Direction discrimination thresholds in binocular, monocular, and dichoptic viewing: Motion opponency and contrast gain control

Goro Maehara Department of Human Science, Kanagawa University, Yokohama, Japan

Robert F. Hess Department of Ophthalmology, McGill University, Montreal, Quebec, Canada

Mark A. Georgeson School of Life and Health Sciences, Aston University, Birmingham, UK

We studied the binocular organization of motion opponency and its relationship to contrast gain control. Luminance contrast thresholds for discriminating direction of motion were measured for drifting Gabor patterns (target) presented on counterphase flickering Gabor patterns (pedestal). There were four presentation conditions: binocular, monocular, dichoptic, and half-binocular. For the half-binocular presentation, the target was presented to one eye while pedestals were presented to both eyes. In addition, to test for motion opponency, we studied two increment and decrement conditions, in which the target increased contrast for one direction of movement but decreased it for the opposite moving component of the pedestal. Threshold versus pedestal contrast functions showed a dipper shape, and there was a strong interaction between pedestal contrast and test condition. Binocular thresholds were lower than monocular thresholds but only at low pedestal contrasts. Monocular and half-binocular thresholds were similar at low pedestal contrasts, but half-binocular thresholds became higher and closer to dichoptic thresholds as pedestal contrast increased. Adding the decremental target reduced thresholds by a factor of two or more—a strong sign of opponency—when the decrement was in the same eye as the increment or the opposite eye. We compared several computational models fitted to the data. Converging evidence from the present and previous studies (Gorea, Conway, & Blake, 2001) suggests that motion opponency is most likely to be monocular, occurring before direction-specific binocular summation and before divisive, binocular gain control.

Introduction

In the study of motion perception, there has been an extended debate over whether the direction-selective mechanisms of motion sensors are monocular or binocular. Anstis and Duncan (1983) found that motion aftereffects can occur separately for the left and right eyes, suggesting that at least some motion sensors are monocular. However, Shadlen and Carney (1986) reported that observers perceived apparent motion while viewing dichoptic motion stimuli. Their stimulus consisted of two monocular flickering patterns in which the phase of one was spatially and temporally shifted by 90° relative to the other. The sum of these two flickering patterns would form a moving one, and because there was no directional component in each eye, Shadlen and Carney concluded that motion sensors must be binocular and capable of integrating dichoptic inputs to encode motion direction. Georgeson and Shackleton (1989) also reported the existence of dichoptic apparent motion but argued that its basis was the spatiotemporal correspondence of visible features ("feature tracking"), not early motion sensors. This may well be one basis for dichoptic motion perception. But later evidence has shown that observers perceived dichoptic motion even when there was no feature to track in either eye, thus supporting the existence of binocular motion sensors (Carney, 1997; Carney & Shadlen, 1993; Derrington & Cox, 1998; Lu & Sperling, 2001; Hayashi, Nishida, Tolias, & Logothetis, 2007). Nevertheless, there is general agreement in these studies that such dichoptic motion is much

Citation: Maehara, G., Hess, R. F., & Georgeson, M. A. (2017). Direction discrimination thresholds in binocular, monocular, and dichoptic viewing: Motion opponency and contrast gain control. *Journal of Vision*, *17*(1):7, 1–21, doi:10.1167/17.1.7.

doi: 10.1167/17.1.7 Received September 23, 2016; published January 10, 2017 ISSN 1534-7362

weaker than the corresponding monocular motion (with which the same stimulus components are physically summed within one eye).

Computational models of motion processing have incorporated motion opponency and divisive gain control (Adelson & Bergen, 1985; Georgeson & Scott-Samuel, 1999; Simoncelli & Heeger, 1998), but binocular processing has received less attention there. Inputs to these models are binocularly presented stimuli and not separated for the left and right eyes. Motion processing models typically include a motion-opponent mechanism that is sensitive only to the difference in contrast or energy between opposite directions. Opponency explains why we cannot perceive two opposite motions at the same time when they are in the same location and the same spatial frequency range (Qian & Andersen, 1994; Qian, Andersen, & Adelson, 1994a, 1994b; Van Doorn & Koenderink, 1982). When two sine wave gratings drift in opposite directions with the same luminance contrast, there is no impression of two opposite, transparent motions, and the grating (Figure 1) typically appears to be counterphase flickering or oscillating (Kelly, 1966; Kulikowski, 1971). Motion opponency is also supported by motion aftereffects in which we perceive motion in the direction opposite to that of adapting motion stimuli.

Our questions here concern the binocular properties of motion opponency and divisive gain control. We address these issues by fitting computational models to threshold data. The present experiment measured luminance contrast thresholds for discriminating direction of motion for drifting Gabor patterns (target) presented on counterphase flickering Gabor patterns (pedestal, equivalent to the superposition of two Gabors drifting in opposite directions). There were four presentation conditions: (a) binocular: all stimuli were presented to both eyes, (b) monocular: all stimuli were presented to one eye and not the other, (c) dichoptic: the target was presented to one eye while the pedestal was presented to the other eye, and (d) half-binocular: the target was presented to one eye while pedestals were presented to both eyes.

In addition, we tested incremental and decremental targets, with which the target increased contrast for one direction of movement but decreased it by the same amount for the opposite moving component of the pedestal. In a motion-opponent mechanism, decreasing the signal strength in one direction should be almost equivalent to increasing it in the other. Hence the combination of incremental and decremental targets should create a much stronger opponent response than the increment alone. In our experiment, the decrement was either in the same eye as the increment or in the other eye, and this might test whether the motion opponency mechanism is capable of binocular integration. According to Stromeyer, Klein, Kronauer, and

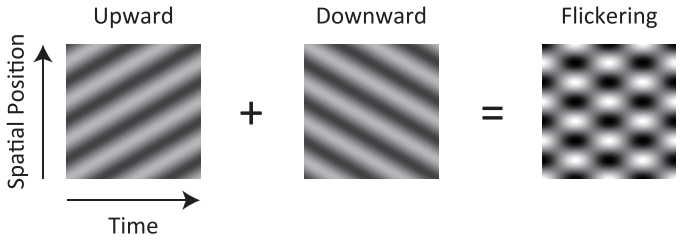

Figure 1. A flickering grating is the sum of two gratings drifting in opposite directions.

Madsen (1984), observers were significantly more sensitive to luminance contrast change (contrast discrimination) when target stimuli consisted of a contrast increment in one direction and a decrement in the opposite direction than when luminance contrast of both motion components was increased (or decreased). This advantage for the increment/decrement condition is strong evidence for motion opponency. Gorea, Conway, and Blake (2001) found that this advantage disappeared when the two opposite directions of movement and the associated increment and decrement were presented separately to the left and right eyes. They concluded that motion opponency must be a monocular process before binocular combination.

In the present experiment, we asked observers to discriminate motion direction instead of discriminating changes in luminance contrast. Because contrast discrimination does not necessarily require perception of motion, especially near threshold, direction discrimination is a more direct way of studying motion processing. Moreover, we measured thresholds over a wide range of flickering pedestal contrast (11 levels between 0% and 40%), and Gorea et al. (2001) tested only one flickering pedestal contrast (40%). This broad range of conditions enabled us to distinguish between several different computational models for the direction discrimination data. We also applied several variants of these models to the Gorea et al. data and, taken together, these analyses point to some fairly firm conclusions about the binocularity (or otherwise) of motion opponency and contrast gain control.

Methods

Observers

There were three observers, JB, GM, and PCH. All had corrected-to-normal visual acuity. GM is one of authors. All observers provided fully informed consent to participate in this study, and the study followed protocols approved by the institutional ethics commit-

tee that were in accordance with the Declaration of Helsinki.

Apparatus

Stimuli were generated using a VSG 2/5 (Cambridge Research System Ltd., Kent, UK), which produces 15-bit gray level resolution and presented on a CRT video monitor (Compaq P1210). The display resolution was set to 1024 × 768 pixels. The refresh rate of the monitor was set to 120 Hz. The highest luminance of the display was 60 cd/m^2. The image on one half of the screen was directed to one eye while the image on the other half was directed to the other eye by means of an eight-mirror stereoscope. Presentation regions on the monitor subtended a visual angle of 10° high × 8.5° wide for each eye. The viewing distance was 57 cm.

Stimuli

Targets were drifting Gaussian-windowed sinusoidal gratings (Gabor patterns). The gratings had a spatial frequency of 1 c/° and were oriented at 90° (horizontal stripes). The standard deviation of the Gaussian window function was 0.6° of visual angle. The gratings drifted upward or downward at a speed of 7.5° of visual angle per second (7.5 Hz) within the stationary Gaussian window.

Pedestals were counterphase flickering Gabor patterns. Their spatial frequency, orientation, and Gaussian window were identical to those of the targets. The flicker rate was 7.5 Hz. As shown in Figure 1, a flickering Gabor pattern is equivalent to the superposition of two Gabor patterns drifting in opposite directions. That is, pedestals can be divided into upward and downward drifting targets whose luminance contrast is half that of the flickering pedestal contrast.

Targets were presented on flickering pedestals. There were two types of targets: incremental targets and incremental and decremental targets. Incremental targets increased luminance contrast for one direction of movement (Figure 2A, left). On the other hand, incremental and decremental targets increased contrast for one direction of movement but decreased it by the same amount for the opposite moving component of the pedestal (Figure 2A, right).

The mean luminance of the stimuli was 30 cd/m^2. Their luminance contrast was defined as Michelson contrast and was expressed in dB re 1%, where 1 dB is 1/20 of a log unit of contrast. That is, 0 dB and 40 dB correspond to 1% and 100% of luminance contrast, respectively. Targets and pedestals were simultaneously presented for 267 ms (two temporal cycles) at the center of the presentation region. We used brief presentations to minimize binocular rivalry.

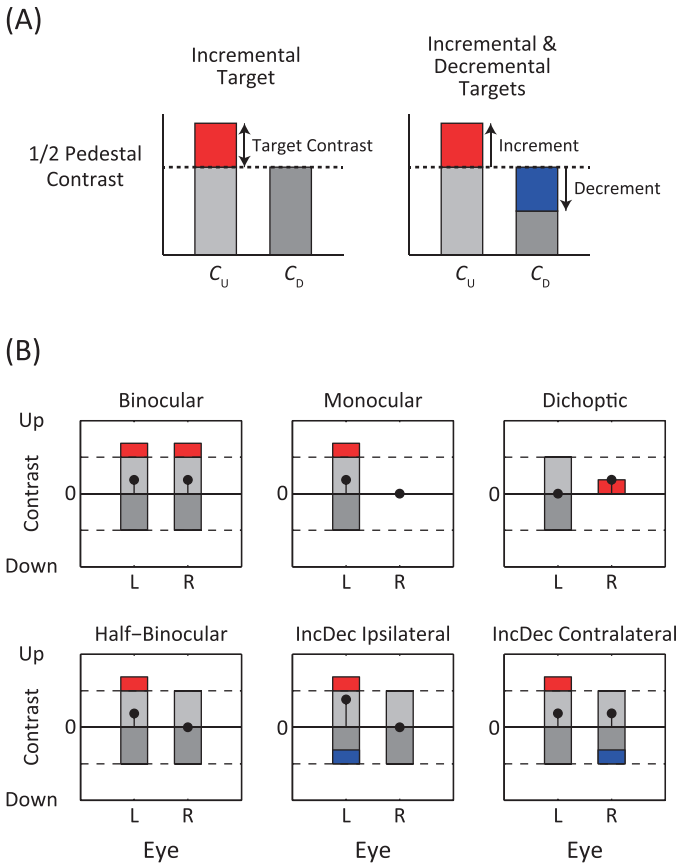

Figure 2. (A) Increment and decrement in moving components of a flickering pedestal. Incremental targets increased luminance contrast of one moving component. Incremental and decremental targets increased one component but also decreased luminance contrast of the opposite moving component. C_U and C_D: luminance contrast of the upward and downward moving component, respectively. (B) Graphical representation of the six test conditions. Gray bars represent contrast of the pedestal components, moving up or down. Red tab indicates a contrast increment, blue tab a contrast decrement. Black symbol is the difference in contrast between the two moving components ($C_U - C_D$). It can be thought of as the net amount of motion in the stimulus for a given eye.

