopic depth embedded in dynamic random-dot stereograms (RDSs). A RDS is a pair of images consisting of a two-dimensional array of randomly placed dots. (The dynamic version is a movie of RDSs.) The images in each pair are identical except that a portion of one image is displaced horizontally. Although the displaced region is invisible if each image is viewed individually, it becomes stunningly clear when the pair of images is viewed under a condition of binocular fusion. Because of this feature, the RDS has been used as an unambiguous test for stereopsis in humans and as a benchmark for computational models of stereovision. Although RDS tests are well established for human vision [4], it has been surprising to discover that neurons relatively early in the visual pathway (i.e. at only the second stage of visual cortical processing) respond to these patterns. Computational models of stereovision, such as those developed by Marr (see [5,6]), have demonstrated that detecting depth in RDSs requires highly complex computations, which seem difficult to accomplish with only two stages of cortical processing.

Models of stereoscopic vision developed over the past year have increased substantially our understanding of the functional neural wiring of V1 complex cells and their ability to respond to RDS patterns. We now also understand how these cells contribute to solving the correspondence problem (i.e. identifying the image feature from the left eye that corresponds to that from the right eye), and with which specific aspects of visual signaling these cells are involved. These issues form the primary focus of this review.

The correspondence problem

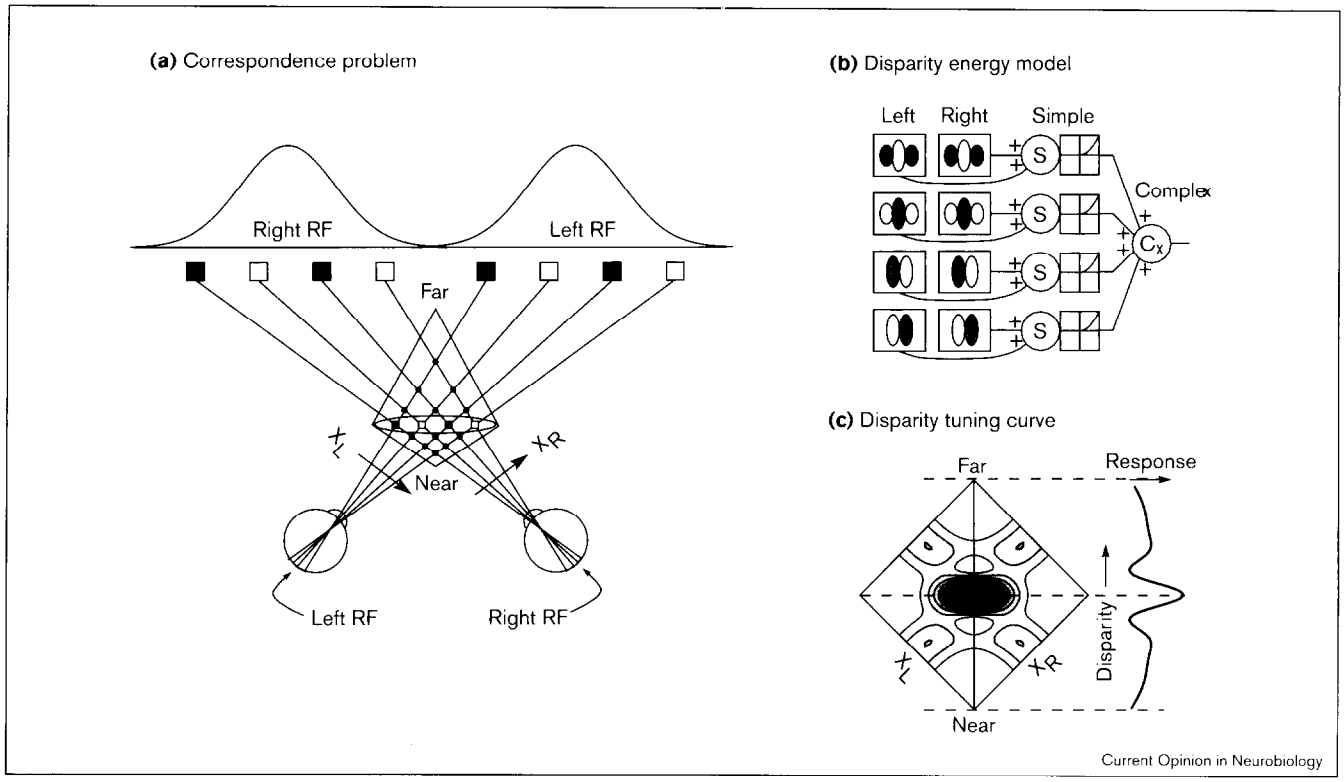
One of the major difficulties of stereoscopic vision relates to the correspondence problem, which is already acute in a complex natural environment but is even more pronounced for artificially constructed random-dot stimuli, which have no conspicuous monocular features to use as guides.

