Report

Perceived Direction of Motion Determined by Adaptation to Static Binocular Images

Keith A. May,[1,3,*] Li Zhaoping,[1] and Paul B. Hibbard[2]
[1]UCL Department of Computer Science, University College London, London WC1E 6BT, UK
[2]School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JP, UK

Summary

In Li and Atick's [1, 2] theory of efficient stereo coding, the two eyes' signals are transformed into uncorrelated binocular summation and difference signals, and gain control is applied to the summation and differencing channels to optimize their sensitivities. In natural vision, the optimal channel sensitivities vary from moment to moment, depending on the strengths of the summation and difference signals; these channels should therefore be separately adaptable, whereby a channel's sensitivity is reduced following overexposure to adaptation stimuli that selectively stimulate that channel. This predicts a remarkable effect of binocular adaptation on perceived direction of a dichoptic motion stimulus [3]. For this stimulus, the summation and difference signals move in opposite directions, so perceived motion direction (upward or downward) should depend on which of the two binocular channels is most strongly adapted, even if the adaptation stimuli are completely static. We confirmed this prediction: a single static dichoptic adaptation stimulus presented for less than 1 s can control perceived direction of a subsequently presented dichoptic motion stimulus. This is not predicted by any current model of motion perception and suggests that the visual cortex quickly adapts to the prevailing binocular image statistics to maximize information-coding efficiency.

Results and Discussion

We ran three experiments to investigate the effect of binocular adaptation on the perceived direction of a dichoptic motion test stimulus based on that of Shadlen and Carney [3]. In their stimulus, each eye received a flickering horizontal grating pattern (with maximum contrast α) that had no net motion in any direction, but the summation (S_+) and difference (S_-) signals were smoothly drifting gratings that moved in opposite directions. Shadlen and Carney reported that observers generally perceived motion in the S_+ direction [3], so, to remove this bias, we boosted the S_- signal by adding drifting gratings of opposite contrast, $\pm\beta$, in each eye. To selectively adapt the summation or differencing channels (which are sketched out in Figure 1) prior to presentation of the test stimulus, we presented static binocular stimuli that were either identical or contrast reversed between the eyes. The procedure for all experiments is illustrated in Figure 2 (see Supplemental Experimental Procedures available online for details).

In experiment 1, the binocular adaptors were pairs of natural images selected from a previously described set [4] and processed to avoid luminance artifacts. There were four adaptation conditions, involving different interocular correlations: correlated+, correlated−, anticorrelated, and uncorrelated (described in Figure 3). The results are shown in Figure 4A. The stimulus contrasts, α and β, were chosen so that, for uncorrelated adaptation, observers were equally likely to report motion in the S_+ or S_- direction, so performance was close to the 50% level on this condition by design. For correlated adaptation, which would selectively adapt (i.e., desensitize) the summation channel, the stimulus was most often seen to move in the S_- direction, whereas for anticorrelated adaptation, which would selectively desensitize the differencing channel, the stimulus was most often seen to move in the S_+ direction.

Most observers found the perceived motion to be subjectively weak, so performance was generally away from ceiling. This is not surprising, because the summation and difference signals provide conflicting information about stimulus direction, so neither inference (motion upward or downward) would be well supported by the pattern of neural activity in the visual cortex. Observers were faced with the task of selecting the more likely of two weakly supported hypotheses under a forced choice. The interobserver differences in effect size may have resulted from differences in neural noise, or "calculation efficiency" [5, 6], during the decision process under conflict. For our purpose (which does not extend to explaining the strength of the motion percept), we only need to assume that the preferred choice (S_+ direction or S_- direction) is influenced by the relative sensitivities to the S_+ and S_- signals.