Procedure

The present experiment measured luminance contrast thresholds for discriminating motion direction of targets presented on pedestals. There were 11 levels of pedestal contrast (−∞, −4, 0, 4, 8, 12, 16, 20, 24, 28, and 32 dB) for each presentation condition described below.

For the incremental targets, there were four presentation conditions: the binocular, monocular, dichoptic, and half-binocular presentations (Figure 2B). All

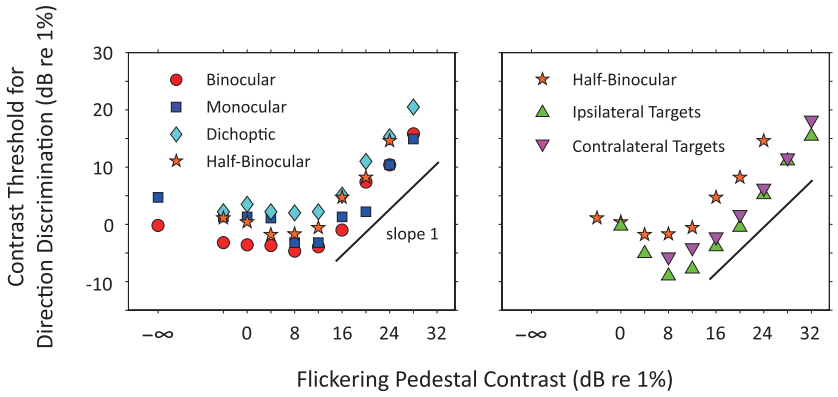

Figure 3. Contrast thresholds for direction discrimination (mean of three observers). Left panel: thresholds for the incremental target. Right panel: thresholds for the incremental and decremental targets (upright and inverted triangles for ipsilateral and contralateral target presentations, respectively).

stimuli were presented to both eyes under the binocular presentation condition whereas they were presented to the same single eye under the monocular presentation condition. For the dichoptic presentation, the target was presented to one eye while the pedestal was presented to the other eye. For the half-binocular presentation, the target was presented to one eye while pedestals were presented to both eyes.

For the incremental and decremental targets, there were two presentation conditions: the ipsilateral and contralateral target presentations (Figure 2B). For the ipsilateral targets, the decrement was presented to the same eye as the increment whereas for the contralateral targets the decrement was presented to the other eye. These presentation conditions share a similarity with the half-binocular presentation in that pedestals were presented to both eyes while the increment was applied to only one eye.

The protocol in this experiment was a single-interval, direction discrimination task. In each trial, the incremental target (Figure 2A, B) drifted upward or downward. Observers judged the direction of motion. Feedback was given after each incorrect response. A one-up/three-down staircase was used to adjust the target contrast, increasing it after one error or decreasing it after three correct responses. The step size of the staircase was initially set at 4 dB and moved to 2 dB after the second reversal. The staircase terminated after seven reversals. Observers completed four staircases for each condition. Target contrast thresholds (at 75% correct) and standard errors were determined by fitting a logistic psychometric function to the response data (the number of correct and incorrect responses) using the Palamedes toolbox (Kingdom & Prins, 2010; Prins & Kingdom, 2009). Another four staircases were conducted for the condition in which the standard error exceeded 4 dB. In such a case, thresholds were based on eight staircases in total.

Results

Figure 3 shows mean target contrast thresholds for direction discrimination as a function of the flickering pedestal contrast (TvC function). Individual results are shown in Figure 4. In some conditions, observers were not able to discriminate the direction even at the highest possible target contrast. Those data points are missing in Figure 4. We averaged thresholds and plotted them in Figure 3 when thresholds were obtained for all three observers. It should be noted that averaging might make the dips shallower because of individual differences in sensitivity.

The direction discrimination thresholds were lower under binocular viewing (red circles) than under monocular viewing (blue squares) at least for low-flickering pedestal contrasts. We calculated binocular summation ratios in the absence of a pedestal by dividing the monocular contrast threshold (not in dB) by the binocular threshold at zero pedestal contrast. These binocular summation ratios were 1.71, 1.80, and 1.82 for JB, GM, and PCH, respectively. Previous research has found binocular summation ratios typically between 1.4 and 2 (Arditi, Anderson, & Movshon, 1981; Legge, 1984a; Maehara & Goryo, 2005; Meese, Georgeson, & Baker, 2006; Rose, 1978).

The TvC functions had a typical dipper shape when thresholds decreased and then increased with pedestal contrast under binocular and monocular viewing (red circles and blue squares in Figures 3 and 4). The amount of dip was much smaller for the dichoptic presentation (light blue diamond) than for the binocular and monocular presentations. Unlike the low-contrast conditions, there was little or no binocular advantage across a wide range of suprathreshold pedestal contrasts. Thresholds for the half-binocular presentation (orange stars in Figure 3 and 4) were close to those for the monocular presentation (blue square)

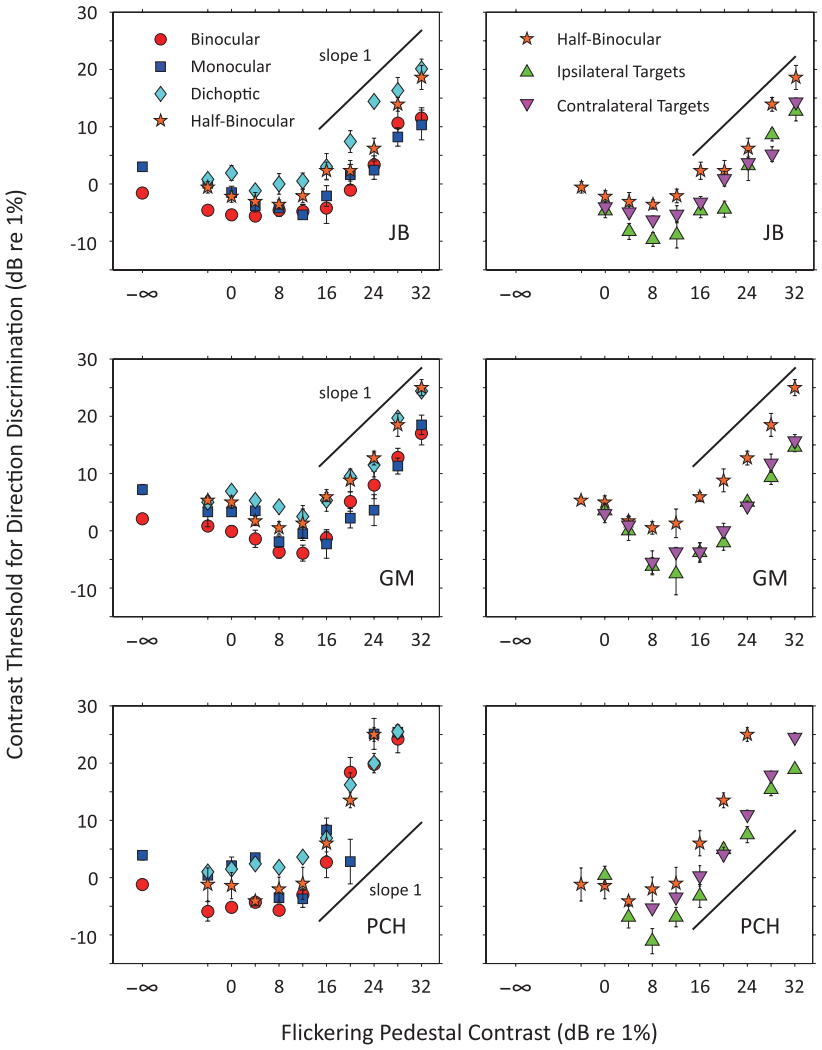

Figure 4. Contrast thresholds for direction discrimination for three observers. Left panels: thresholds for the incremental target. Right panels: thresholds for the incremental and decremental targets (upright and inverted triangles for ipsilateral and contralateral target presentations, respectively). Error bars shows standard errors estimated by maximum likelihood fitting.

at low pedestal contrasts. But at intermediate-to-high pedestal contrasts, thresholds were higher for the half-binocular presentation than for the monocular one.

Slopes of the present TvC functions were close to one or slightly higher than one at high pedestal contrasts (Figures 3 and 4) whereas they were consistently lower than one (about 0.5 to 0.7) for contrast discrimination of stationary stimuli (Legge, 1984a; Maehara & Goryo, 2005; Meese et al., 2006).

It can be seen from Figures 3 and 4 (right-hand panels) that thresholds were about 6–8 dB lower for the combination of monocular incremental and decremental targets (green triangles and purple inverted triangles) than for incremental targets alone (orange stars, half-binocular). This opponency advantage is consistent with the results of the previous studies (Gorea et al., 2001; Stromeyer et al., 1984). Observers could not discriminate the direction at low pedestal contrasts (missing data points at –4 dB for all observers; 0 dB and 4 dB for PCH's contralateral target presentation). At these low pedestal contrasts, the decremental target is not always well defined: If the decremental target contrast exceeds the pedestal contrast, then spatial phase reverses, and the net target plus pedestal contrast increases instead of continuing to decrease.

To assess any difference in threshold between the ipsilateral and contralateral targets (green triangles and purple inverted triangles in Figures 3 and 4), we subjected the data at intermediate and high pedestal contrasts to two-way ANOVA with factors of Target (ipsilateral or contralateral) and Pedestal Contrast (8, 12, 16, 20, 24, 28, or 32 dB). Although average thresholds were slightly lower (2.2 dB) for the ipsilateral targets than for the contralateral targets, the main effect of Target was not significant, $F(1, 2) = 13.9$, $p = 0.0651$. The interaction with Pedestal Contrast was also not significant, $F(6, 12) = 1.02$, $p = 0.457$.

One might argue that, if binocular rivalry takes place, dichoptic thresholds should be measured separately for the suppressed eye and the dominant eye. However, in our dichoptic presentations, the pedestal was flickering in one eye while the test was drifting in the other. Thus the test component can sum binocularly with the same direction component of the pedestal in the other eye. According to Blake and Boothroyd (1985) summation takes precedence over rivalry, and so it seems unlikely that rivalry will be invoked under these conditions, especially for our brief presentations. Moreover, Gorea et al. (2001) found that dichoptic thresholds were not significantly different between the suppressed and the dominant eyes when they used drifting and flickering gratings as stimuli. Therefore, it seems reasonable to pool the data over all trials for the dichoptic presentation in the present experiment.

Modeling

The aim of the present study is to construct binocular versions of motion processing models and to explain the threshold data using them. For this purpose, we incorporate binocular processing into the motion contrast model (Georgeson & Scott-Samuel, 1999) on the basis of binocular processing models of luminance contrast (Maehara & Goryo, 2005; Meese et al., 2006).

