Center-surround motion interaction between low and high spatial frequencies under binocular and dichoptic viewing

Omar Bachtoula Department of Experimental Psychology, Universidad Complutense de Madrid, Madrid, Spain

Ignacio Serrano-Pedraza Department of Experimental Psychology, Universidad Complutense de Madrid, Madrid, Spain

Motion discrimination of a stimulus that contains fine features is impaired when static coarser features are added to it. Previous findings have shown that this cross-scale motion interaction occurs under dichoptic presentation, where both components are spatially overlapped. Here, we used a center-surround spatial configuration where both components do not spatially overlap. We measured the strength of this motion interaction by assessing the cancellation speeds (i.e., the speed needed to cancel out the motion discrimination impairment) for different combinations of spatial frequencies, temporal frequencies, contrasts, durations, and under binocular and dichoptic presentations. The experiments revealed that cancellation speed is bandpass tuned to spatial frequency, increases with temporal frequency up to 12 Hz before slightly decreasing, and intensifies with contrast before stabilizing at higher levels. We found similar patterns of results for both dichoptic and binocular presentations, although the interaction was stronger in the binocular condition. These results confirm that this interaction mechanism can integrate fine and coarse scales when presented to different eyes, even when motion signals do not spatially overlap. Finally, we explain the differences between dichoptic and binocular cancellation speeds using a motion-sensing model that includes a cross-scale interaction stage. The model simulations suggest that an interocular gain control, followed by binocular summation and then by cross-scale interaction, accounts for the differences observed between binocular and dichoptic viewing.

Introduction

Natural scenes contain moving elements with features in different spatial and temporal scales. Previous research has shown that humans systematically report the opposite direction to the actual direction of motion of brief complex stimuli that combine static coarse and moving fine scales (Derrington, Fine, & Henning, 1993; Derrington & Henning, 1987; Henning & Derrington, 1988; Serrano-Pedraza, Goddard & Derrington, 2007; Serrano-Pedraza & Derrington, 2010 (see section "Interaction across different spatial scales" in Nishida 2011). The existence of this interaction between coarse and fine scales questions the classical notion of parallel processing that motion energy models propose (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Watson & Ahumada, 1985). These models define the early stages of visual motion processing, where the direction of motion is computed by simulating direction selective simple cells. They include localized and oriented motion sensors tuned to spatial and temporal frequency, and their output is independent from the activity of other motion sensors. Experiments that used masking and spatial summation support these properties of the motion sensors (Anderson & Burr, 1985; Anderson & Burr, 1987; Anderson & Burr, 1989; Anderson & Burr, 1991; Anderson, Burr, & Morrone, 1991; Levinson & Sekuler, 1975). Despite the success of the motion energy model explaining some perceptual phenomena, like the missing fundamental illusion, reverse phi, or apparent motion (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Watson & Ahumada, 1985), it fails to predict the result of the interaction that occurs when static coarse scales are added to moving fine scales. Serrano-Pedraza et al., (2007) proposed a model of motion sensing based on the motion energy model that includes a cross-scale interaction stage that computes a subtractive interaction between the outputs of the motion sensors tuned to fine and coarse scales. This model accounts for the systematic errors that humans make when judging the direction of motion of a short duration complex pattern comprising a static low frequency grating and a moving high spatial frequency (HF) grating (Luna & Serrano-Pedraza, 2018;

Citation: Bachtoula, O., & Serrano-Pedraza, I. (2025). Center-surround motion interaction between low and high spatial frequencies under binocular and dichoptic viewing. *Journal of Vision*, *25*(8):15, 1–18, https://doi.org/10.1167/jov.25.8.15.

https://doi.org/10.1167/jov.25.8.15 Received September 26, 2024; published July 17, 2025 ISSN 1534-7362

Luna & Serrano-Pedraza, 2020; Serrano-Pedraza et al., 2007; Serrano-Pedraza & Derrington, 2010).