Experiment 2 examined the orientation selectivity of the adaptation mechanisms. It was the same as experiment 1, except that the adaptors were 1D vertical or horizontal noise. The test stimulus was always horizontal. Figure 4B shows that, for two of the four observers (AB and KAM), the size of the adaptation effect was barely affected by the orientation of the adaptor, suggesting that the adaptation for these observers depended mainly on non-orientation-selective neurons in the visual cortex. The summation channel would be implemented by cells with identical isotropic receptive fields in the two eyes, and the differencing channel would be implemented by cells with isotropic receptive fields in which the polarity in one eye was opposite to that in the other. A few examples of the latter type of cell have been reported [7, 8]; as required by Li and Atick's theory [1, 2], Livingstone and Hubel found that "the responses from the two eyes, although opposite in sign, were algebraically additive" ([7], p. 324). They remarked that they were "at a loss to imagine any plausible benefit these collector's items could offer the animal" ([7], p. 324). We suggest that these cells form part of the implementation of our proposed differencing channel.

The other two observers in experiment 2 showed either a weaker (PS) or nonexistent (KMZ) effect of interocular correlation for the vertical adaptor. We used a chi-square test to see whether interocular correlation had had a significant effect for each observer and adaptor orientation (see Table 1). Although KMZ showed an overall bias to perceive motion in

[3]Present address: Department of Optometry and Vision Science, City University London, London EC1V 0HB, UK
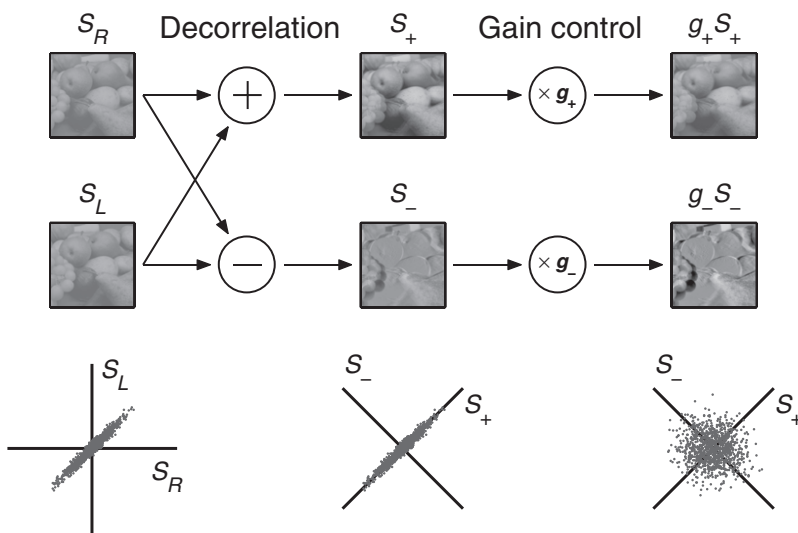*Correspondence: keith@keithmay.org

Figure 1. Li and Atick's Theory of Efficient Stereo Coding

If the visual scene contains objects at various distances from the observer, including some at very short distances, the two retinal images will be quite different from each other. On the other hand, if the visual scene consists entirely of faraway objects (such as a distant landscape), the two retinal images will be highly correlated and form a redundant, inefficient, representation. Li and Atick [1] showed that this inefficient representation could be transformed into an efficient one in two conceptual stages. First, the signals from the left and right eyes, S_L and S_R, are transformed into two uncorrelated signals: the sum, S_+, and the difference, S_-. Second, gains applied to the summation and differencing channels (g_+ and g_-, respectively) are dynamically optimized to the prevailing interocular correlation. As long as the signal-to-noise ratio is not too low, this optimization amounts to equalization of the channel amplitudes (i.e., whitening) [30, 31]. This maximizes information-coding efficiency because information is a compressive function of signal-to-noise ratio [32], which gives a diminishing return of information as signal strength increases; so, for a given energy budget, information capacity is maximized by attenuating strong signals and boosting weak ones until all channels have the same signal amplitude. When the images in the two eyes are weakly correlated, the S_+ and S_- signals will have similar amplitudes and the optimal gains will be similar; when the images in the two eyes are strongly correlated, the S_+ signal will be stronger than S_-, so the gain in the summation channel should be lower. The bottom row of this figure represents the channels as Cartesian axes, with scatter plots representing hypothetical distributions of responses in the two channels.

the S_+ direction with vertical adaptors, this bias was not significantly affected by the interocular correlation. This suggests that, for KMZ, adaptation to interocular correlation is orientation selective and our effect depends on neurons tuned to the same orientation as the test stimulus. KMZ's overall S_+ bias for vertical adaptors is consistent with this account, as we explain later, in our discussion of experiment 3.