Performance on various visual tasks is known to be better with two eyes than with one eye (binocular summation; Blake & Fox, 1973; Blake, Sloane, & Fox, 1981). Research on luminance contrast perception has addressed binocular processing. Legge (1984b) proposed quadratic summation as a rule that describes binocular summation in luminance contrast detection of static patterns. Quadratic summation means that monocular signals are squared and added to form a binocular signal. Maehara and Goryo (2005) revised Foley's (1994) divisive gain control model of luminance contrast processing to account for detection and discrimination thresholds of luminance contrast under binocular, monocular, and dichoptic viewing. The revised model, called the twin summation model, receives inputs from the left and right eyes separately. There is a similarity between quadratic summation and the twin summation model in that monocular signals are accelerated exponentially before their summation for generating binocular signals. This summation is followed by divisive inhibition among processing units tuned to different orientations and spatial frequencies. Meese et al. (2006) proposed a related model with two stages of divisive gain control. The two monocular processing pathways have a suppressive interaction at the first stage, and this is followed by the divisive gain control at the second, binocular stage. Research on binocular rivalry has also suggested that there are two stages of inhibition for monocular and binocular processing (Blake, 1989; Lehky, 1988; Wilson, 2003).

Spatiotemporal filters

The first processing stage of the present models is spatial and temporal filters, which were originally proposed by Adelson and Bergen's (1985) motion energy model. The models convolve the image sequence with two spatial filters, which differ in position, and two temporal filters, one of which is delayed relative to the other. Outputs from the filtering process are summed or subtracted to create direction-selective responses. The responses are then squared and summed, giving phase-invariant, direction-specific signals called *motion energy*.

Although our models, in principle, also apply these filters to the image sequence, the process can be simplified here. We just assume that there are spatiotemporal filters that yield a motion signal proportional to luminance contrast of motion components at the monocular processing stage. That is, the monocular excitatory signal E_{ij} for the target motion direction i in eye j is

$$E_{ij} = C_{tj}S_E + C_{pj}S_E/2,$$

where C_{tj} and C_{pj} are target and pedestal luminance contrast, respectively, expressed as Michelson contrast, and S_E is the excitatory sensitivity. Because a flickering pedestal is the sum of two opposite motion components, we divide the pedestal contrast by two to get the contrast of its moving components. The target luminance contrast, C_{tj}, equals the increment or decrement in motion components (Figure 2). When no target is presented, $C_{tj} = 0$.

We assume another output, inhibitory signals, from the spatiotemporal filters for the denominator of the divisive gain control. The monocular inhibitory signal I_{ij} for the target motion direction is

$$I_{ij} = C_{tj}S_I + C_{pj}S_I/2,$$

where S_I is the inhibitory sensitivity. $C_{tj} = 0$ when no target is presented as for the calculation of excitatory signals.

The twin summation model of motion processing (TS1)

As mentioned earlier, we consider two contrast processing models that describe how monocular signals are combined to yield binocular signals: the twin

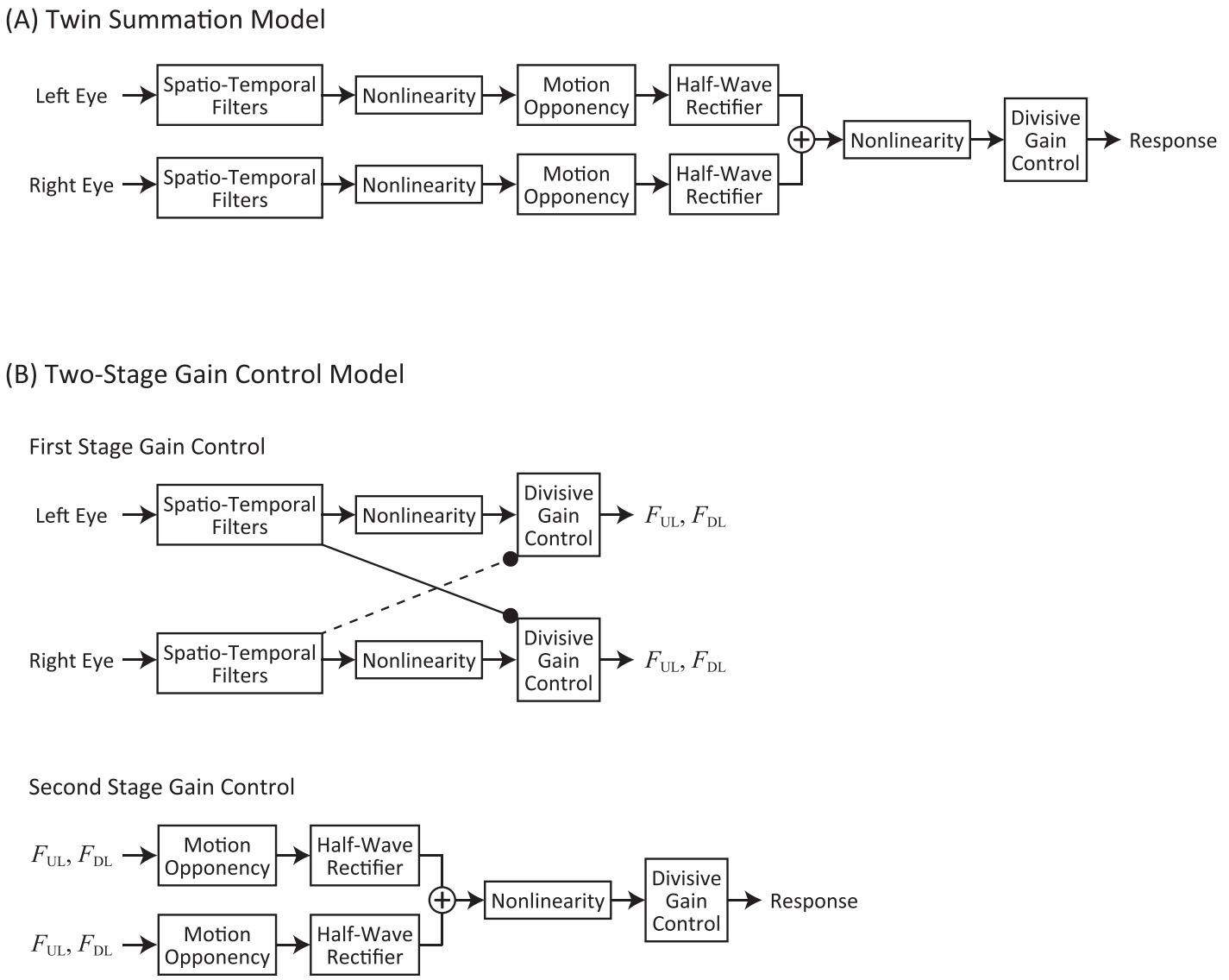

Figure 5. Schematic illustrations of binocular versions of motion processing models. (A) The twin summation model. (B) The two-stage gain control model. These diagrams show motion opponency within the monocular pathways. In the text, we also consider motion opponency at a late stage after binocular combination.

summation model (Maehara & Goryo, 2005) and the two-stage divisive gain control model (Meese et al., 2006). Our goal here was to develop plausible extensions of both these models to handle motion signals. First, we describe an opponent-motion model based on the twin summation model because this model has the simpler structure.

Figure 5A shows a schematic illustration of the twin summation model with monocular opponency, which we shall call *TS1*. Spatiotemporal filters produce four types of monocular excitatory signals—E_{UL}, E_{UR}, E_{DL}, and E_{DR}—and four types of monocular inhibitory signals—I_{UL}, I_{UR}, I_{DL}, and I_{DR}—for combinations of two motion directions (upward or downward, U or D) and two eyes (left or right, L or R). Monocular excitatory signals for the left and right eyes are raised to power m (nonlinear transducer) and subjected to motion opponency followed by half-wave rectification. The rectified opponent signals are summed between two eyes and raised again to power p before the divisive inhibition. In a similar way, inhibitory signals are raised to power n, summed, and raised again to power q. However, we assume no opponency for the inhibitory signals because for flickering pedestals the contrast gain control effect would be nullified through cancellation. Then, the divisive inhibition is applied to yield a binocular motion response M_i. These calculations are conducted for a specific direction i and expressed as

$$M_U = \frac{(hwr\{E^m_{UL} - E^m_{DL}\} + hwr\{E^m_{UR} - E^m_{DR}\})^p}{(I^n_{UL} + I^n_{UR})^q + z} \quad (1a)$$

$$M_D = \frac{(hwr\{E^m_{DL} - E^m_{UL}\} + hwr\{E^m_{DR} - E^m_{UR}\})^p}{(I^n_{DL} + I^n_{DR})^q + z}, \quad (1b)$$

where z is a constant, and direction $i = U$ or D. The function $hwr\{x\}$ is half-wave rectification, i.e., $max(x,0)$, serving to prevent negative responses. Note that we have two directional channels, each with opponent input from the other direction in the same eye, followed by direction-specific binocular summation. The constant z in the denominator is required to prevent division by 0 at zero contrast. More generally it (a) controls response gain at low contrasts with higher z giving lower responses, and (b) it controls the pedestal contrast level at which a low-threshold (or facilitation) regime gives way to the rising (masking) branch of the TvC function: higher z shifts that transition to higher contrasts. This description holds true for both a drifting grating, with which the response to the pedestal increases with contrast, and a flickering grating with which the opponent-mechanism response to such a pedestal is always zero (see Figure A4B).

Both these channels will be silent when the upward and downward inputs are balanced (no net motion in either eye), and so it is reasonable to suppose that direction will be discriminable when a response to the target direction is reliably nonzero. Thus, if the target direction is upward, that direction will be just detectable if $M_U = 1$.

The two-stage gain control model of motion processing

Figure 5B shows a schematic illustration of the two-stage gain control model. The characteristic of this model is that monocular processing pathways for the left and right eyes mutually suppress each other (Meese et al., 2006). The inclusion of interocular suppression is an advantage of this model because research on eye rivalry has suggested similar processing (Blake, 1989; Lehky, 1988; Wilson, 2003).

The processing starts with spatiotemporal filtering that is similar to the twin summation model. There are four monocular motion signals—E_{UL}, E_{UR}, E_{DL}, and E_{DR}—as output. The two-stage model uses them for both the numerator and denominator of the divisive inhibition.

The first stage of the divisive gain control implements interocular suppression. Specifically, the monocular motion signals are raised to power m and divided by the sum of the two monocular motion signals and a constant s, yielding the first-stage outputs F_{ij} for motion direction i in eye j:

$$F_{ij} = E^m_{ij}/(E_{iL} + E_{iR} + s). \quad (2)$$

The first-stage outputs are subjected to motion opponency, half-wave rectified, summed between two eyes, and then subjected to the second-stage divisive gain control, yielding the binocular motion response M_i. This calculation is expressed as

$$M_U = \frac{(hwr\{F_{UL} - F_{DL}\} + hwr\{F_{UR} - F_{DR}\})^p}{(F_{UL} + F_{UR})^q + z} \quad (3a)$$

$$M_D = \frac{(hwr\{F_{DL} - F_{UL}\} + hwr\{F_{DR} - F_{UR}\})^p}{(F_{DL} + F_{DR})^q + z}, \quad (3b)$$

where p and q are exponents of the nonlinearity for the numerator and denominator, respectively, and z is a constant.