Previous research has characterized this cross-scale motion interaction mechanism with several parameters. The effect has been reported for coarse scales lower than 1 c/deg and fine scales higher than 1.5 c/deg (Derrington & Henning, 1987; Henning & Derrington, 1988; Luna & Serrano-Pedraza, 2018; Luna & Serrano-Pedraza, 2020). The strength of the interaction increases as the stimulus presentation shortens (<100 ms) (Derrington & Henning, 1987; Henning & Derrington, 1988; Luna & Serrano-Pedraza, 2020; Serrano-Pedraza et al., 2007; Serrano-Pedraza & Derrington, 2010), but it is still present for longer durations (Serrano-Pedraza et al., 2007). Other studies found that the activity of this mechanism is tuned to the temporal frequency of the moving component (Derrington & Henning, 1987; Luna & Serrano-Pedraza, 2018), showing a bandpass shape with maximum intensity at approximately 6 to 12 Hz. Finally, the contrast level is also a relevant variable, because the strength of the motion mechanism depends on the relative contrast of the components. Specifically, the reversals in motion direction discrimination occur when the contrast ratio between the fine features and the coarser features falls between 0.8 and 4.0 (Serrano-Pedraza & Derrington, 2010).

Interestingly, using spatially overlapped components, it has been found that the strength of the interaction is decreased slightly in dichoptic presentations (presenting a static low-frequency grating in one eye and the moving high-frequency grating in the other eye) compared with monocular and binocular presentations (Derrington et al., 1993). The presence of the interaction in dichoptic presentations suggests that this inhibitory mechanism integrates motion signals from both eyes. However, we do not know whether the spatial arrangement of the fine and coarse scales involved in this motion illusion is relevant in dichoptic presentations. Henning and Derrington (1988) measured the strength of the illusory motion for different combinations of test and inducing spatial frequencies presented on different spatial positions. These studies used binocular presentations, but little is known about the strength of this interaction mechanism in dichoptic presentations when the different spatial features are presented on different retinal positions. According to those previous studies, we expect that the interaction between scales presented on different retinal positions will be similar for both dichoptic and binocular viewing conditions.

Here, using a cancellation-speed technique in a center-surround configuration, we explore the properties of the interaction between spatial scales under dichoptic and binocular presentations with the components of the stimulus presented on different retinal positions. In total, we have performed three experiments where the main objectives are to 1) measure the strength of the motion interaction between spatial frequencies when the components are presented to different retinal positions under binocular and dichoptic presentations and 2) characterize this inhibitory mechanism, with spatially separated components in both viewing conditions, testing different stimulus durations, spatial frequencies, temporal frequencies, and contrast levels.

Finally, we performed simulations using a motion-sensing model that includes a cross-scale interaction stage, based on the model by Serrano-Pedraza et al., (2007), but with interocular gain control and binocular summation stages preceding the cross-scale interaction. This model successfully predicted the observed difference in cancellation speed between the dichoptic and binocular viewing conditions.

Methods

Participants

Four human participants (2 females and 2 males) with experience in psychophysical experiments took part in all the experiments of the study, with ages in the range of 23 to 27 years old (25.5 ± 1.73 years old; mean ± SD). All participants had normal or corrected-to-normal vision. We assessed their spatial visual acuity with the SLOAN ETDRS 2000 letter series (Precision Vision, Woodstock, IL) on each eye and at two distances, 3 m, and 40 cm. The limit to take part in the experiments was 0.2 logarithm of the minimum angle of resolution (logMAR), so getting a visual acuity above that value would result in the removal from the study. Likewise, we determined their stereoscopic visual acuity with the Graded circle test (Randot Stereotest, Stereo Optical Company, Inc., Chicago, IL) at a distance of 40 cm. The averaged spatial visual acuity at 3 m was −0.1 ± 0.12 logMAR on the right eye and −0.1 ± 0.12 logMAR on the left eye. At 40 cm, the results were −0.08 ± 0.05 logMAR on the right eye and −0.05 ± 0.06 logMAR on the left eye. On the other hand, the averaged stereoscopic visual acuity with the Randot test was 25 ± 5.77 arcsec. We also ran a random-dot stereotest developed by our laboratory (Serrano-Pedraza, Manjunath, Osunkunle, Clarke, & Read, 2011). This stereotest uses random gaussian-dots (dot density of 23 dots/deg^2), and presents a flat three-dimensional (3D) circle (diameter of 8 deg) inside a flat square (size of 23.7 deg × 23.7 deg) at a distance from the observer of 150 cm. The 3D circle appears randomly in front of or behind the screen, and the task is to indicate its position (e.g., close or far). To obtain the disparity thresholds we used a Bayesian staircase procedure with characteristics described by Serrano-Pedraza et al., (2020). For each participant,