Experiment 3 investigated the time course of adaptation. The initial adaptation period was identical to that in the uncorrelated adaptation condition of experiment 1, and the test stimulus was also the same. Unlike experiment 1, the top-up adaptation switched between correlated+ and anticorrelated on alternate trials within a session. The top-up adaptation stimuli were the same processed natural image pairs as in experiment 1, but the duration varied between sessions, and only one top-up adaptation stimulus was used on each trial.

The results of experiment 3 are shown in Figure 4C. The data for correlated+ and anticorrelated adaptation clearly diverged
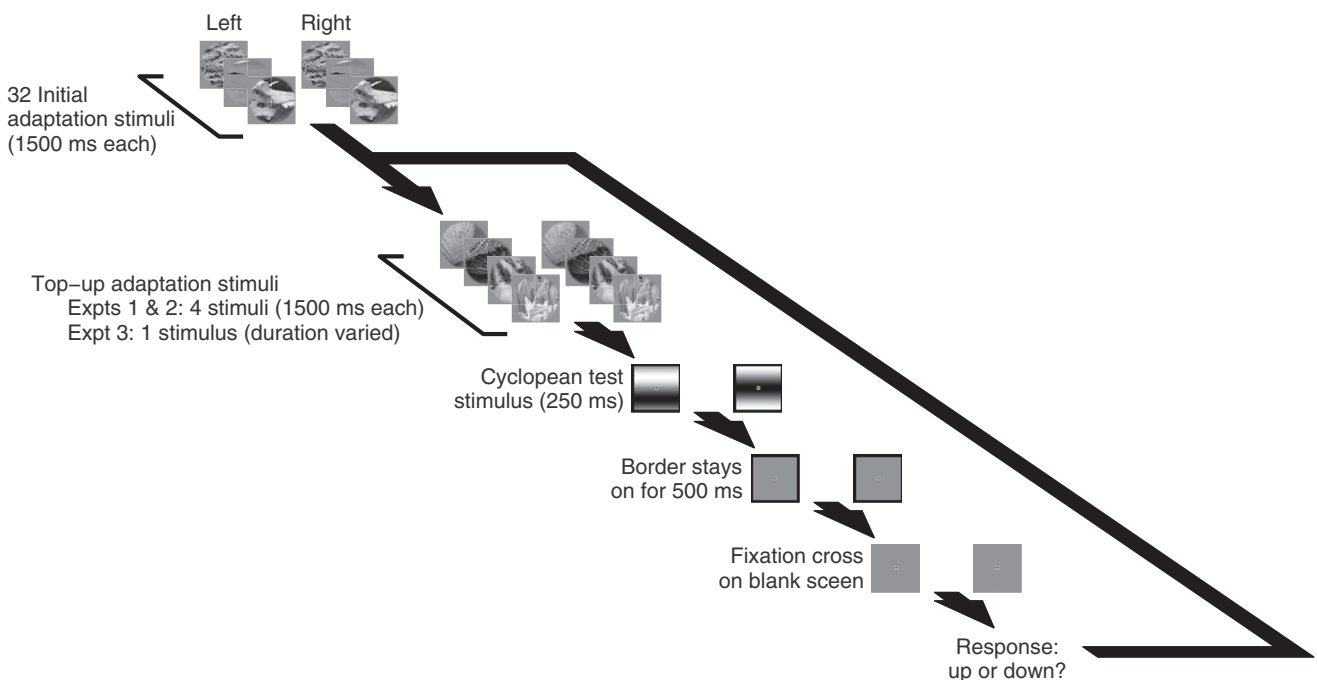