Target contrast will be at threshold when a response to the target direction equals a constant value d. This constant, representing internal noise, is a free parameter in the two-stage model. It was fixed to be one in the twin summation model, in which internal noise is effectively bundled into the sensitivity terms, S_E, S_I. The two models are not formally identical, but they have many similarities.

Fitting the models to the data

The fitting procedure was as follows. Parameter values that gave a rough fit to data were found by trial and error as a starting point for least-squares fitting. Then the Matlab 'fminsearch' function (the *Simplex* algorithm) was used to fit the models. We computed 30 fits. Each fit started with a different set of parameter values randomly sampled from a normal distribution. Mean values of the normal distributions were set to be the rough fit values with a *SD* of 30%. The reported fits are those that achieved the lowest squared errors between model and data in dB. Numbers of data points to be fitted were 53, 60, 60, and 53 for mean data, JB, GM, and PCH, respectively.

The smooth curves in Figures 6 and 7 correspond to the best fits of the twin summation model to mean and individual data. Even though motion opponency is assumed to be before binocular summation, the *TS1* model predicts that there is no difference in thresholds between the ipsilateral and contralateral targets. The green and purple curves overlap completely in Figures 6 and 7. Errors and estimated parameters are given in

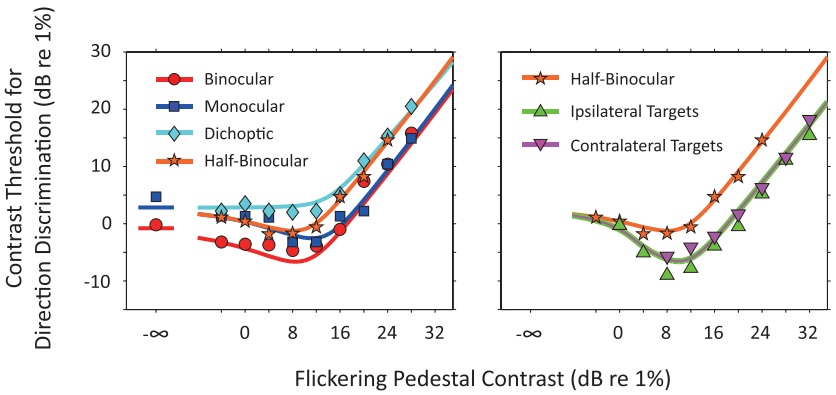

Figure 6. Fitting the twin summation model (*TS1*) to mean data of the three observers. Smooth curves correspond to the best fit. Two curves overlap for the ipsilateral and contralateral targets (green and purple lines).

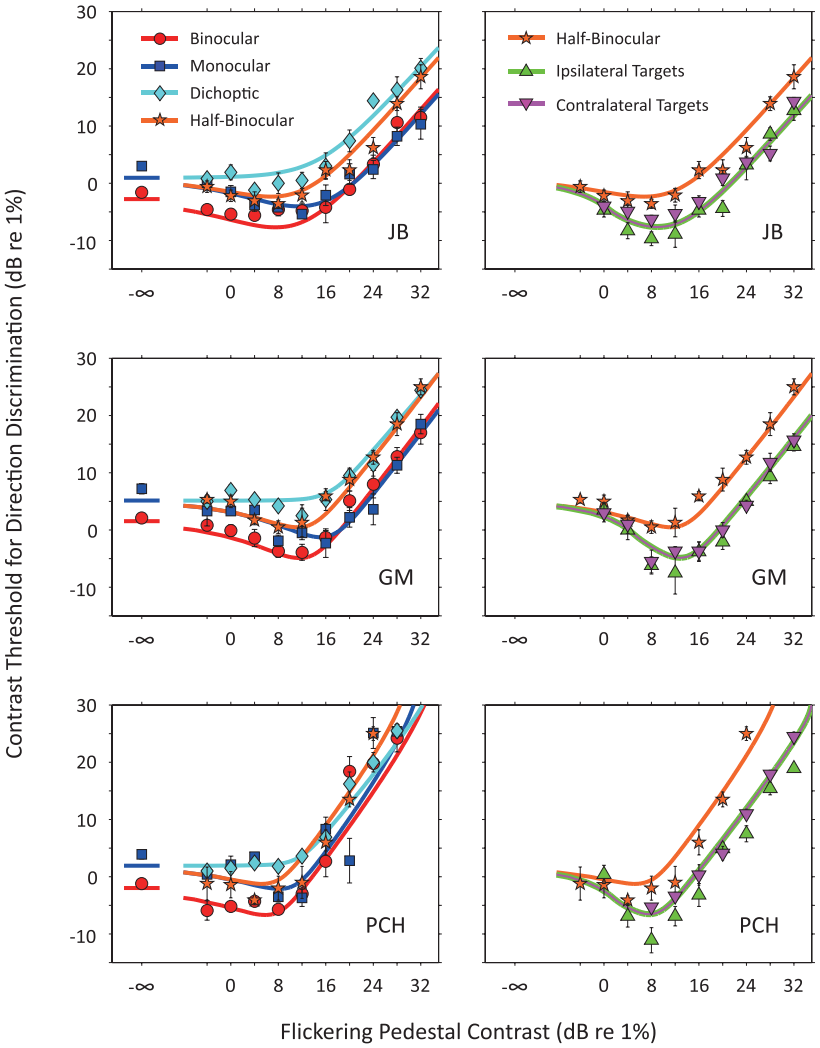

Figure 7. Fitting the twin summation model to individual data. Smooth curves correspond to the best fit. Two curves overlap for the ipsilateral and contralateral targets (green and purple lines).

(A) Twin summation model (*TS1*)	S_E	S_I	m	n	p	q	z	$m.p$	$n.q$	$m.p - n.q$	SSE	RMSE
Mean	100	48.2	1.66	1.97	2.57	2.42	3.89	4.27	4.77	−0.50	97.5	1.36
JB	100	36.2	1.62	1.58	1.33	1.47	1.14	2.15	2.32	−0.17	144	1.55
GM	100	42.3	1.68	1.71	2.81	3.10	16.0	4.72	5.30	−0.58	134	1.50
PCH	100	54.3	1.55	1.98	3.57	3.19	3.33	5.53	6.32	−0.78	528	3.12

(B) Two-stage gain control model	S_E	m	s	p	q	z	d	SSE	RMSE
Mean	100	1.79	0.129	3.12	3.30	3.25e-7	0.0490	86.6	1.28
JB	100	1.65	1.07	1.43	1.49	1.30e-4	0.157	143	1.54
GM	100	1.64	0.595	3.28	3.48	4.88e-8	0.0391	149	1.57
PCH	100	1.79	0.141	3.59	4.07	1.97e-9	0.173	534	3.17

Table 1A, B. Estimated free parameters and fitting errors. *Notes*: S_E was a fixed parameter. Numbers of data points were 53, 60, 60, and 53 for mean data, JB, GM, and PCH, respectively. SSE = sum of squared errors.

Table 1A. S_I, m, n, p, q, and z were free parameters; S_E was fixed to be 100 for compatibility with previous publications (Foley, 1994; Maehara & Goryo, 2005). The root mean squared errors (RMSEs) were 1.36 dB for group mean data, 1.55 dB for JB, 1.50 dB for GM, 3.12 dB for PCH. The fits were reasonably good and captured the major trends and many of the more subtle interactions in the data.

Table 1B shows errors and estimated parameters for fitting the two-stage model. The RMSEs were 1.28 dB for mean data, 1.54 dB for JB, 1.57 dB for GM, and 3.17 dB for PCH. The fits were as good as those with the twin summation model, and the fitted curves were almost identical (Figure S1 in the Supplementary Materials).

Discussion

The present experiment measured luminance contrast thresholds for direction discrimination of drifting targets presented on flickering pedestals. The stimuli were presented under binocular, monocular, or dichoptic viewing. First, we found that thresholds were lower for the binocular presentation than for the monocular presentation at the low pedestal contrast range, consistent with binocular summation in motion detection (Arditi et al., 1981; Rose, 1978). Second, thresholds were lowered and then elevated as pedestal contrast increased. This threshold reduction was much smaller for the dichoptic presentation than for other presentation conditions. Third, we found that when a contrast increment in the target direction was combined with a contrast decrement in the opposite direction, the contrast threshold for detecting the target direction improved by a factor of 2 to 2.5 (6–8 dB) compared with the increment alone. This form of synergy or cooperation between opposite directions strongly implies motion opponency. Put simply, if the upward (U) and downward (D) contrasts are $c + \delta c$ and $c - \delta c$, respectively, then (ignoring any nonlinearities) their opponent combination is $U - D = 2\delta c$, a factor of two gain.

Importantly, the added decremental targets reduced thresholds in *both* cases: when the decrement was in the same eye as the increment (ipsilateral) *and* when it was in the opposite eye (contralateral). This can be explained by two factors: (a) the presence of bidirectional (flickering) pedestals in both eyes and (b) the idea that binocular summation follows monocular opponency. Again, put simply, an upward increment in the left eye creates an opponent signal $(U - D) = (c + \delta c) - c = \delta c$ for the left eye, and a downward decrement in the right eye creates an opponent signal $(U - D) = c - (c - \delta c) = \delta c$ for the right eye. Binocular summation then renders a combined signal $2\delta c$, as before, even though the opponency itself precedes binocular summation. We examine this more formally below.

Binocular summation in motion

Our models assumed that motion detection and direction discrimination depend on responses from binocular processing. This supports the notion that the later stages of motion sensing are binocular. If there were separate monocular motion sensors for each eye without binocular summation, then the binocular advantage should not exceed what we expect from probability summation. However, the binocular summation ratios for motion detection without a pedestal (1.71, 1.80, and 1.82 for JB, GM, and PCH) were much higher than the values typically expected from probability summation (about 1.2). Rose (1978) found that binocular contrast sensitivity was twice as high as monocular sensitivity when gratings were flickering at 3.5 Hz. Arditi et al. (1981) examined the effects of

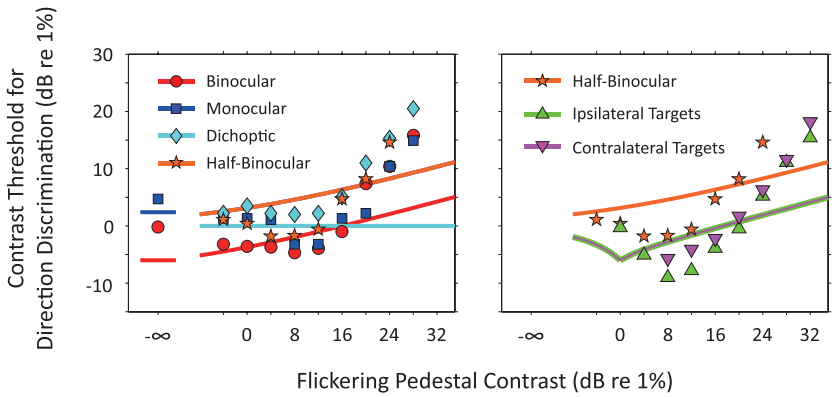

Figure 8. Fitting an alternative model in which the divisive gain control was removed from the twin summation model. The fits to mean data are shown here. There are substantial deviations between experimental thresholds (symbols) and model fits (curves).

spatial frequency on binocular and monocular detection of motion. The binocular summation ratios were nearly two for 0.6 c/° but about 1.6 for 9.6 c/°. Because the spatial frequency of our stimuli (1 c/°) was between these two, the present results are consistent with previous findings.