Using a diagram similar to the one depicted in Figure 1a, Julesz [4] described the complexity of the binocular matching process, which is used in stereoscopic vision. Even for the receptive field (RF) of a single complex cell, as illustrated at the top of Figure 1a, multiple image features are probably present within the RF. For a single row of actual targets, as shown enclosed in an elongated ellipse in the fronto-parallel plane, many other matches are possible at all intersections of the rays contained in the diamond-shaped region in Figure 1a. The possible matches that fall outside the ellipse are called false matches. Any stereoscopic vision mechanism must be able to resolve this ambiguity, eliminating false matches and coming up with a globally consistent solution. For the case shown in Figure 1a, an appropriate response for a cell may be to use a filtering operation that will make it respond only to targets that lie within the horizontal ellipse. A map of such selectivity to a stimulus placed at any point in the diamond-shaped area may be defined as the binocular RF of the cell. Being a two-dimensional map, a binocular RF is a more complete description of binocular properties than a disparity tuning curve, which most previous studies have relied upon [2,3,7–9].

The disparity energy model

We [10] and others [11–13] have devised and elaborated a model for a binocular complex cell, called the disparity energy model, which is shown schematically in Figure 1b. The model predicts well the shape of the binocular RF for complex cells in the cat [14•]. It has also been used

Figure 1



The correspondence problem and the disparity energy model. **(a)** Complex cells generally have large RFs that cover multiple targets in visual scenes. There are a large number of possible binocular matches (all intersections of rays), of which only those contained in the horizontal ellipse are correct. Those outside are called false matches. If a complex cell responds to any conjunction of left and right targets within its RF, it will not be able to discriminate correct matches from false targets. The problem is even more difficult when targets in the entire stimulus area must be considered, because the number of possible matches grows with the square of the number of targets in each eye's image. Redrawn from [4]. **(b)** The disparity energy model for complex cells, as described in [10,14•]. This model is based on highly specific hierarchical connections from selected simple cells to a complex cell (see text). **(c)** Predicted responses for a binocular RF are shown of the model in (b) to a target placed within the diamond-shaped area in (a) at a variety of positions. By integrating the profile along horizontal lines (parallel to dashed lines), the disparity tuning curve may be derived (right). X_L and X_R are the stimulus positions for the left and right eyes, respectively.

successfully to provide the initial binocular components for computational models that are capable of solving RDS tests [11–13,15,16•,17,18].

The disparity energy model consists of two stages. The first stage consists of an array of binocular simple cells with RFs for both eyes. The output of the simple cells goes through a half-squaring nonlinearity (i.e. rectification + squaring), and is summed by the complex cell stage. The predicted binocular RF for a complex cell is shown in Figure 1c, within a diamond-shaped domain similar to that of Figure 1a. There is a horizontally elongated area of high sensitivity, which is a desirable characteristic, as noted above. The traditional form of a disparity tuning curve may be derived from this binocular RF by integrating the RF along the horizontal lines (i.e. projections to the near-far axis) as shown at the right of Figure 1c.