Figure 2. Procedure
The test stimulus was 4° visual angle wide, with a square black border. Experiments 1 and 2 had four top-up adaptation stimuli before each test stimulus; experiment 3 had only one. Only natural image adaptors (experiments 1 and 3) are shown here (actual diameter 12.8° visual angle); experiment 2 used 1D noise adaptors of the same size and shape as the test stimuli. An opaque fixation cross remained in view throughout each session.
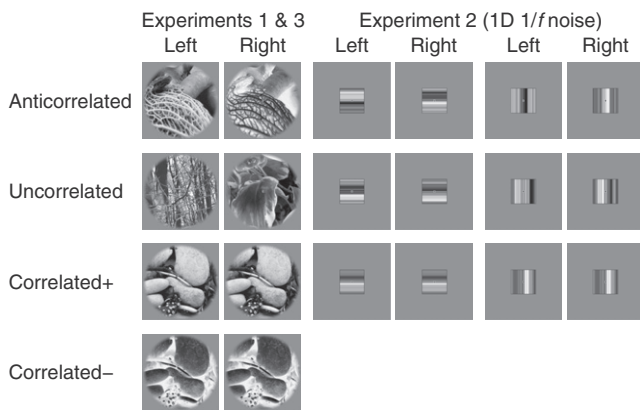
Figure 3. Adaptation Stimuli

Examples of the different adaptation types. Each row shows binocular stimulus pairs for a different adaptation condition. Adaptation stimuli for all experiments are shown to the same scale. "Left" and "Right" headings indicate left and right eyes. Experiment 1 had four types of adaptation, all using processed natural images. With anticorrelated adaptation, each eye saw the photographic negative of the other eye's image; these stimuli would cancel each other out in the summation channel, so only the differencing channel was stimulated. With uncorrelated adaptation, each eye received a completely different, randomly selected, positive image; this condition would stimulate both channels equally, on average. Finally, there were two correlated adaptation conditions: correlated+, in which each eye saw the same positive image, and correlated−, in which each eye saw the same negative image. The correlated conditions would stimulate only the summation channel. In experiment 2, the adaptors were 1D 1/f noise patterns, either identical in each eye (correlated), contrast reversed between the eyes (anticorrelated), or a different, random pattern in each eye (uncorrelated).

for top-up durations greater than around 0.25–1 s, depending on the observer. As in experiment 1, the test stimulus had an added drifting grating that was calibrated to give unbiased performance for uncorrelated adaptation in experiment 1. In experiment 3, the average correlation across all types of adaptation within every session was zero, like the uncorrelated condition of experiment 1, yet there was an overall bias toward S_+, especially for short or zero top-up durations. This probably reflects adaptation to the brief test stimulus, with the differencing channel adapting more quickly than the summation channel (a likely possibility given that, in natural vision, the strength of the S_- signal fluctuates more rapidly than S_+); the brief test stimulus preceded and followed by a blank screen would then have little effect on adaptation of the summation channel, but the differencing channel might adapt quite strongly within the first few milliseconds of the stimulus, allowing the summation channel to dominate perception of it. This would explain the perceptual bias toward the S_+ direction in Shadlen and Carney's original experiment. It also explains the S_+ bias with vertical adaptors in KMZ's data in experiment 2: our analysis of that experiment suggested that the effects of adaptation were mediated largely by horizontally tuned cells for KMZ, so the vertical adaptation condition would be similar to having no adaptation at all, which should give a bias in the S_+ direction.

Our results follow naturally from the proposal of independently adaptable binocular summation and differencing channels that are central to Li and Atick's [1, 2] theory of efficient stereo coding. We know of only two other papers to have argued for the existence of binocular differencing channels [9, 10], and this feature is missing from most contemporary

models of binocular combination [11–16], which sum the ocular signals. In many of these models, binocular summation is preceded by divisive suppression of each eye by the other [12–16], and one might wonder whether this interocular suppression mechanism, rather than a binocular differencing channel, could explain our results. In fact, no model of this kind can account for our results, because the manipulations in our study would have the same effect on each eye, and this symmetry would ensure that any change in interocular suppression before summation due to adaptation would be identical in the two eyes; the effect would be an overall change in effective stimulus contrast, but the effective contrast would not differ between the eyes, and the two eyes' signals would still sum together to produce motion in the S_+ direction, as in Equation 6 of the Experimental Procedures section below. The reversal of perceived direction requires a change of sign of one eye's signal (Equation 7), implying the existence of a binocular differencing channel. However, our conclusions are not fundamentally incompatible with current models of binocular combination [11–16], and we suggest that these models could usefully be extended by including both a binocular differencing channel and gain-control mechanisms that are sensitive to the interocular correlation.