Relationship to divisive gain control and motion contrast

The TvC functions had a dipper shape for all conditions except the dichoptic presentation. The divisive inhibition is required to account for such a dipper function. If the divisive gain control is removed from the present models, the fits deviate enormously from the threshold data. As an example of this kind of alternative model, motion response, M_i, can be calculated as

$$M_U = (hwr\{E^m_{UL} - E^m_{DL}\} + hwr\{E^m_{UR} - E^m_{DR}\})^p \quad (4a)$$

$$M_D = (hwr\{E^m_{DL} - E^m_{UL}\} + hwr\{E^m_{DR} - E^m_{UR}\})^p. \quad (4b)$$

This alternative twin summation model failed to fit the data (RMSE was 6.19 dB for mean; Figure 8), and its failure supports the notion that the encoding of visual motion includes the divisive gain control.

Georgeson and Scott-Samuel (1999) found that *motion contrast* $(E_U - E_D)/(E_U + E_D)$ was a better predictor of direction discrimination than *opponent energy* $(E_U - E_D)$ proposed by Adelson and Bergen (1985). Because motion contrast incorporates both motion opponency and divisive gain control, our models do not contradict the concept of motion contrast. Actually, the present model (*TS1*) is similar to motion contrast except that we introduce half-wave rectification of the opponent signals followed by summation across the two eyes.

Model TS2: Opponency could be binocular?

One could argue that motion opponency might be binocular rather than purely monocular. To test this possibility, we fitted an alternative version of the twin-summation model (dubbed *TS2*) in which motion opponency takes place after binocular summation. The motion response of the twin-summation model (Equation 1) was re-expressed as

$$M_i = \frac{(E^m_{iL} + E^m_{iR})^p}{(I^n_{iL} + I^n_{iR})^q + z}, \quad (5)$$

where direction i is U or D. We assume that the binocular motion responses are subjected to motion opponency. The mechanism response R is given by

$$R = M_U - M_D.$$

Direction will be reliably discriminated when the mechanism response R is higher or lower than zero by a constant value. Here, R equals 1 or -1 at the threshold.

This model with late binocular opponency (*TS2*) was fitted to the group mean data of Figure 3, and the RMSE (1.37 dB) was almost identical to that for the early, monocular opponency model (1.36 dB; see Table A1). The present experiment alone therefore does not reveal whether motion opponency occurs before or after the binocular integration of monocular signals. We aim to resolve this ambiguity below (see *Monocular opponency*).

Is opponency a sensory process or a decision strategy?

Because the late opponency model fits our data well, we must consider another interpretation of that

idea: that motion opponency operates at a decision stage rather than as a sensory process. Suppose that observers had separate upward and downward signals (M_U, M_D) available without sensory opponency. Both mechanisms are active in a given trial, driven by the counterphase flickering pedestal, and so to make a decision about motion direction, the observer *must* compare the upward and downward motion signals and choose the larger. Such a comparison at the decision stage yields a model that is functionally identical to late, binocular opponency (Equation 7). Nevertheless, there are other arguments in favor of the sensory opponent mechanism. With both directional channels active and no opponency, we should expect the counterphase grating to look like two opposite transparent motions, and the lack of such transparency has long been argued as evidence for opponency. According to Qian et al. (1994a), observers perceived transparent motion only when stimuli contain locally unbalanced motion signals, suggesting that motion opponency is a spatially localized operator. It is also well known that, after viewing a motion stimulus, a stationary stimulus appears drifting in the opposite direction (motion aftereffect). Taking these findings together, it seems reasonable to conclude that motion opponency is a sensory process.

Monocular opponency

We saw previously that our results strongly implicate opponency but are consistent with either monocular or binocular opponency. To resolve this ambiguity, we applied our models to results obtained by Gorea et al. (2001). They tested the case in which pedestals had opposite directions in the two eyes, and this revealed a lack of dichoptic opponency. Their key finding was that performance (d') in detecting a contrast increment in one direction combined with a contrast decrement in the opposite direction ("inc/dec") was two to three times better than detecting the increment alone. But this strong signature of opponency disappeared when the two motion directions were seen by opposite eyes; inc/dec performance was then similar to that for the increment alone. Gorea et al. argued in favor of monocular opponency followed by direction-specific binocular summation but did not support their verbal argument with quantitative modeling. We therefore applied our models (*TS1*, *TS2*) to their results (as described in Appendix 1). Five out of six parameters were fixed from the fits to our data, and with just one free parameter, we found that monocular opponency (*TS1*) was strongly supported (i.e., it predicted both the advantage of opponency and its failure in dichoptic viewing). But the model fit was much less good when binocular opponency (*TS2*) was assumed instead (see Appendix 1 for details). We therefore conclude that the balance of evidence favors early, monocular opponency followed by direction-selective binocular summation (Gorea et al., 2001).

An extended model for motion and flicker

To account for other findings of Gorea et al., (2001) we devised two optional extensions to the *TS1* model to incorporate the possibility of (a) nondirectional flicker channels and (b) monocular channels. The inclusion of nondirectional flicker channels was also proposed by Wilson (1985) and Gorea et al. (2001). These extensions did not increase the number of model parameters, and five out of six parameters were again fixed in advance by fitting to our own data (Figure 3). We show in Appendix 1 that including flicker channels as well as motion-opponent channels gave an excellent quantitative account of the Gorea et al. data. The flicker channels played a key role in detecting contrast change in a drift-balanced condition in which there was no net motion. The monocular channels played little part for this data set but may play a larger role when a larger range of conditions is considered (Georgeson, Wallis, Meese, & Baker, 2016). We recognize that the data supporting this extended model are as yet very limited, but now that the theoretical structure is developed, a way forward to future experimental tests is clear.

Noise sources

Solomon, Chubb, John, and Morgan (2005) reported that psychophysical characteristics for direction discrimination at very low contrasts are inconsistent with a late-noise Reichardt model, in which noise is added only at the very end of processing. Based on this finding, Solomon et al. suggested that the early noise added to the output from spatial filters is also required to account for the psychometric functions. Although there must be many sources of noise in visual processing, those noises are simplified as the late noise in the present models, and this was sufficient to explain the threshold data. However, as pointed out by Solomon et al., we need to consider the nature of noise in more detail to account for how accuracy changes as a function of stimulus intensity. Indeed, we were surprised to find that the *TS1* model parameters derived from our flickering pedestal data produced implausible predictions for the contrast discrimination of drifting gratings. We

describe this anomaly in Appendix 2 and show that it can be fully resolved by supposing that an extra noise source (flicker-induced "motion noise") affects direction discrimination but not contrast discrimination. Further experimental work is needed to test the motion noise hypothesis.

Conclusion

This paper has addressed how motion sensing unfolds over monocular and binocular stages of processing. We constructed and compared computational models to explain direction discrimination thresholds under binocular, monocular, and dichoptic viewing. Converging evidence from two studies (ours and that of Gorea et al., 2001) suggests that motion opponency is most likely to be monocular, occurring before direction-specific binocular summation and before divisive, binocular gain control. Luminance-based motion perception depends on a chain of events in monocular and binocular pathways, and the ordering and functional description of those events is slowly becoming clearer.

Keywords: motion perception, motion opponency, binocular interactions, gain control, direction discrimination, computational models

Acknowledgments

This work was supported by a grant from the Japan Society for Promotion of Science to G. M.

Commercial relationships: none.
Corresponding author: Goro Maehara.
Email: goro@kanagawa-u.ac.jp.
Address: Department of Human Sciences, Kanagawa University, 3-27-1 Rokkakubashi, Kanagawa-ku, Yokohama, Kanagawa 221-8686, Japan.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A: Optics and Image Science, 2*(2), 284–299.

Anstis, S., & Duncan, K. (1983). Separate motion aftereffects from each eye and from both eyes. *Vision Research, 23*(2), 161–169.

Arditi, A. R., Anderson, P. A., & Movshon, J. A. (1981). Monocular and binocular detection of moving sinusoidal gratings. *Vision Research, 21*(3), 329–336.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review, 96*(1), 145–167.

Blake, R., & Boothroyd, K. (1985). The precedence of binocular fusion over binocular rivalry. *Perception and Psychophysics, 37*(2), 114–124.

Blake, R., & Fox, R. (1973). The psychophysical inquiry into binocular summation. *Perception and Psychophysics, 14*(1), 161–185.

Blake, R., Sloane, M., & Fox, R. (1981). Further developments in binocular summation. *Perception and Psychophysics, 30*(3), 266–276.

Carney, T. (1997). Evidence for an early motion system which integrates information from the two eyes. *Vision Research, 37*(17), 2361–2368.

Derrington, A., & Cox, M. (1998). Temporal resolution of dichoptic and second-order motion mechanisms. *Vision Research, 38*(22), 3531–3539.

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 11*(6), 1710–1719.

Georgeson, M. A., & Scott-Samuel, N. E. (1999). Motion contrast: A new metric for direction discrimination. *Vision Research, 39*(26), 4393–4402.

Georgeson, M. A., & Shackleton, T. M. (1989). Monocular motion sensing, binocular motion perception. *Vision Research, 29*(11), 1511–1523.

Georgeson, M. A., Wallis, S., Meese, T. S., & Baker, D. H. (2016). Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision Research, 129*, 98–118.

Gorea, A., Conway, T. E., & Blake, R. (2001). Interocular interactions reveal the opponent structure of motion mechanisms. *Vision Research, 41*(4), 441–448.

Green, D. M., & Svets, J. A. (1966). *Signal detection theory and psychophysics*. Newport Beach, CA: Peninsula Publishing.

Hayashi, R., Nishida, S., Tolias, A., & Logothetis, N. K. (2007). A method for generating a "purely first-order" dichoptic motion stimulus. *Journal of Vision, 7*(8):7, 1–10, doi:10.1167/7.8.7. [PubMed] [Article]

Kelly, D. (1966). Frequency doubling in visual re-

sponses. *The Journal of the Optical Society of America, 56*(11), 1628–1633.

Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. London: Academic Press.

Kulikowski, J. (1971). Effect of eye movements on the contrast sensitivity of spatio-temporal patterns. *Vision Research, 11*(3), 261–273.

Legge, G. E. (1984a). Binocular contrast summation—I. Detection and discrimination. *Vision Research, 24*(4), 373–383.

Legge, G. E. (1984b). Binocular contrast summation—II. Quadratic summation. *Vision Research, 24*(4), 385–394.

Lehky, S. R. (1988). An astable multivibrator model of binocular rivalry. *Perception, 17*(2), 215–228.

Lu, Z.-L. & Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. *The Journal of the Optical Society of America A, 18*(9), 2331–2370.

Maehara, G., & Goryo, K. (2005). Binocular, monocular and dichoptic pattern masking. *Optical Review, 12*(2), 76–82.

Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision, 6*(11):7, 1224–1243, doi:10.1167/6.11.7. [PubMed] [Article]

Meier, L., & Carandini, M. (2002). Masking by fast gratings. *Journal of Vision, 2*(4):2, 293–301, doi:10.1167/2.4.2. [PubMed] [Article]

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. Retrieved from www.palamedestoolbox.org

Qian, N., & Andersen, R. A. (1994). Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *The Journal of Neuroscience, 14*(12), 7367–7380.

Qian, N., Andersen, R. A., & Adelson, E. H. (1994a). Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *The Journal of Neuroscience, 14*(12), 7357–7366.