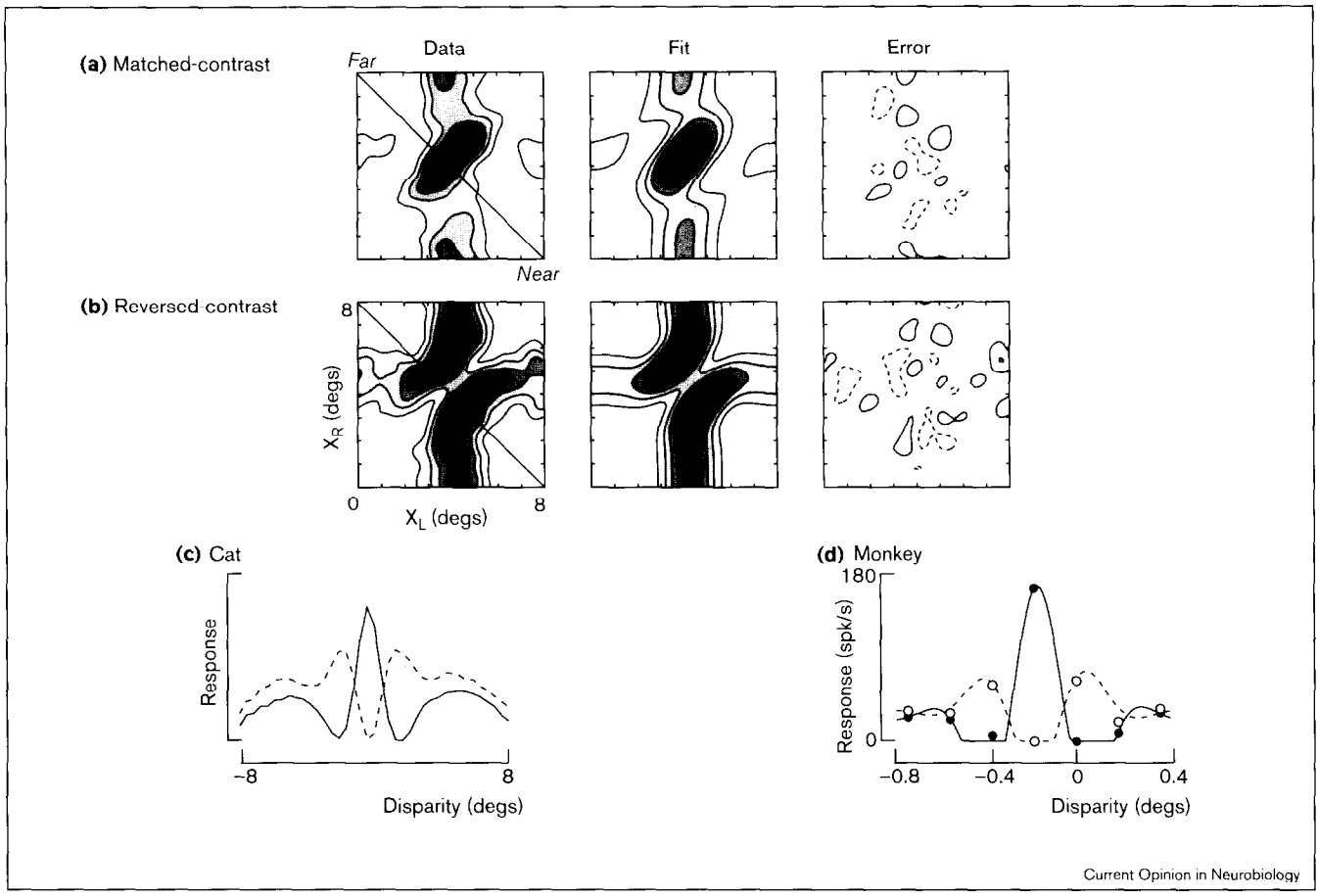
We [14•] have shown that the disparity energy model describes accurately binocular RFs of complex cells in cat striate cortex. Figure 2a depicts a representative binocular RF measured using a pair of dark bar-shaped stimuli and 'fitted' using the disparity energy model. The small amplitude and lack of structure in the residual error profile (panels on the right) show the goodness of fit. A reversal of contrast of the stimulus for one eye caused an inversion of the central portion of the binocular RF, as shown in Figure 2b (measured using a bright bar and dark bar for the left and right eyes, respectively). Again, the same disparity energy model predicts the cell responses accurately. The inversion is more clearly visible in the disparity tuning curves (Figure 2c) obtained by projecting the two-dimensional profiles from Figure 2a and 2b onto the near-far axis: the central excitatory peak for the matched-polarity condition (solid curve) becomes a trough when reversed-contrast stimuli are used (dashed curve).

Very similar results were obtained in a recent study by Cumming and Parker [19••] of V1 complex cells in the alert behaving monkey. Using dynamic RDS stimuli consisting of randomly placed bright and dark dots, they measured disparity tuning curves while the monkey maintained fixation. An example is depicted in Figure 2d (filled circles), to which a tuning curve

predicted by the disparity energy model has been fitted (solid curve). When bright and dark dots in the stimuli for one eye were reversed in polarity, the tuning curve inverted (dashed curve). Although the presence of a threshold for firing apparently clips away the negative portions of the curves, the inversion of the curve is clear. The authors emphasize that these responses to reversed-contrast stimuli do not in fact reflect 'conscious' perception of depth (see [20]), because there is no perception of depth with reversed-contrast RDS [21,22]. Therefore, they conclude that neurons in V1 do not solve the correspondence problem globally. The lack of a global solution at this stage is somewhat expected, given the generally accepted notion that most neurons in V1 operate as localized filters that see only a portion of the stimuli that happen to fall within their RF. It is important to note, however, that the lack of a global solution does not mean that V1 cells do not play a role in solving the correspondence problem (see below).