we obtained three thresholds, and the average of all thresholds of all participants was 1.05 ± 0.23 $\log_{10}$(arcsec) (the geometric mean was 11.22 arcsec). The Ethics Committee of Universidad Complutense de Madrid approved the experimental procedures, and they comply with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Before taking part in these studies, all participants gave written informed consent, and we anonymized their identity with a random alphanumeric code.

Apparatus

We ran the experiments on a Linux-operated computer (Ubuntu 20.04 LTS, 64 bits) with an Intel Core i7-9700K processor (8 cores, 3.6 GHz, 32 GB RAM) and an AMD Radeon RX 590 Fatboy graphics card (8 GB of GDDR5 memory). We created the stimuli and implemented the experiments in MATLAB (The MathWorks, Natick, MA) with the Psychtoolbox extension libraries (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Display-wise, a DataPixx Video I/O Hub (VPixx Technologies Inc., Saint-Bruno, Canada) controlled the output of the graphics card and sent it to a PROPixx projector (VPixx Technologies, Saint-Bruno, Quebec, Canada), that we used to display the experimental stimuli. The projector resolution was 1,920 × 1,080 pixels and the framerate was 120 Hz. This projector worked in RB3D mode, allowing us to present the stimuli at 120 Hz in each eye independently. The output of the projector passed through a DepthQ 3D polarizing filter (Lightspeed Design, Bellevue, WA) and was projected on a Stewart Filmscreen 150 rear projection screen (Stewart Filmscreen, Torrance, CA), where we presented the stimuli.

During the experiments, participants wore DepthQ passive polarized glasses (Lightspeed Design), allowing independent image presentation on each eye. The crosstalk of the light passing through the screen and the glasses ranged between 2% and 3%. The mean luminance of the screen through these glasses was 28.58 cd/m^2. We fixed the head position of the participants at 150 cm from the screen with a chin rest (UHCOTech HeadSpot, Houston, TX), and participants used a RESPONSEPixx button pad (VPixx Technologies) to respond.

Stimuli

The images of the stimuli had a resolution of 800 × 800 pixels and appeared on the center of the screen, where they measured 50.42 cm × 50.42 cm. They subtended 19.08 deg × 19.08 deg of visual angle (deg), resulting in 41.93 pixels per deg. The stimuli consisted of a circular central component surrounded by an annulus. With this structure, we could study induced motion, because the drift of the outer component in one direction produces an illusory percept on the static central component, that seems to move in the opposite direction. We created the central part by multiplying a vertical sinewave grating with a low-pass two-dimensional (2D) Butterworth spatial window of order 10 and a cut-off visual angle of 1 deg. Thus, the diameter of this central component was 2 deg. This part of the stimulus was the test, that is, the part in which the illusion occurs. In contrast, the annulus consisted of another vertical grating multiplied by a bandpass 2D Butterworth window of order 10, a central visual angle located at 2 deg and a bandwidth of 2 deg. This component of the stimulus was the inducer, the stimulus that produces the illusory motion on the test. Both gratings drifted horizontally (e.g., leftward [L] or rightward [R]) during the experiments within the spatial windows delimited by the Butterworth filters. The remainder of the screen outside of the stimulus structure had mean luminance. Figure 1a shows the spatial structure of an example stimulus, and Figure 1b contains the profile of its corresponding Butterworth windows (González & Wintz, 1987; see also Appendix A in Sierra-Vázquez, Serrano-Pedraza, & Luna, 2006).