Qian, N., Andersen, R. A., & Adelson, E. H. (1994b). Transparent motion perception as detection of unbalanced motion signals. III. Modeling. *The Journal of Neuroscience, 14*(12), 7381–7392.

Rose, D. (1978). Monocular versus binocular contrast thresholds for movement and pattern. *Perception, 7*(2), 195–200.

Shadlen, M., & Carney, T. (1986). Mechanisms of human motion perception revealed by a new cyclopean illusion. *Science, 232*(4746), 95–97.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research, 38*(5), 743–761.

Solomon, J. A., Chubb, C., John, A., & Morgan, M. (2005). Stimulus contrast and the Reichardt detector. *Vision Research, 45*(16), 2109–2117.

Stromeyer, C., Klein, S., Kronauer, R., & Madsen, J. (1984). Opponent-movement mechanisms in human vision. *The Journal of the Optical Society of America, A, 1*(8), 876–884.

Van Doorn, A., & Koenderink, J. (1982). Temporal properties of the visual detectability of moving spatial white noise. *Experimental Brain Research, 45*(1–2), 179–188.

Wilson, H. R. (1985). A model for direction selectivity in threshold motion perception. *Biological Cybernetics, 51*(4), 213–222.

Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences, USA, 100*(24), 14499–14503, doi:10.1073/pnas.2333622100.

Appendix 1

Detecting contrast change: Modeling the results of Gorea et al. (2001)

Gorea et al. (2001) measured detectability (d') for a variety of dichoptic and binocular contrast increments and decrements for moving sine wave gratings. Figure A1 represents the nine tested conditions that we consider here. For clarity and brevity, we shall refer to these test conditions as *t1* through *t9*. These conditions and the detection task were different from our experiment, and so we hope to converge on models that are consistent with both data sets, and reject those that are not.

Gorea et al. (2001) stimuli and methods

Pedestal components were moving sinusoidal gratings of 1 c/° at a 20-Hz drift rate, each with 40% contrast in a 9° × 9° field. Increment/decrement contrast was also fixed for each of two subjects (at 4%, 5%), and performance was measured as detectability d'. An unusual feature of the procedure was that Gorea et al. (2001) followed Stromeyer et al. (1984) in having the pedestal grating present continuously. A trial was then defined as a 200-ms period in which the contrast change either did or did not

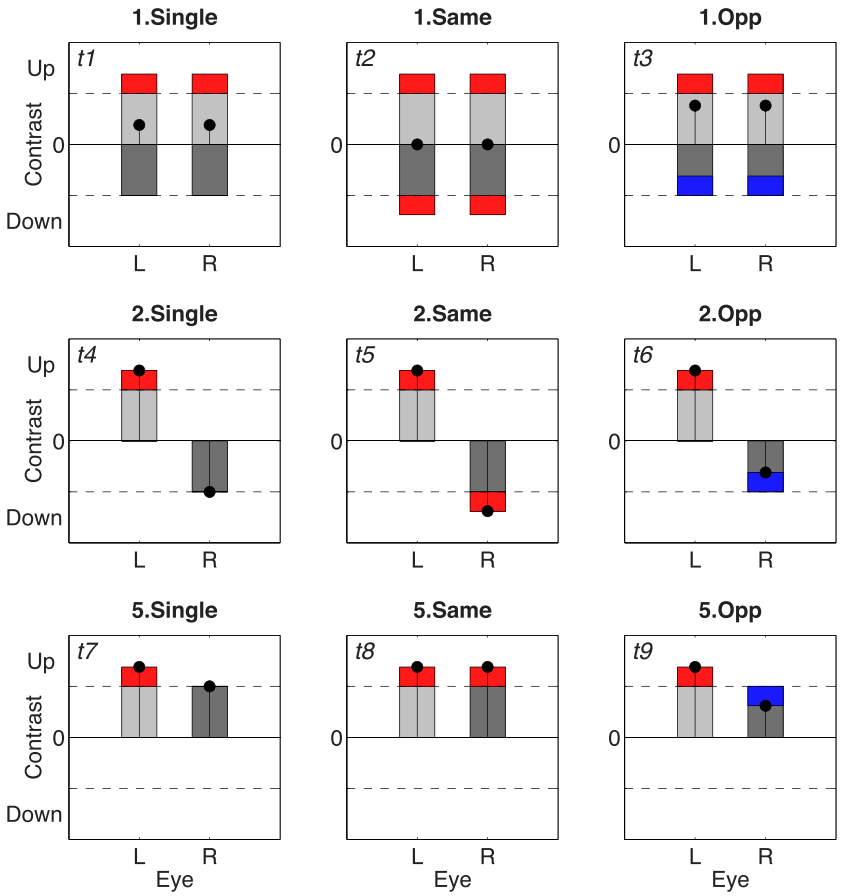

Figure A1. Like Figure 2B but representing nine of the stimulus conditions used by Gorea et al. (2001) in their study of contrast change detection for a variety of binocular and dichoptic moving gratings. Gray bars are components of the background (pedestal) grating, moving up or down. Red tabs: test contrast increment; blue tabs: test contrast decrement. Pedestal 1 (top row) had two *binocular* components: two equal-contrast, horizontal, binocular gratings (indicated here by light and dark gray), drifting up and down, respectively. Pedestal 2 (second row) had two monocular components, drifting in opposite directions in the two eyes. Pedestal 5 (third row) had two monocular components drifting in the same direction in the two eyes. (Pedestal conditions 3 and 4, not shown, were not relevant to the present paper.) Column 1 shows the "Single" condition, in which just one pedestal component (light gray) was incremented in contrast (red) or decremented (not shown). Column 2 shows the "Same sign" condition, in which both pedestal components were incremented (red) or both decremented (not shown). Column 3 shows the "Opposite sign" condition, in which one pedestal component (light gray) was incremented (red), and the other component (dark gray) was decremented (blue). Note that in Gorea et al.'s (2001) terminology, "opposite sign" refers to the direction of contrast changes, not to directions of motion.

occur. The task was thus a single-interval, yes/no detection task with 50% signal trials and 50% nonsignal trials (no contrast change) from which d' was derived from at least 800 trial responses in the standard way [as $z(Hits) - z(False\ alarms)$] for each condition *t1* through *t9*.

Experimental results

Two key findings can be seen in Figure A2. With a binocular, bidirectional pedestal (such as ours), d' values for a combined binocular increment and decrement (condition *t3*) were two to three times higher than for the increment alone (*t1*), analogous to our results. But when the pedestal components were separated between the eyes (*t4*, *t6*), there was little difference between the detectability of the increment/decrement (*t6*) and the increment alone (*t4*). This pair of results seems more consistent with monocular opponency, and we now test that idea in a model extended to cope with the conditions tested by Gorea et al. (2001).

The twin summation model: TS1

The model *TS1* has monocular opponency, followed by direction-specific binocular summation and binocular, direction-specific suppression (Equations 1a, b). The binocular channel responses (now indexed by B) are repeated here:

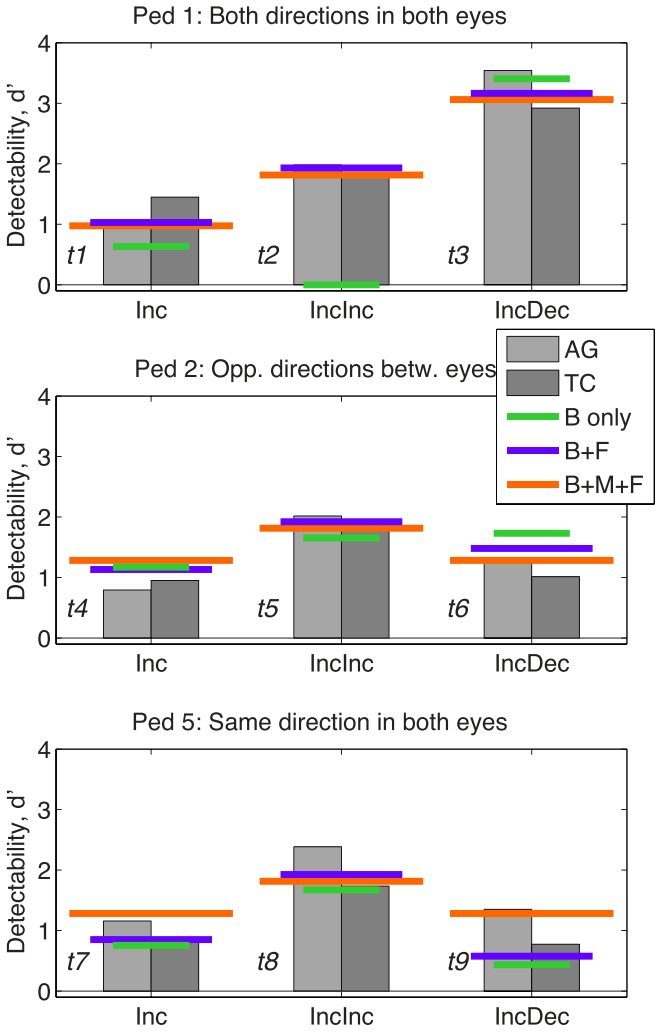

Figure A2. Detectability (d') for contrast change in the nine conditions of Gorea et al. (2001) (see Figure A1). Gray bars show the d' values for their two observers (AG, TC). Colored horizontal lines (short, medium, and long) mark the d' values that form three variants of the *TS1* model: green, short: binocular motion channels only; purple, medium: binocular motion and flicker channels; orange, long: binocular and monocular motion channels and nondirectional flicker channels. For all three models, opponency in the motion channels was at the monocular level (Equation A1). Note how the B-only model worked well for all conditions except *t2*, and B + F and B + M + F did well in all cases. Monocular opponency was a key feature in accounting for these results. Nondirectional flicker (F) channels were needed to capture information that was invisible to the binocular motion channels (B) alone.

$$M_{UB} = \frac{(hwr\{E^m_{UL} - E^m_{DL}\} + hwr\{E^m_{UR} - E^m_{DR}\})^p}{(I^n_{UL} + I^n_{UR})^q + z} \quad \text{(A1a)}$$

$$M_{DB} = \frac{(hwr\{E^m_{DL} - E^m_{UL}\} + hwr\{E^m_{DR} - E^m_{UR}\})^p}{(I^n_{DL} + I^n_{DR})^q + z}, \quad \text{(A1b)}$$

where $hwr\{x\}$ is half-wave rectification, i.e., $max(x,0)$.

Optional "flicker" channel: A nonopponent, nondirectional binocular channel

Not surprisingly, observers are sensitive to contrast change even when there is no net motion (*t2*). This is important because opponent channels are silent in response to drift-balanced flicker, and this implies that any general model should include either nonopponent or nondirectional mechanisms to account for this sensitivity. We therefore include the option of a binocular "flicker" channel (indexed by F) that has the same parameters as the motion channels but lacks opponency and responds to both directions in both eyes:

$$M_{FB} = \frac{(E^m_{UL} + E^m_{UR} + E^m_{DL} + E^m_{DR})^p}{(I^n_{UL} + I^n_{UR} + I^n_{DL} + I^n_{DR})^q + z}. \quad \text{(A2)}$$

To keep track of the different model variants, we denote the first model with binocular motion channels (Equation A1) as *TS1*(B), and when flicker channels are included, it becomes *TS1*(B + F).