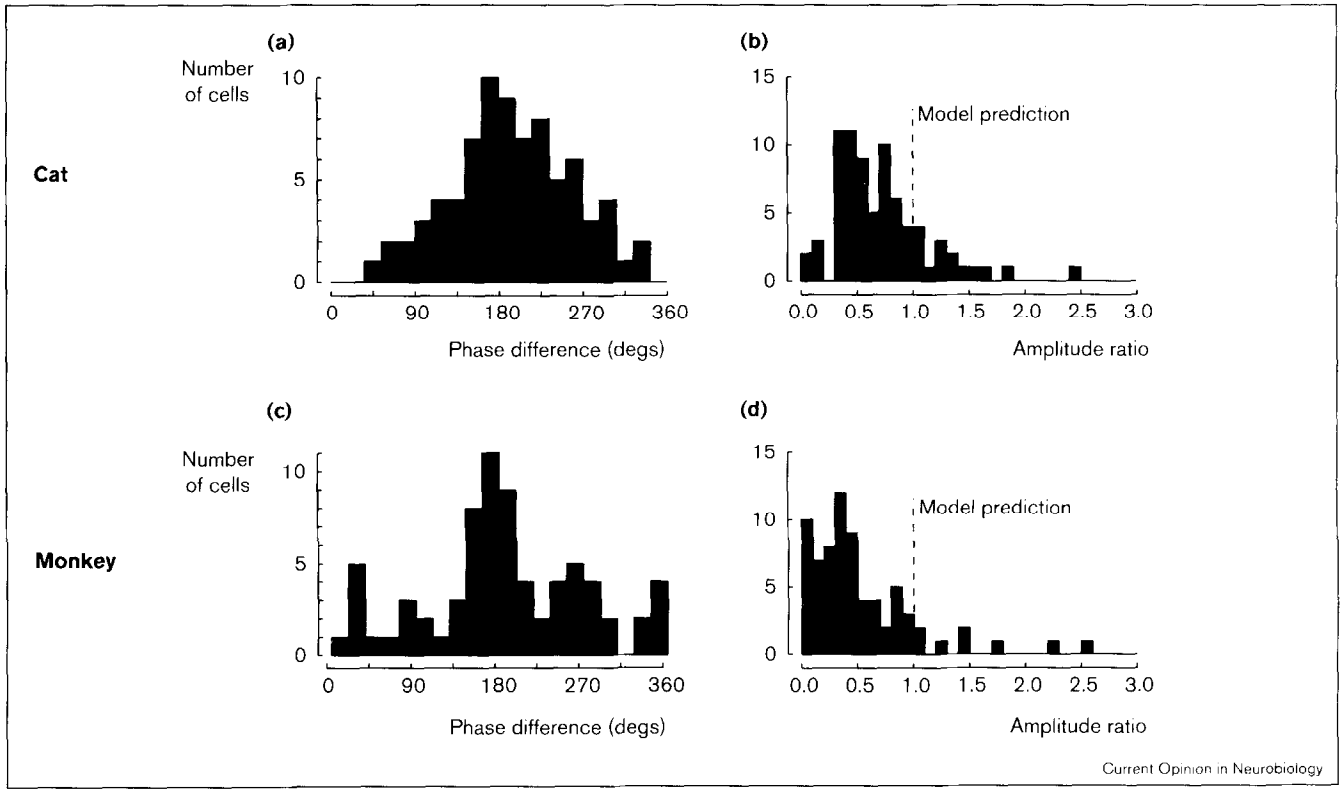
Perhaps a more important aspect of Cumming and Parker's [19••] results is that the disparity energy model predicts the responses of V1 complex cells to RDS almost completely. Not only are responses to same-contrast stereograms predicted well, but also those to reversed-contrast stereograms. Therefore, their results confirm the validity of the disparity energy model as an accurate description of the functional circuitry for V1 complex cells in the monkey. This is the first physiological confirmation that such a model can predict neural responses to dynamic RDSs, as originally proposed by Qian [11] in a computational study. The results from the physiological studies may be examined quantitatively by comparing the phase and amplitude of the disparity tuning curve for the