In this study, we used horizontal stimuli because vergence adjustments (which would disrupt our binocular stimuli) are smaller for horizontal than vertical stimuli [17, 18]. However, Li and Atick's theory of efficient stereo coding predicts summation and differencing channels at all orientations and uses the vertical summation and differencing channels for disparity or 3D depth computation. Other models of disparity computation do not explicitly recognize the statistical separability of the summation and differencing channels, or how crucial it is for efficient coding that the gains on these channels be independently adaptable.

There has been an intense debate about whether "cyclopean motion" of the kind used in this study is detected by dedicated motion sensors, or by tracking the changing positions of features over time [19–27]. But to date, all models of motion perception that can accommodate cyclopean motion have a key omission in common: they lack a binocular differencing channel. Our results provide strong evidence for the existence of binocular summation and differencing channels in human vision; the gains on these channels can adapt strongly to the prevailing interocular correlation in less than 1 s, to maximize coding efficiency. This fast adaptation may explain why the difference between the effects of adapting to natural versus nonnatural stereograms was very weak in a previous study [28], in which a test stimulus appeared 472 ms after the offset of the adapting stimulus. When the binocular summation and difference signals move in opposite directions, the perceived direction of motion can be reversed by changing the interocular correlation of the adaptation stimuli, even if the adaptation stimuli are static. This finding provides particularly strong support for Li and Atick's theory of efficient stereo coding [1, 2], because it is difficult to imagine any other reason why adaptation to static stimuli would lead to a reversal of perceived motion direction. With the recent surge in popularity of 3D movies and video games, there is currently a great deal of interest in the effects of viewing stereo displays, given that they are likely to differ from the normal visual environment. Our study shows one way in which human binocular vision mechanisms adapt to stereo displays, and how this could lead to perceptual biases. Our finding that adaptation is fast suggests that at least some effects of stereoscopic viewing may be short lived.
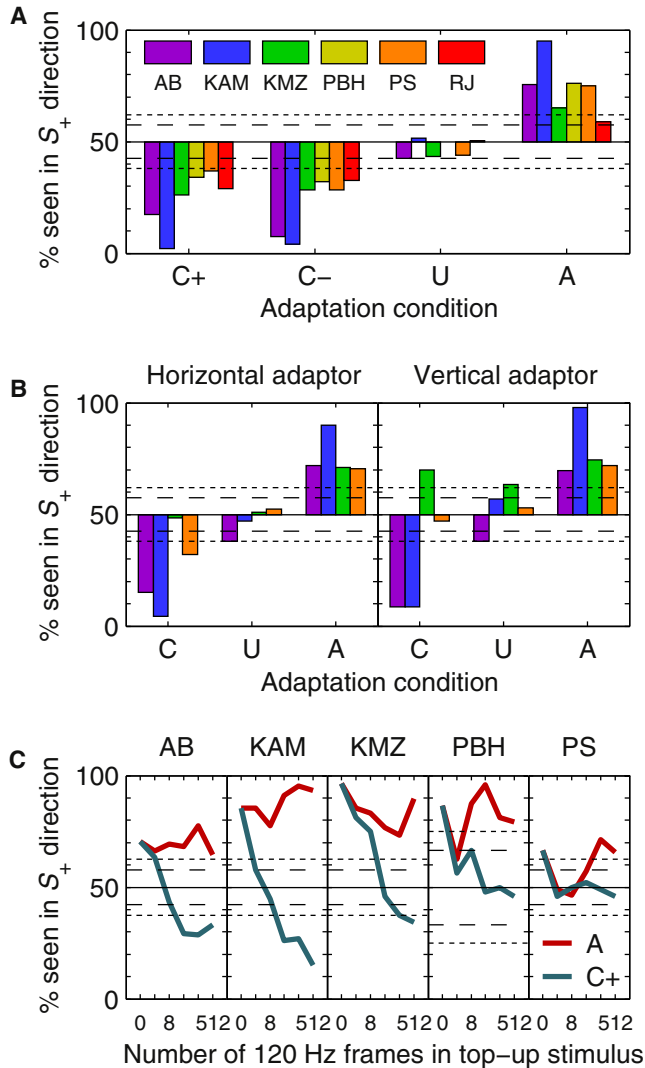
Figure 4. Results