The following equation describes the experimental stimuli:

$$f(x, y) = L_0 \times [1 + \text{Test} + \text{Inducer}], \quad (1)$$

where L_0 is the mean luminance of the screen, and

$$\text{Test} = m_T \times \cos[2\pi u_T (x - v_T t) + \varphi_T] \times \left[1 + \left(\frac{\sqrt{(x^2 + y^2)}}{\rho_T}\right)^{2r_T}\right]^{-1} \quad (2)$$

$$\text{Inducer} = m_I \times \cos[2\pi u_I (x - v_I t) + \varphi_I] \times \left[1 + \left(\frac{\sqrt{(x^2 + y^2)} - \rho_I}{B/2}\right)^{2r_I}\right]^{-1} \quad (3)$$

where m is the Michelson's contrast of the grating, u is the spatial frequency of the grating (c/deg), v is the drifting speed of the grating (deg/s), t is the time (s), φ is the phase of the grating (rad), ρ_I is the central visual angle in the bandpass Butterworth window ($\rho_I = 2$ deg), and ρ_T is the cut-off visual angle in the low-pass Butterworth window ($\rho_T = 1$ deg), B is the bandwidth of the bandpass Butterworth window ($B = 2$ deg), and r is the order of the Butterworth window ($r_T = r_I = 10$). Subindices I and T indicate whether the parameter corresponds to the inducer or the test, respectively.

Using our polarizing system (see Apparatus section), we presented the stimuli to the participants in two

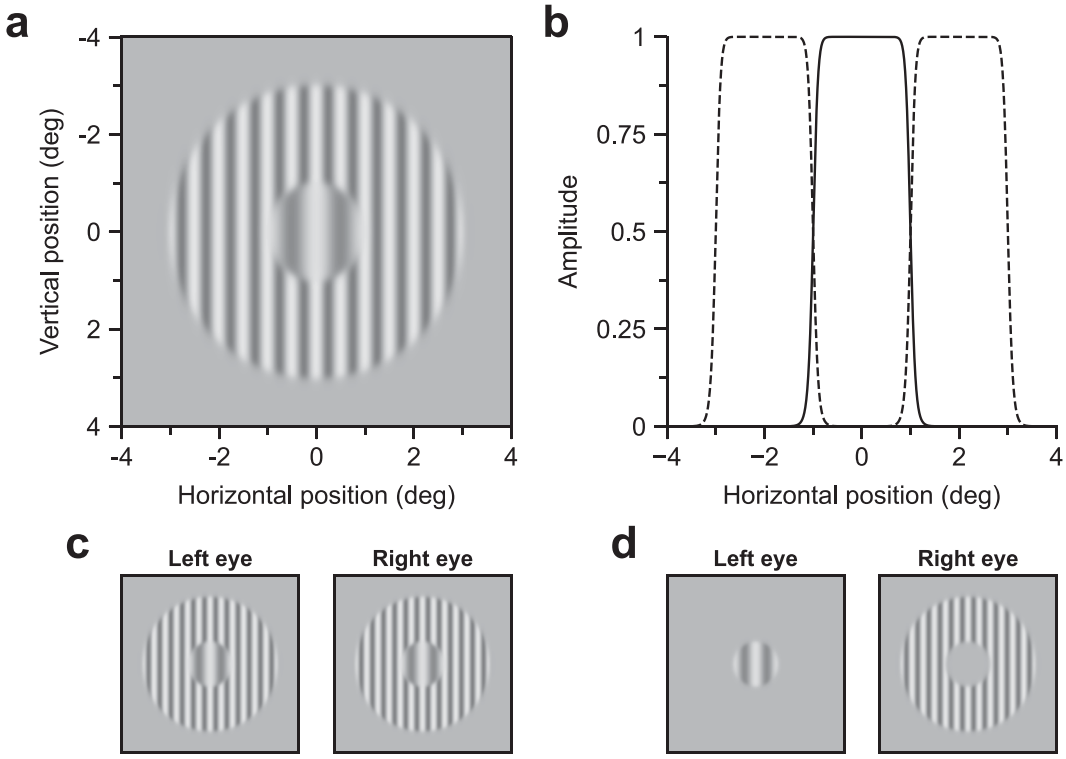

Figure 1. Center-surround structure of one of the experimental stimuli. (**a**) Image of the stimulus. The spatial frequency is 1 c/deg for the test (i.e., central region), and 2 c/deg for the inducer (i.e., surrounding annulus). Note that the contrast of both components is higher than the actual contrast in the experiments. (**b**) Amplitude profile of the Butterworth spatial windows used to create the test (black line) and the inducer (dashed line). (**c**) Example of the stimulus that each eye receives in the binocular condition. (**d**) Example of the stimulus that each eye receives in the dichoptic condition.