Figure 2



Responses of complex cells to matched-contrast and reversed-contrast stimuli. **(a)** Responses of a complex cell in cat V1 to matched-contrast stimuli. The stimulus was a dark bar (0.4 × 20 degrees) presented to both eyes at various combinations of locations. Each profile is defined by a grid consisting of 20 × 20 data points. A least-square fit to the data and residual errors are shown in the middle and right panels, respectively. Dashed contours indicate negative values. **(b)** Responses to reversed-contrast stimuli (bright and dark bars for left and right eyes, respectively). **(c)** Disparity tuning curves were derived from (a) and (b). Solid and dashed curves represent disparity tuning curves for matched-contrast and reversed-contrast stimuli, respectively. Data for (a–c) from [14•]. **(d)** For comparison, the disparity tuning curves from an alert monkey (redrawn from Cumming and Parker [19••]). Data are shown as filled (matched-contrast) and open (reversed-contrast) symbols. Curves represent fits based on the disparity energy model.

Figure 3



Summary of data and predictions of the disparity energy model. Data from **(a,b)** the cat (compiled from [14•]; n=40, of which 38 cells are represented twice for bright-dark and dark-bright conditions) and **(c,d)** the monkey (compiled from [19••]; n=72) are presented for comparison with the predictions of the disparity energy model. (a,c) Distributions of the phase of the disparity tuning curve for the reversed-contrast condition with respect to that for the matched-contrast condition for cat and monkey, respectively. The model predicts the phase difference to be 180°. (b,d) Distributions of the ratio of response amplitude for the reversed-contrast condition to that for the matched-contrast condition are shown. The model predicts a ratio of 1.0.

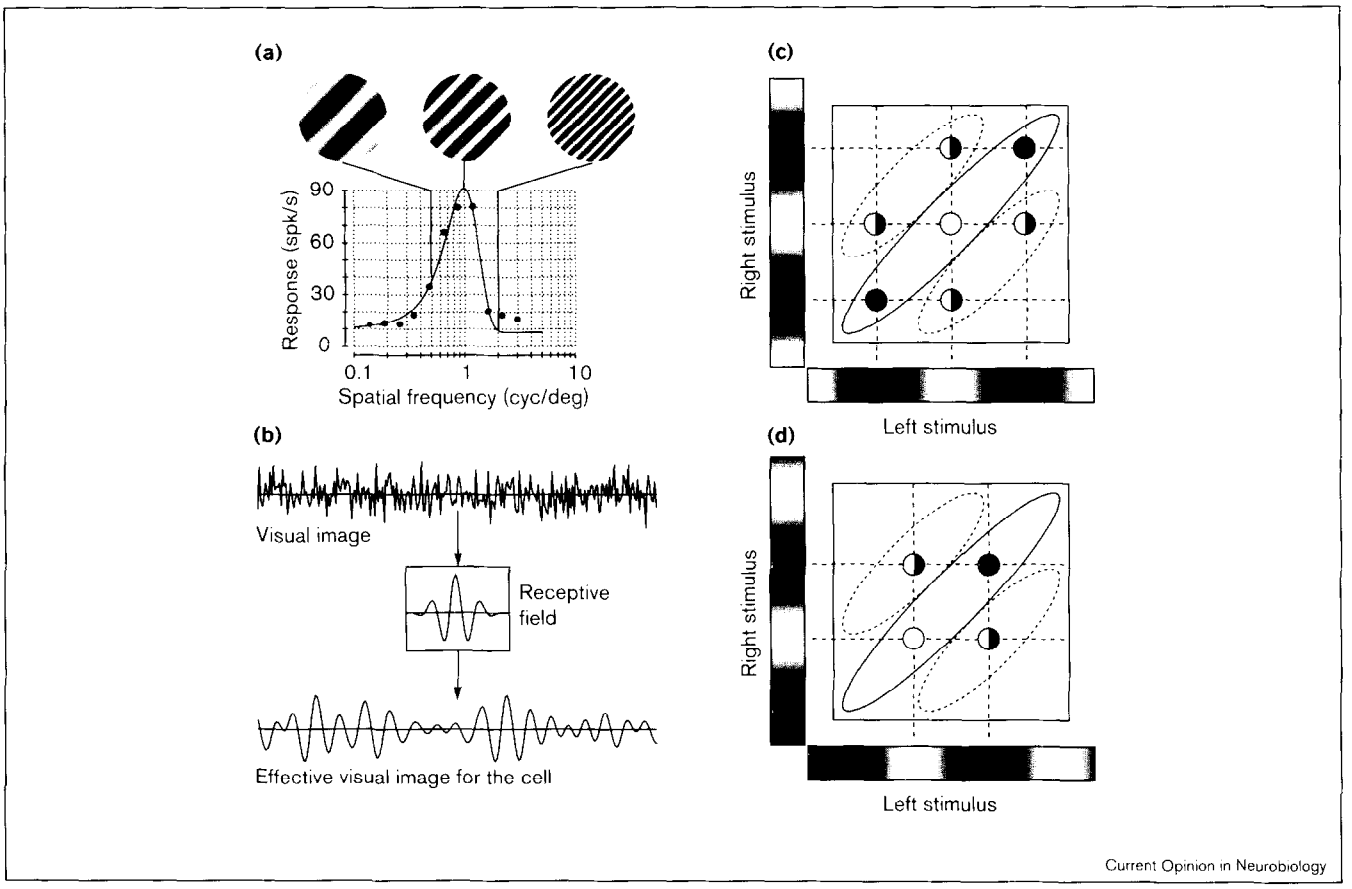
reversed-contrast case to those for the matched-contrast condition. The disparity energy model predicts that there will be a phase inversion (180° shift) while the amplitude remains the same (amplitude ratio = 1). Figure 3 summarizes the data from cat [14•] and monkey [19••]. The phase of the reversed-contrast disparity tuning curve is clearly clustered around 180°, as predicted by the model, for both cat (Figure 3a) and monkey (Figure 3c), although the distribution for the cat is somewhat broader. Interestingly, however, there is an unexpected systematic bias in the amplitude distributions. For the vast majority of complex cells, responses to reversed-contrast stimuli are substantially weaker than those to matched-contrast stimuli, for both cat (Figure 3b) and monkey (Figure 3d). This tendency appears to be more pronounced for the monkey (mean ratio = 0.52 ± 0.46 sd) than for the cat (mean ratio = 0.79 ± 0.60 sd). Some neurons responded very weakly or not at all to reversed-contrast stimuli. This is clearly a deviation from the prediction of the disparity energy model. However, it is also a deviation in the desired direction, in the sense that, ideally, there should be no response to reversed-contrast stimuli if these neurons support conscious perception of depth [19••,20]. The source of this deviation is not clear. It is possible that the bias that favors matched-contrast stimuli is based on feedback signals from higher-order areas that implement global stereo-matching. However, the fact that the bias is observed for cells in anesthetized, paralyzed cats suggests that such mechanisms may lie at a stage not involved in the conscious perception of stereoscopic stimuli.