Vertical axes give the percentage of times that the test stimulus was reported to be moving in the S_+ direction.

(A) Experiment 1. Each color represents data from a different observer, as indicated along the top of the panel. Bars are grouped according to adaptation condition (C+, correlated+; C−, correlated−; U, uncorrelated; A, anticorrelated). Dashed lines with long dashes indicate performance levels significantly different from chance at the 0.05 level; dashed lines with short dashes indicate a significant difference from chance at the 0.001 level. These were calculated according to a two-tailed binomial test with 200 trials, so they do not apply to observer PBH, who did 100 trials per condition; however, PBH's performance on the correlated+, correlated−, and anticorrelated conditions was still significantly different from chance, the highest p value from these three conditions being 0.00179, for correlated+.

(B) Experiment 2. Each color is assigned to the same observer as in (A). Left and right panels give data for horizontal and vertical adaptors, respectively. Dashed lines have the same meaning as in (A) and apply to all observers.

(C) Experiment 3. Each panel gives data for a different observer. The horizontal axis represents the duration of the top-up stimulus (tick marks represent durations of 0, 2, 8, 32, 128, and 512 frames of 120 Hz). The two colored lines represent data for trials preceded by different types of top-up adaptation stimulus (A, anticorrelated; C+, correlated+). The lines necessarily converge at zero (0) frames top-up duration because then the two adaptation conditions are identical and the data were pooled. For nonzero top-up adaptation durations, adaptation switched between correlated+ and anticorrelated on alternate trials within a session. Dashed lines have the same meaning as in (A) and (B) and are based on 48 trials per condition for observer PBH and 196 trials per condition for all other observers.

Table 1. Effect of Adaptor Interocular Correlation on Perceived Direction Judgments in Experiment 2

Observer	Horizontal Adaptor		Vertical Adaptor	
	χ^2	p	χ^2	p
AB	135	$<1.11 \times 10^{-16}$	157	$<1.11 \times 10^{-16}$
KAM	293	$<1.11 \times 10^{-16}$	324	$<1.11 \times 10^{-16}$
KMZ	24.8	4.13×10^{-6}	5.75	0.0563
PS	59.4	1.24×10^{-13}	27.9	8.95×10^{-7}

For each observer and adaptor orientation, the trials were classified according to reported direction (S_+ or S_- direction) and interocular correlation of adaptor. The resulting 2×3 contingency table was subjected to a chi-square test to see whether interocular correlation of the adaptor had significantly affected the responses. KMZ showed no significant effect for vertical adaptors (p > 0.05); in every other case, the effect was highly significant.

Experimental Procedures

Apparatus and Observers
Stimuli were displayed on a Clinton Monoray CRT monitor with fast-decaying (DP104) phosphor and frame rate 120 Hz, driven by a ViSaGe stimulus generator (Cambridge Research Systems). Images were presented on a 768 × 768 pixel, 19 cd/m^2 background with dark surround. Human observers (four male, two female) viewed the screen through FE-1 stereo goggles (Cambridge Research Systems) from a distance of 120 cm, giving 1 arcmin of visual angle per pixel. Informed consent was obtained from all observers, and approval of the study was obtained from the UCL Psychology Ethics Committee.

Test Stimuli
In the following exposition, the stimulus signals are expressed in terms of the contrast signal [29], $c(x, y, t)$, at each point in space (x, y) and time t, so the stimulus luminance, $L(x, y, t)$, is obtained from the contrast as

$$L(x, y, t) = L_0[1 + c(x, y, t)], \qquad (1)$$

where L_0 is the background luminance.