Optional monocular channels

For completeness, we also explored a possible contribution from monocular channels. These are the same as the binocular ones above except that all input from the other eye is deleted. Hence, from Equations A1 and A2, we get monocular opponent motion channels and a monocular flicker channel for the left eye:

$$M_{UL} = \frac{(hwr\{E^m_{UL} - E^m_{DL}\})^p}{(I^n_{UL})^q + z} \quad \text{(A3a)}$$

$$M_{DL} = \frac{(hwr\{E^m_{DL} - E^m_{UL}\})^p}{(I^n_{DL})^q + z} \quad \text{(A3b)}$$

$$M_{FL} = \frac{(E^m_{UL} + E^m_{DL})^p}{(I^n_{UL} + I^n_{DL})^q + z} \quad \text{(A4)}$$

and similarly for right-eye channels M_{UR}, M_{DR}, and M_{FR}.

The max(L,B,R) operator

In the experiment of Gorea et al. (2001), the *test* signal was a brief, abrupt change in the ongoing pedestal. It is reasonable to suppose that each channel senses that temporal change, expressed as

Model:	*TS1*	*TS2*	*TS3*
Opponency	Early, monoc	Late, binoc	Early, monoc
Noise	Late, fixed	Late, fixed	Late, varies with contrast
Equation	Equation A1	Equations 6 and 7	Equations A1 and A9
S_E *(fixed)*	100	100	100
S_I	48.215	43.799	*48.215*
m	1.664	1.580	*1.664*
n	1.969	1.575	*1.969*
p	2.567	1.386	*2.567*
q	2.418	1.237	*2.060*
z	3.886	1.976	*3.886*
k	-	-	0.341
t	-	-	0.476
m.p − n.q	−0.50	0.24	0.21
SSE	97.49	99.96	101.5
RMSE, dB	1.356	1.373	1.384

Table A1. Parameters obtained from fitting two "binocular channel only" models to our direction discrimination data (group mean; main text, Figure 3), using either early, monocular opponency (*TS1*, Equation A1) or late binocular opponency (*TS2*, Equations 6 and 7 of the main text). *Notes*: Model *TS3* is an elaboration of *TS1*, discussed in Appendix 2.

$$M'_{U\mathrm{B}} = M_{UB}(test + pedestal) - M_{UB}(pedestalonly) \quad \text{(A5)}$$

and similarly for all nine combinations of directions *U,D,F* with ocularities *L,B,R*. Georgeson et al. (2016) introduced a scheme—very successful in the context of binocular and dichoptic contrast discriminations—that we followed here. We reduce the multiplicity of signals by taking the *max* over the monocular and binocular channels (although this feature plays no part when monocular channels are excluded). Thus,

$$R_U = \max\left\{M'_{U\mathrm{L}}, M'_{U\mathrm{B}}, M'_{U\mathrm{R}}\right\} \quad \text{(A6)}$$

and the corresponding sensitivity (d') for this channel will be

$$d'_U = R_U/\sigma, \quad \text{(A7)}$$

where σ is the standard deviation of R_U, and similarly for d'_D, d'_F In the present model $\sigma = 1$.

Decision-level processes

For a given stimulus configuration, the sensitivity (d') to contrast change will in general be different for the three responses R_U, R_D, R_F. But if the observer is able to use these three cues independently and efficiently, then the observed sensitivity d'_{OBS} may be predicted from the ideal observer, whose performance is given by the quadratic sum of the discriminabilities for each signal (Green & Swets, 1966):

$$d'_{OBS} = \sqrt{d'^2_U + d'^2_D + d'^2_F}. \quad \text{(A8)}$$

These modifications introduce a more complex architecture to the *TS1* model, but the number of free parameters is unchanged. We think the more complex structure is plausible and successful so far on this limited data set.

Modeling the data

We can define a variety of models within this scheme (Equations A1 through A8) simply by including or excluding some of the channels. For example, the simplest version, *TS1*(B), has only the binocular channels (with motion opponency at the monocular input level; Equation A1). The monocular and flicker channel responses (A2, A3, A4) were set to zero. For the *TS1*(B + F) model, both the motion and flicker channels (Equations A1 and A2) were switched on, and for the *TS1*(B + M + F) model, the monocular channels (Equations A3 and A4) were enabled as well.

These models were fitted to the data of Gorea et al. (2001) in two stages. First, we fitted the "B only" model to our present data (group mean, Figure 3 in the main text). This allowed us to hold fixed five of the six parameters via this independent dataset (Table A1). Then we derived predictions for the Gorea et al. data from the B, B + F, and B + M + F models and found that a relatively small adjustment of just one parameter (q; via Matlab's *fminsearch* as usual) was sufficient. The value of parameter q decreased from 2.4 (Table A1) to about 2.06 (Table A2). The surprisingly strong implications of this small parameter change are discussed in Appendix 2.

	Model variant				
Condition	B only	B + M	B + F	B + M + F	Experiment *d′*
t1	0.63	0.64	1.03	0.98	1.23
t2	0.00	0.00	1.93	1.82	1.91
t3	3.41	3.44	3.17	3.06	3.23
t4	1.17	1.20	1.13	1.28	0.87
t5	1.65	1.69	1.92	1.81	1.92
t6	1.73	1.20	1.48	1.28	1.17
t7	0.75	1.20	0.85	1.28	0.99
t8	1.67	1.71	1.93	1.81	2.06
t9	0.44	1.20	0.58	1.28	1.06
	5.130	4.398	0.485	0.493	*SSE*
	9	9	9	9	*N*
	-0.116	0.043	0.894	0.893	R^2
	2.048	2.047	2.062	2.066	*q*, fitted

Table A2. Summary of twin summation (*TS1*) model, showing *d′* values from model fits to the experimental data of Gorea et al. (2001) (*d′* values, mean of 2 Ss). *Notes*: Last four rows are goodness of fit statistics and the fitted value of parameter *q*. The other five parameters were fixed from Table A1, Equation A1.

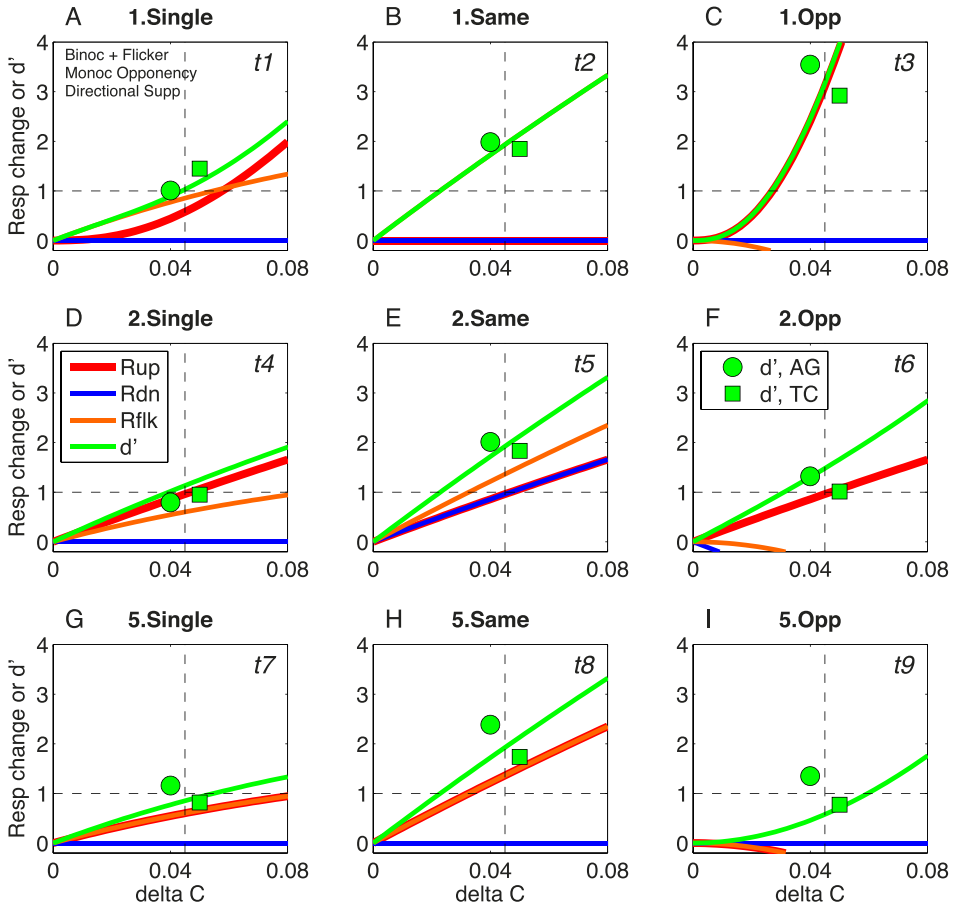

Figure A3. *TS1* model with binocular motion and flicker channels, *TS1*(B + F), showing how responses from different mechanisms contribute to performance. Curves show the response to contrast change (ΔC) in conditions *t1* to *t9* (panels A to I; cf. Figure A1) for the upward channel (R_U, red), the downward channel (R_D, blue), and the flicker channel (R_F, orange) along with the *d′* values (green curve) predicted by efficient use of all three cues (Equation A8). Symbols show data from two observers (Gorea et al., 2001) close to the predicted curve (green). In *t2*, the response is carried entirely by the flicker channel (orange, but hidden behind green).

Results

Let us first consider the model *TS1*(B) that has binocular motion channels only. Figure A2 shows the d' values for two observers (gray bars) along with the predictions of the "B only" model (green lines). The fit for eight of the nine conditions was good or fairly good; in particular, this model with monocular opponency explains why performance was much higher for condition *t3* than *t6*. In both cases, gratings drifting in opposite directions are incremented and decremented respectively (Figure A1). The difference in outcome arises because, with monocular opponency, "Up" increments and "Down" decrements reinforce each other to increase the opponent response M_{UB} when they are in the same eye (*t3*) but not when they are in opposite eyes (*t6*). Both results are well predicted by the *TS1*(B) model. But, as expected, this model incorrectly predicts no sensitivity at all for condition *t2* because monocular (or binocular) opponency yields no response to the drift-balanced motion components (Figure A3).

The next step was therefore to add the flicker channels and refit the model, adjusting only q. Figure A2 ("B + F" model; purple lines) shows that the fit for *t2* was now excellent, and the fit for all other conditions remained as good or better than before. Overall goodness-of-fit was high ($R^2 = 0.894$, RMSE = 0.232 d' units).

The final step was to add the monocular channels ("B + M + F"). Figure A2 (orange lines) shows that all nine conditions again fit well but with no improvement in the fit ($R^2 = 0.893$, RMSE = 0.234 d' units).

We also tested, in a similar way, the viability of late opponency, located after binocular summation (Equations 1 and 2, incorporated into Equations A1 through A8). We'll call this the *TS2* model. Five of the six parameters were fixed from the fit to our own data (Table A1, center column), and q was again adjusted for a least-squares fit. Unlike the *TS1* model, we found no version of this late opponency model (*TS2*) that fit well. For the four variants (B, B + M, B + F, B + M + F), the R^2 values were unimpressive: −1.312, −0.293, 0.278, and 0.365, respectively. We also tried the same approach, but using the same fixed parameters as *TS1*. The R^2 values were −0.971, −0.131, 0.682, and 0.713, somewhat improved for the last two (B + F, B + M + F) but markedly poorer than for *TS1*. In short, even though it fit our own data (Figure 6), we were unable to find a good fit of the late opponency model (*TS2*) to the Gorea et al. (2001) data set without resorting to a larger number of free parameters with which, with only nine data points, the danger of overfitting was severe. By contrast, the fit of *TS1* to both data sets was excellent but with one change in parameter value whose implications are discussed in Appendix 2.