It may be argued that the bias for matched-contrast stimuli may simply reflect preferences of cells to either bright or dark stimuli, as found for V1 complex cells [10,23] and in higher-order areas [24,25]. However, this possibility is unlikely because monocular biases in the responsiveness to bright and dark stimuli do not generally predict reduced responses to dichoptically reversed-contrast stimuli. In particular, even in cases in which both bright and dark stimuli elicited equally strong responses, reversed-contrast stimuli usually generated much weaker responses (see figure 6 in [14•]).

Responses to RDS stimuli

The results presented in Figure 2 suggest that individual neurons are unable to signal information as to whether they are excited by a stimulus of optimal disparity or by reversed-contrast stimuli at another disparity. At

Figure 4



Complex cells play a role in signalling interocular image registration. **(a)** A typical spatial frequency tuning curve for a V1 complex cell. **(b)** A consequence of such band-pass filtering by the RF is that any complex visual stimuli (depicted here by noise, top) is equivalent to a filtered version (bottom) that is highly periodic. **(c)** Equivalent periodic stimuli will produce combinations of appropriate contrasts, both matched and reversed, that are excitatory anywhere within the binocular RF. **(d)** This relationship is not sensitive to absolute monocular positions of stimuli.

first glance, such a response pattern would appear to be undesirable because it implies that the cells signal an ambiguous message. However, this is not the case. Consider another important property of these complex cells: each is tuned to a specific spatial frequency, as shown by the tuning curve in Figure 4a. Because of this band-pass characteristic, any complex image becomes equivalent to a filtered version of the image. This means that, for a given cell, the image is effectively highly periodic, having the main frequency of the RF profile. Binocularly, a combination of such periodic stimuli creates a pattern of matches that is perfectly optimal everywhere within the binocular RF, such as depicted in Figure 4c: the diagonally elongated region of excitation for matched-contrast stimuli (solid contour) receives appropriate stimuli (filled and open symbols); likewise, the two flanking regions excitatory to reversed-contrast stimuli (dashed contour) also receive optimal stimuli, a combination of bright and dark segments of stimuli (half-filled symbols). Therefore, the sensitivity to reversed-contrast stimuli of V1 complex cells is beneficial in that it enhances signals indicating binocular registration of left and right images at the preferred spatial frequency. Figure 4d shows that this registration process is not sensitive to the absolute position of bright and dark segments in each eye's view, as long as the binocular disparity is appropriate for the cell.