viewing conditions. In the binocular condition, both eyes received the whole stimulus simultaneously (Figure 1c), whereas, in the dichoptic condition, one eye received the test stimulus, and the other eye received the inducer (Figure 1d). Furthermore, to avoid the effect on the results of a possible systematic issue in one of the eyes, we tested the dichoptic conditions on both eyes.

In experiment 1, the test component had a Michelson's contrast of 0.1, 1 c/deg of spatial frequency, and drifted rightward or leftward with a speed determined by a Bayesian adaptive staircase (see Procedure). The inducer, in contrast, had a Michelson's contrast of 0.3 and one of seven possible spatial frequencies: 0 (no inducer, control condition), 0.5, 1.0, 2.0, 3.0, 4.0, or 6.0 c/deg, and drifted at a fixed speed of 4 deg/s. We tested two durations in this experiment: 50 and 100 ms, for binocular and dichoptic presentations. Thus, the total number of conditions for this experiment was 42 (7 spatial frequencies × 2 presentation durations for the binocular condition plus 7 spatial frequencies × 2 presentation durations × 2 eyes for the dichoptic condition).

In experiment 2, the test had the same parameters as in experiment 1. The inducer had a Michelson's contrast of 0.3, one of two possible spatial frequencies: 2 or 3 c/deg, and drifted with one of five temporal frequencies: 2, 6, 12, 16, or 24 Hz. We only tested the duration of 50 ms for binocular and dichoptic presentations, getting a total of 30 conditions (5 temporal frequencies × 2 spatial frequencies for the binocular condition plus 5 temporal frequencies × 2 spatial frequencies × 2 eyes for the dichoptic condition).

Last, in experiment 3, the central component had 1 c/deg of spatial frequency and three possible contrast levels: 0.1, 0.3, or 0.6. The inducer had a spatial frequency of 2 c/deg, one of five possible contrast levels: 0, 0.1, 0.3, 0.6, or 0.9, and drifted at 4 deg/s. The duration was always 50 ms and the viewing conditions were binocular and dichoptic. Therefore, we tested 45 conditions (3 test contrasts × 5 inducer contrasts for the binocular condition plus 3 test contrasts × 5 inducer contrasts × 2 eyes for the dichoptic condition).

All four participants completed the experiments. In experiment 1, we measured 126 staircases, that is, participants went through 3,780 trials. In experiment 2, there were 90 staircases, or 2,700 trials. Finally, in experiment 3, there were 135 staircases, adding up to 4,050 trials. In total, each participant completed 10,530 trials throughout the 117 conditions tested in all the

experiments, resulting in approximately 10 hours of experimentation per participant.

Procedure

All the experiments consisted of a single presentation forced choice procedure, where the task was to indicate the direction of motion of the test component (e.g., left or right). At the start of a given trial, a fixation cross, that subtended 0.49 deg × 0.49 deg, appeared on the center of the screen for 160 ms. After the cross had disappeared, we presented the stimulus. A Gaussian temporal function, centered and truncated in a window of 750 ms, defined its presentation throughout the trial, providing a gradual onset and fading of the contrast of the stimulus. The nominal duration was twice the standard deviation of this Gaussian function ($2 \times \sigma_t$), and the drifting direction of the inducer and the phase of both gratings (test and inducer) were random in every trial.

With the stimulus layout described in the Stimuli section, and under a particular selection of parameters for the test and inducer, the motion of the inducer in one direction originates an illusory motion percept on the test component when it is static. Despite not physically moving, it appears to drift in the opposite direction of the inducer. In the three experiments presented here, we measured the drifting speed of the test stimulus, in the opposite direction of the illusion, that cancelled the illusory percept (i.e., cancellation speed). To do that, we controlled the speed of the test component with a Bayesian adaptive staircase that increased the speed whenever the participant responded in the direction of the illusion and decreased it when the participant responded in the direction of the inducer.