In Shadlen and Carney's stimulus [3], one eye receives a horizontal counterphase grating, S_1 (i.e., a stationary sine wave grating with contrast that reverses sinusoidally over time); the other eye receives a similar counterphase grating, S_2, which is shifted in spatial and temporal phase by $\pi/2$ radians relative to S_1. We can express S_1 as

$$S_1 = \alpha \cos Y \cos T,$$

where α is the stimulus contrast and Y and T are vertical distance and time, converted to instantaneous phase:

$$Y = 2\pi f_y y + \varphi_y \qquad (2)$$

$$T = 2\pi f_t t + \varphi_t, \qquad (3)$$

where y is vertical distance in degrees of visual angle, t is time in seconds, f_y and f_t are the spatial and temporal frequencies, respectively, and φ_y and φ_t are the spatial and temporal initial phases (at $y = 0$ and $t = 0$, respectively). S_2 is given by

$$S_2 = \alpha \sin Y \sin T.$$

Our proposed summation and differencing channels would see gratings drifting in opposite directions:

$$S_+ = S_1 + S_2 = \alpha \cos(Y - T)$$

$$S_- = S_1 - S_2 = \alpha \cos(Y + T).$$

Shadlen and Carney reported that subjects generally perceived motion in the summation direction [3], so, to remove this bias, we added drifting gratings, $\pm\beta\cos(X + T)$, with opposite contrast in each eye:

$$S_1 = \alpha \cos Y \cos T + \beta \cos(Y + T) \qquad (4)$$

$$S_2 = \alpha \sin Y \sin T - \beta \cos(Y + T) \qquad (5)$$

The drifting gratings cancel out in the summation channel but boost the difference signal:

$$S_+ = S_1 + S_2 = \alpha \cos(Y - T) \tag{6}$$

$$S_- = S_1 - S_2 = (\alpha + 2\beta)\cos(Y + T) \tag{7}$$

The test stimulus was a horizontal dichoptic motion stimulus as described in Equations 2–5, with $\alpha = 0.1$, $\beta = 0.02$ (experiments 1 and 3) or $\beta = 0.025$ (experiment 2), $f_y = 0.25$ cycles/degree, and $f_t = \pm 6$ Hz. We used the same stimulus parameters (including α and β) for all observers. φ_y and φ_t were randomized on each trial. S_+ direction (upward or downward, controlled by the sign of f_t) and assignment of signals S_1 and S_2 to left and right eyes were counterbalanced across trials within each session.