Comparisons of Figure 4c and 4d to Figure 1a indicate that a V1 complex cell solves the correspondence problem as best as it can within the localized image area that it sees through the RFs. It signals a registration of left and right images within a localized area. It has also been shown that the disparity energy model implements a computation mathematically equivalent to interocular cross-correlation [13], an operation highly suitable for registration tasks. Of course, being a localized detector, it cannot solve the global stereo correspondence problem alone, because of an aperture problem for stereopsis [26,27•]. However, by signaling patch-wise registration of images over a larger area containing multiple dots and lines, a V1 complex cell substantially reduces the complexity of the correspondence problem [4–6].

Conclusions

In the past year, there have been significant advances in our understanding of the roles of V1 complex cells for stereoscopic vision. They are detectors for interocular image registration that operate within a limited RF area. In this context, it makes perfect sense that vergence eye movements — the purpose of which is to bring left and right images into a best possible registration — are generated by both matched-contrast and reversed-contrast stimuli, as if they are driven by these complex cells [28••]. However, for stereoscopic vision, image registration is only the beginning. Further processing must be performed to extract borders defined by depth differences, which is necessary for detecting a shape embedded in a RDS. This process does not take place in V1. One way in which such a 'cyclopean edge detector' may be constructed is by an antagonistic (subtractive) convergence of output of V1 complex cells tuned to different disparities, just as simple cell RFs (once thought of as luminance edge detectors) are composed of antagonistic subregions.

Acknowledgements

The author is grateful to Keiji Tanaka, Ralph D Freeman, Aki Anzai, and Mike Menz for reading the manuscript and providing valuable comments, and to Aki Anzai for suggestions on interpretations of the reversed-contrast responses. Preparation of this manuscript was supported by a research grant from the National Institutes of Health (EY01175) to Ralph D Freeman.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Barlow HB, Blakemore C, Pettigrew JD: **The neural mechanism of binocular depth discrimination.** *J Physiol (Lond)* 1967, **193**:327-342.

2. Poggio GF, Fischer B: **Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving Rhesus monkey.** *J Neurophysiol* 1977, **40**:1392-1405.

3. Poggio GF, Motter BC, Squatrito S, Trotter Y: **Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms.** *Vision Res* 1985, **25**:397-406.

4. Julesz B: *Foundations of Cyclopean Perception.* Chicago: University of Chicago Press; 1971.

5. Marr D, Poggio T: **A computational theory of human stereo vision.** *Proc R Soc Lond [Biol]* 1979, **204**:301-328.

6. Marr D: *Vision.* New York: WH Freeman; 1982.

7. Ferster D: **A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex.** *J Physiol (Lond)* 1981, **311**:623-655.

8. LeVay S, Voigt T: **Ocular dominance and disparity coding in cat visual cortex.** *Vis Neurosci* 1988, **1**:395-414.

9. Poggio GF, Gonzalez F, Krause F: **Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity.** *J Neurosci* 1988, **8**:4531-4550.

10. Ohzawa I, DeAngelis GC, Freeman RD: **Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors.** *Science* 1990, **249**:1037-1041.

11. Qian N: **Computing stereo disparity and motion with known binocular cell properties.** *Neural Computation* 1994, **6**:390-404.

12. Zhu YD, Qian N: **Binocular receptive field models, disparity tuning, and characteristic disparity.** *Neural Computation* 1996, **8**:1611-1641.

13. Fleet DJ, Wagner H, Heeger DJ: **Encoding of binocular disparity: energy models, position shifts and phase shifts.** *Vision Res* 1996, **36**:1839-1857.

14. • Ohzawa I, DeAngelis GC, Freeman RD: **Encoding of binocular disparity by complex cells in the cat's visual cortex.** *J Neurophysiol* 1997, **77**:2879-2909.
Extensive analyses of binocular RFs of V1 complex cells and examinations of the validity of the disparity energy model are presented. It describes a highly specific neural wiring scheme by which output of multiple simple cells are combined to produce an efficient sensor for binocular disparity. Comparisons of the physiological data and predictions of the model show generally excellent agreement.