To obtain the cancellation speed we used a Bayesian adaptive staircase (Treutwein, 1995) with characteristics from the ZEST procedure (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994). Each staircase consisted of 30 trials and started with a speed of 3 deg/s. The prior probability distribution function was uniform (Emerson, 1986), and the model function was a logistic function, with a spread of 1, a delta of 0.01, and lapse and guessing rates of 0.01 (see details of this type of Bayesian staircase in Serrano-Pedraza et al., 2020). The target probability for the staircase was 0.5, because we wanted to get the speed that produced chance performance, that is, the speed that cancelled the motion illusion. On a given trial, the speed presented was the mean of the posterior probability distribution, and the cancellation speed was the mean of the final probability distribution function. We divided each experiment into blocks of three randomly selected conditions. In each block of 90 trials, we randomly interleaved the trials of the three staircases (each with 30 trials) corresponding with the three conditions selected. After completing all the blocks, we had collected three cancellation speeds of each condition.

Results

Experiment 1: Effect of the spatial frequency of the inducer and the duration of the stimulus

In this experiment, we studied the effect of the spatial frequency of the inducer on the cancellation speed of the test, using two short durations (50 and 100 ms), and two viewing conditions (binocular and dichoptic). Figure 2 shows the results from the experiment. The panels show the cancellation speed as a function of the spatial frequency of the inducer for binocular (black dots) and dichoptic (red diamonds) viewing conditions. The spatial frequency of the test was always 1 c/deg. Positive cancellation speeds indicate that, to cancel the motion illusion, the test stimulus had to drift in the direction of the inducer, and, therefore, opposite to the motion illusion.

Focusing on the shortest duration, displayed in Figure 2a, both viewing conditions show a bandpass tuning of induced motion to the spatial frequency of the inducer. The cancellation speed increases with the spatial frequency up to the peak effect at 2 c/deg, from which it decreases as spatial frequency continues to increase. Comparing the peak of both types of viewing conditions, the cancelation speed for the binocular condition is 2.5 deg/s and for the dichoptic is 1.66 deg/s. That is, the cancellation speed is 1.51 times faster for the binocular than for the dichoptic viewing condition. Despite some spatial frequencies producing very similar results between viewing conditions, most of the points in the binocular presentations are very similar or above the dichoptic data (e.g., between 2 and 4 c/deg). Additionally, the strongest motion illusion occurs when the inducer has a higher spatial frequency than the test stimulus. Specifically, in the binocular condition, the cancellation speed for the inducer with 2, 3, and 4 c/deg was higher than when the spatial frequency of the inducer matched the test (1 c/deg). In the dichoptic presentation, the spatial frequency of 3 c/deg produces very similar results to the matching spatial frequency, and only 2 c/deg produces a stronger cancellation speed than the matching spatial frequency. These results reveal a strong interaction between different spatial frequencies when located on different retinal positions. More specifically, they show that the strongest influence of the inducer on the test occurs when the peripheral spatial frequency is higher than the one in the center.

Increasing the duration of the stimulus, as shown in Figure 2b, not only decreases the intensity of the illusory motion, but also shifts the tuning of