Summary

Applying model *TS1* to the data set of Gorea et al. (2001), we found that the binocular channels with monocular opponency gave a good account of all the conditions in which a motion signal was present and that the inclusion of a nondirectional flicker channel with no extra free parameters added the necessary sensitivity to contrast change in a condition in which motion energy was balanced. Monocular channels were not necessary to explain performance for this experimental data set.

Appendix 2

Resolution of an anomaly

The twin summation model (*TS1*) fitted our data well (Figures 6 and 7), but further exploration of its properties revealed an inconsistency that we now describe, then attempt to resolve.

Model TS1 creates an apparent paradox

To understand the shape of a TvC curve, we need to understand how responses to the pedestal and to the added target are related to contrast. In Figure A4B, red filled symbols represent model responses to pedestal contrasts for a drifting grating, but because of opponency, the response to counterphase flickering pedestals (green symbols) is zero. Thick curve segments projecting from each pedestal point are responses to increasing contrast increments for a drifting target component. The upper tip of each curve segment represents the just-detectable contrast increment. Corresponding TvC curves are shown in panel A.

Responses to simple contrast increments (δc) of a *drifting* pedestal grating are especially diagnostic. When the effective exponent of excitation ($m.p$) is greater than that of the suppressive term ($n.q$), then the response increases monotonically with contrast in a compressive fashion if $m.p$ exceeds $n.q$ by less than one (solid red curve in B; $m.p - n.q = 0.21$). The corresponding TvC function shows a characteristic dipper shape (red curve, A). However, if $m.p - n.q = 0$, the contrast response saturates, and if $m.p - n.q < 0$, *the* response declines markedly at higher contrasts (two dashed red curves, B; for the lower curve $m.p - n.q = -0.5$, from Table A1, *TS1*). In such cases, contrast discrimination would be impossible in the saturated region and implausibly reversed in the declining region (a contrast increment produces response decrement). Experiments on contrast discrimination for drifting gratings have revealed no such catastrophes and instead showed conventional

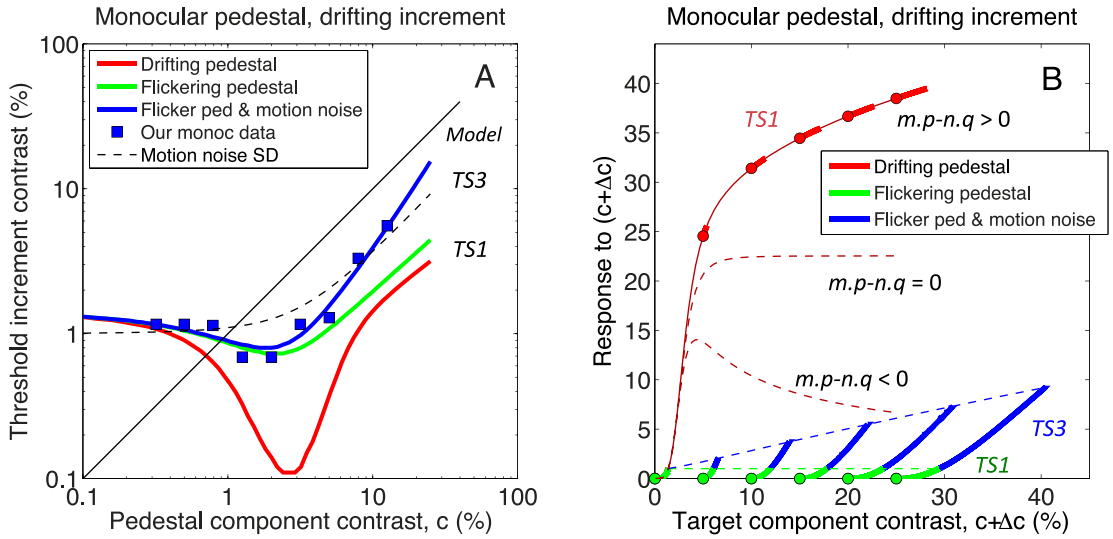

Figure A4. How the detectability of contrast increments (A) is related to the underlying responses (B) of two versions of the twin-summation model (*TS1, TS3*). *TS3* is the same as *TS1* but with the assumption of contrast-dependent, flicker-induced "motion noise" that compromises motion direction discrimination but not contrast discrimination (see Appendix 2). Dashed curve (in panel A) plots the standard deviation of the proposed motion noise as a function of contrast. Filled symbols (in panel B) represent responses to pedestal contrasts. Because of opponency, the response to counterphase flickering pedestals is zero. Thick curve segments projecting from each pedestal point are responses to increasing contrast increments for a drifting target component. The upper tip of each curve segment represents the just-detectable contrast increment. Dashed blue curve illustrates the growth of noise with contrast in model *TS3,* compared with constant noise in *TS1* (green dashed curve).

dipper-shaped TvC curves (Meier & Carandini, 2002), such as those well known for stationary gratings. Hence we can be sure that to account for contrast discrimination of simple drifting gratings, the *TS1* model should have $m.p - n.q > 0$. Indeed, our fitting of *TS1* to the Gorea et al. (2001) data set (Appendix 1, Table A2) gave $q = 2.06$, yielding $m.p - n.q = 0.21$. But to fit the rather steep masking curves seen in our own experiment with slopes ≥ 1, consistently required $m.p - n.q < 0$ (Table 1) with an average $m.p - n.q = -0.5$. And, as we have just seen, this leads to thoroughly implausible predictions about contrast discrimination. Some other factor may therefore be at work to make the masking curves with flickering pedestals steeper than they otherwise would be. One interesting possibility is that the limiting noise in our task (direction discrimination) might increase with contrast, leading the masking curves to be steeper as described next. If this factor is ignored, then q has to rise instead, and $m.p - n.q$ goes negative, leading to the inconsistency just described.

Model TS3: "Motion noise" induced by flicker is added to TS1

Flickering gratings in the spatiotemporal frequency range that we used can appear to jitter, move, or oscillate (Kelly, 1966, his figure 5; Kulikowski, 1971, his figure 3a) in a way that might affect direction discrimination but not contrast discrimination. We therefore propose that the motion task may be compromised by some form of motion noise induced by the flickering pedestal and that this noise increases with contrast. We note that the *product* of the upward and downward contrasts (or excitatory signals, E, in a given eye) represents the degree to which the pedestal is flickering rather than drifting. The product is zero for a drifting grating (because contrast in the other direction is zero), rising to E^2 for the flickering grating (because both signals have the same value E). This product therefore reflects both the "flickeriness" of the grating and its contrast and may be a useful index of the proposed motion noise. Computing this product for each eye, then summing them, we define the standard deviation σ_m of the motion noise to be a power function of that sum:

$$\sigma_m = k(E_{UL}E_{DL} + E_{UR}E_{DR})^t \quad \text{(A9)}$$

with two free parameters k and t. Assuming statistical independence, we can sum the variances of the unit variance internal noise and the flicker-induced motion noise to get the standard deviation σ_c of the combined noise:

$$\sigma_c = \sqrt{1 + \sigma_m^2}. \quad \text{(A10)}$$

Note that at zero contrast, $\sigma_c = 1$, as in model *TS1*. We assume for consistency with *TS1* that the threshold for direction discrimination is reached when $d' = 1$ for

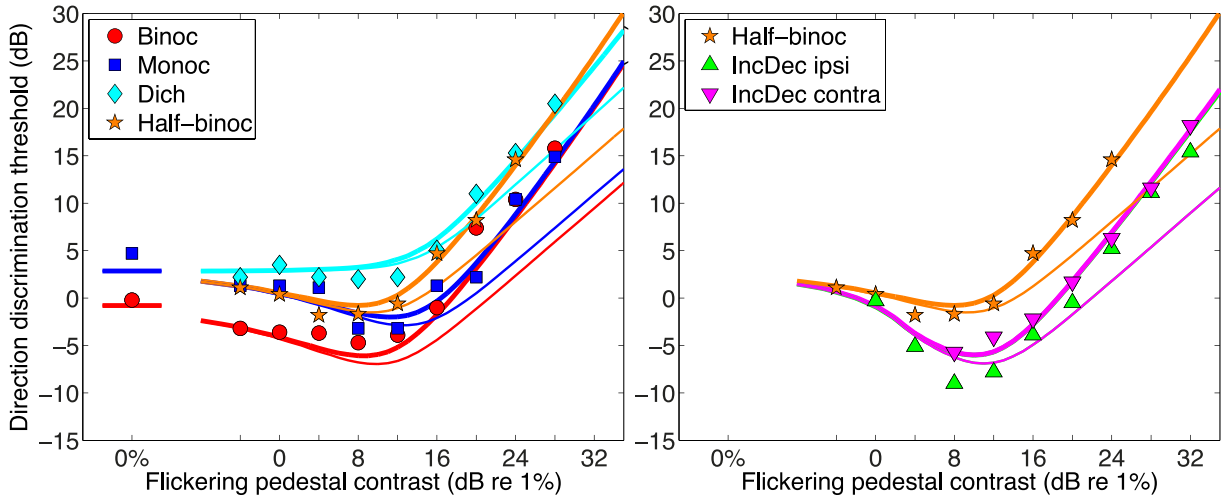

Figure A5. Like Figure 7 except that the thin curves show predictions of the *TS1* model with the parameter values used to fit the Gorea et al. (2001) data (Figure A2); in particular, $q = 2.06$ and other parameters as in Figure 7. Note the relatively shallow masking curves as expected from this choice of q. Thick curves show model *TS3*: These steeper, well-fitting curves were obtained when the contrast-dependent "motion noise" was added to the model (to convert *TS1* into *TS3*).

the target channel (e.g., upward):

$$M_U/\sigma_c = 1. \quad \text{(A11)}$$

In short, model *TS3* is a simple extension of *TS1* in which constant noise is replaced by contrast-dependent noise σ_c, which includes the motion noise σ_m. The motion noise falls to zero for a drifting grating, and Equation A9 allows it to be calculated automatically in all cases. For a monocular counterphase grating with component contrasts c, Equation A9 simplifies to $\sigma_m = k(S_E c)2t$. We call it *motion noise* because we assume at present that it does not affect contrast discrimination. This kept our analysis of Gorea et al. (2001) (Appendix 1) unchanged and fixed six of the eight parameters in *TS3* (italicized in Table A1). *TS3* was then fitted to the data of Figure 3 by adjusting only the new parameters k, t. Thick curves in Figure A5 show that *TS3* fits the group mean data just as well as *TS1* did (RMSE = 1.38 dB for *TS3*, 1.36 dB for *TS1*). But it has the clear advantage that with no change in parameter values it also predicts a plausible dipper function for contrast discrimination (red curve in Figure A4A).

Because the exponent t emerged as close to 0.5 (Table A1), hence $2t$ close to 1, this implies that motion noise in our experiment rose almost in directional proportion to contrast (blue dashed line in Figure A4B). Model *TS3* offers an interpretation of *TS1* and resolves the apparent paradox that *TS1* otherwise creates (above). In this view, the rise in motion masking with flickering pedestal contrast is partly due to divisive suppression, which reduces contrast gain as it does for a contrast discrimination task. But motion masking rises more steeply because it also includes a rise in noise that is specific to motion discrimination.