15. Qian N, Zhu Y: **Physiological computation of binocular disparity.** *Vision Res* 1997, **37**:1811-1827.

16. • Qian N, Andersen RA: **A physiological model for motion-stereo integration and a unified explanation of the Pulfrich-like phenomena.** *Vision Res* 1997, **37**:1683-1698.
This theoretical study unifies the models for V1 complex cells in the joint domain of motion and stereoscopic perception. The authors view complex cells as sensors for both motion energy [29,30] and disparity energy [10–13,14•]. Physiological evidence in support of the joint model has begun to emerge (A Anzai, I Ohzawa, RD Freeman, *Soc Neurosci Abstr* 1997, 23:567).

17. McLoughlin NP, Grossberg S: **Cortical computation of stereo disparity.** *Vision Res* 1998, **38**:91-99.

18. Gray MS, Pouget A, Zemel RS, Nowlan SJ, Sejnowski TJ: **Reliable disparity estimation through selective integration.** *Visual Neurosci* 1998, **15**:511-528.

19. •• Cumming BG, Parker AJ: **Responses of primary visual cortical neurons to binocular disparity without depth perception.** *Nature* 1997, **389**:280-283.
This is the first demonstration that the disparity energy model predicts well the responses of monkey V1 cells to dynamic random-dot stereograms (RDSs). Qian and colleagues [11,12,15] were the first to predict this outcome. An emphasis is placed on conscious perception of depth, motivated by a recent surge in interest in our awareness of neural activities [20]. This paper shows that V1 complex cells do not support the conscious perception of depth, as they respond to reversed-contrast RDS for which we cannot perceive any depth [21].

20. Crick F, Koch C: **Are we aware of neural activity in primary visual cortex?** *Nature* 1995, **375**:121-123.

21. Cogan AI, Lomakin AJ, Rossi AF: **Depth in anticorrelated stereograms: effects of spatial density and interocular delay.** *Vision Res* 1993, **33**:1959-1975.

22. Cogan AI, Kontsevich LL, Lomakin AJ, Halpern DL, Blake R: **Binocular disparity processing with opposite-contrast stimuli.** *Perception* 1995, **24**:33-47.

23. Hubel DH, Wiesel TN: **Receptive fields and functional architecture of monkey striate cortex.** *J Physiol (Lond)* 1968, **195**:215-243.

24. Tanaka K, Saito H, Fukada Y, Moriya M: **Coding visual images of objects in the inferotemporal cortex of the macaque monkey.** *J Neurophysiol* 1991, **66**:170-189.

25. Ito M, Fujita I, Tamura H, Tanaka K: **Processing of contrast polarity of visual images in inferotemporal cortex of the macaque monkey.** *Cereb Cortex* 1994, **4**:499-508.

26. Adelson EH, Movshon JA: **Phenomenal coherence of moving visual patterns.** *Nature* 1982, **300**:523-525.

27. • Morgan MJ, Castet E: **The aperture problem in stereopsis.** *Vision Res* 1997, **37**:2737-2744.
This work represents the first detailed psychophysical examination of an aperture problem for stereopsis. The aperture problem refers to an ambiguity that results from viewing images through an aperture, and has been studied extensively in relation to motion perception. For example, one cannot determine the true direction of motion of a grating pattern if viewed through an aperture (e.g. see Figure 4a, top). An RF of a visual neuron is an aperture. Therefore, if the RF truly has no sensitivity to stimuli outside the RF, it suffers from an aperture problem. In general, however, the visual system is able to solve the aperture problem because it is able to utilize responses of many neurons.

28. •• Masson GS, Busettini C, Miles FA: **Vergence eye movements in response to binocular disparity without depth perception.** *Nature* 1997, **389**:283-286.

Although we do not perceive depth from reversed-contrast stereograms, the vergence eye movement system, remarkably, responds to such stimuli. Short latencies in the response suggest that V1 complex cells may be an important source of the signal for this vergence response.

29. Adelson EH, Bergen JR: **Spatiotemporal energy models for the perception of motion.** *J Opt Soc Am [A]* 1985, **2**:284-299.

30. Emerson RC, Bergen JR, Adelson EH: **Directionally selective complex cells and the computation of motion energy in cat visual cortex.** *Vision Res* 1992, **32**:203-218.