Supplemental Information

Acknowledgments

References

1. Li, Z., and Atick, J.J. (1994). Efficient stereo coding in the multiscale representation. Network *5*, 157–174.
2. Li, Z. (1995). Understanding ocular dominance development from binocular input statistics. In The Neurobiology of Computation: The Proceedings of the Third Annual Computation and Neural Systems Conference, J. Bower, ed. (Boston: Kluwer), pp. 397–402.
3. Shadlen, M., and Carney, T. (1986). Mechanisms of human motion perception revealed by a new cyclopean illusion. Science *232*, 95–97.
4. Hibbard, P.B. (2008). Binocular energy responses to natural images. Vision Res. *48*, 1427–1439.
5. Pelli, D.G. (1990). The quantum efficiency of vision. In Vision: Coding and Efficiency, C. Blakemore, ed. (Cambridge: Cambridge University Press), pp. 3–24.
6. Pelli, D.G. (1991). Noise in the visual system may be early. In Computational Models of Visual Processing, M. Landy and J.A. Movshon, eds. (Cambridge, MA: MIT Press), pp. 147–152.
7. Livingstone, M.S., and Hubel, D.H. (1984). Anatomy and physiology of a color system in the primate visual cortex. J. Neurosci. *4*, 309–356.
8. Snodderly, D.M., and Gur, M. (1995). Organization of striate cortex of alert, trained monkeys (*Macaca fascicularis*): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. J. Neurophysiol. *74*, 2100–2125.
9. Cohn, T.E., and Lasley, D.J. (1976). Binocular vision: two possible central interactions between signals from two eyes. Science *192*, 561–563.
10. Yoonessi, A., and Kingdom, F.A. (2009). Dichoptic difference thresholds for uniform color changes applied to natural scenes. J. Vis. *9*, 3.1–3.12.
11. Maehara, G., and Goryo, K. (2005). Binocular, monocular and dichoptic pattern masking. Opt. Rev. *12*, 76–82.
12. Ding, J., and Sperling, G. (2006). A gain-control theory of binocular combination. Proc. Natl. Acad. Sci. USA *103*, 1141–1146.
13. Meese, T.S., Georgeson, M.A., and Baker, D.H. (2006). Binocular contrast vision at and above threshold. J. Vis. *6*, 1224–1243.
14. Baker, D.H., Meese, T.S., and Summers, R.J. (2007). Psychophysical evidence for two routes to suppression before binocular summation of signals in human vision. Neuroscience *146*, 435–448.
15. Baker, D.H., Meese, T.S., and Georgeson, M.A. (2007). Binocular interaction: contrast matching and contrast discrimination are predicted by the same model. Spat. Vis. *20*, 397–413.
16. Baker, D.H., and Meese, T.S. (2007). Binocular contrast interactions: dichoptic masking is not a single process. Vision Res. *47*, 3096–3107.
17. Kertesz, A.E. (1981). Effect of stimulus size on fusion and vergence. J. Opt. Soc. Am. *71*, 289–293.
18. Howard, I.P., Fang, X., Allison, R.S., and Zacher, J.E. (2000). Effects of stimulus size and eccentricity on horizontal and vertical vergence. Exp. Brain Res. *130*, 124–132.
19. Georgeson, M.A., and Shackleton, T.M. (1989). Monocular motion sensing, binocular motion perception. Vision Res. *29*, 1511–1523.
20. Carney, T., and Shadlen, M.N. (1992). Binocularity of early motion mechanisms: comments on Georgeson and Shackleton. Vision Res. *32*, 187–191.
21. Georgeson, M.A., and Shackleton, T.M. (1992). No evidence for dichoptic motion sensing: a reply to Carney and Shadlen. Vision Res. *32*, 193–198.
22. Carney, T., and Shadlen, M.N. (1993). Dichoptic activation of the early motion system. Vision Res. *33*, 1977–1995.
23. Lu, Z.-L., and Sperling, G. (1995). The functional architecture of human visual motion perception. Vision Res. *35*, 2697–2722.
24. Patterson, R. (1999). Stereoscopic (cyclopean) motion sensing. Vision Res. *39*, 3329–3345.
25. Lu, Z.-L., and Sperling, G. (2001). Three-systems theory of human visual motion perception: review and update. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *18*, 2331–2370.
26. Patterson, R. (2002). Three-systems theory of human visual motion perception: review and update: comment. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *19*, 2142–2143, discussion 2144–2153.
27. Lu, Z.-L., and Sperling, G. (2002). Stereomotion is processed by the third-order motion system: reply to comment on "Three systems theory of human visual motion perception: review and update". J. Opt. Soc. Am. A Opt. Image Sci. Vis. *19*, 2144–2153.
28. Chen, D., and Li, Z. (1997). A psychophysical experiment to test the efficient stereo coding theory. In Theoretical Aspects of Neural Computation: A Multidisciplinary Perspective, K.-Y.M. Wong, I. King, and D.-Y. Yeung, eds. (New York: Springer-Verlag), pp. 225–235.
29. Watson, A.B., Barlow, H.B., and Robson, J.G. (1983). What does the eye see best? Nature *302*, 419–422.
30. Atick, J.J., and Redlich, A.N. (1992). What does the retina know about natural scenes? Neural Comput. *4*, 196–210.
31. Atick, J.J. (1992). Could information theory provide an ecological theory of sensory processing? Network *3*, 213–251.
32. Shannon, C.E. (1948). A mathematical theory of communication. Bell Sys. Tech. J. *27*, 379–423, 623–656.