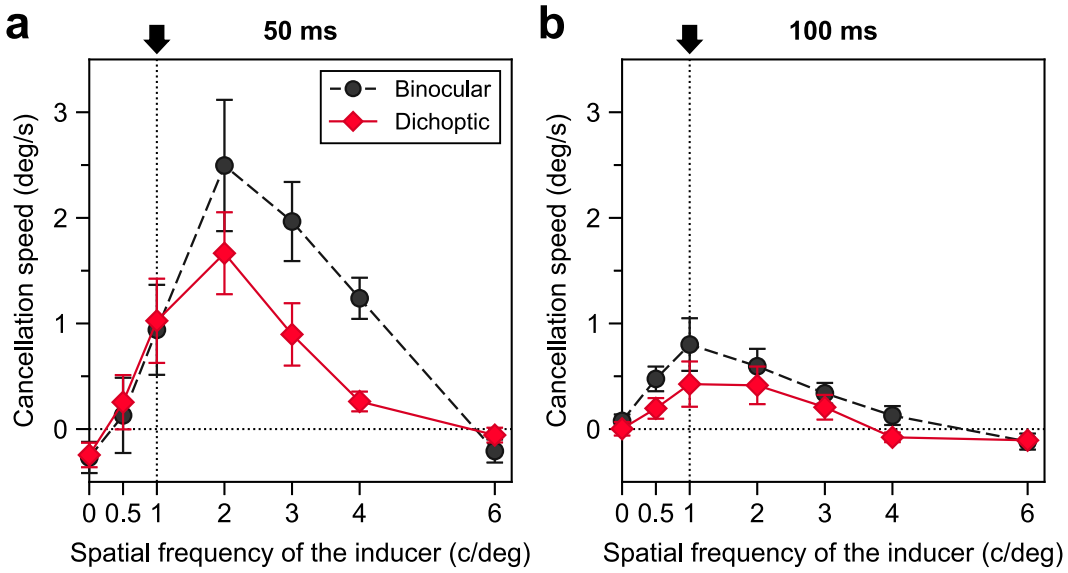

Figure 2. Results from experiment 1. The panels show the cancellation speed (mean ± standard error of the mean) as a function of the spatial frequency of the inducer for four observers. The arrows on top of the panels indicate the spatial frequency of the test stimulus (1 c/deg). Black dots show the results for the binocular viewing condition. Red diamonds show the results for the dichoptic viewing condition. (**a**) Results for the duration of 50 ms. (**b**) Results for the duration of 100 ms. All inducer stimuli drifted at a fixed speed of 4 deg/s.

the illusion to the spatial frequency of the inducer. Now, the higher cancellation speed is closer to the spatial frequency of the inducer that matches the test. Higher spatial frequencies, namely, 2 and 3 c/deg, still influence the perception of the test, but their effect is weaker than the matching spatial frequency. This implies that the intensity of the interaction between spatial frequencies depends on the duration of the stimulus and is more intense for shorter durations. Also, in agreement with the observations in the shortest duration, the results show that the illusory motion originated by the drift of the inducer was more intense for binocular than dichoptic presentations.

Our results suggest that the shorter the duration, the higher the spatial frequency that produces the maximum cancellation speed. This analysis is based on empirical data, and the values are limited by the discrete sampling of the inducer's spatial frequency. However, if we fit a smooth curve to our data, our conclusions could change, and the maximum might be similar for both durations. To test this hypothesis, we fitted a lognormal function with three free parameters to the data from Figure 2 (see fits in Appendix A). The parameter f_{max}, which indicates the spatial frequency at which the maximum occurs, supports our conclusions. The theoretical f_{max} for the shortest duration (50 ms) was 2.061 c/deg for the binocular condition and 1.648 c/deg for the dichoptic condition; and for the longest duration (100 ms), it was 1.088 c/deg for binocular and 1.257 c/deg for dichoptic. We can also analyze the ratio between the maxima (A parameter) of the binocular and dichoptic conditions. The ratios are similar to our previous analysis. For the short duration, the ratio was 2.579/1.801 = 1.432; and 0.838/0.514 = 1.63 for the longer duration. Interestingly, the bandwidth in octaves (full width at half height) of the fitted function was greater for the longer duration (2.375 octaves for binocular, 1.924 for dichoptic) than for the shorter duration (1.679 for binocular, 1.646 for dichoptic). Thus, although the strength of the cancellation speed increases at the shorter duration, the bandwidth is narrower compared with the longer duration.

Experiment 2: Effect of the temporal frequency of the inducer

In this experiment, we studied the effect of the temporal frequency of the inducer on the cancellation speed of the test. We tested two spatial frequencies for the inducer (2 and 3 c/deg) and two viewing conditions (binocular and dichoptic) for a duration of 50 ms. The spatial frequency of the test was always 1 c/deg. We selected the combination of 2 and 3 c/deg with 50 ms because they were the parameters in experiment 1 that produced the strongest effect. Figure 3 shows the cancellation speed as a function of the temporal frequency of the inducer for binocular (black dots) and dichoptic (red diamonds) viewing conditions.

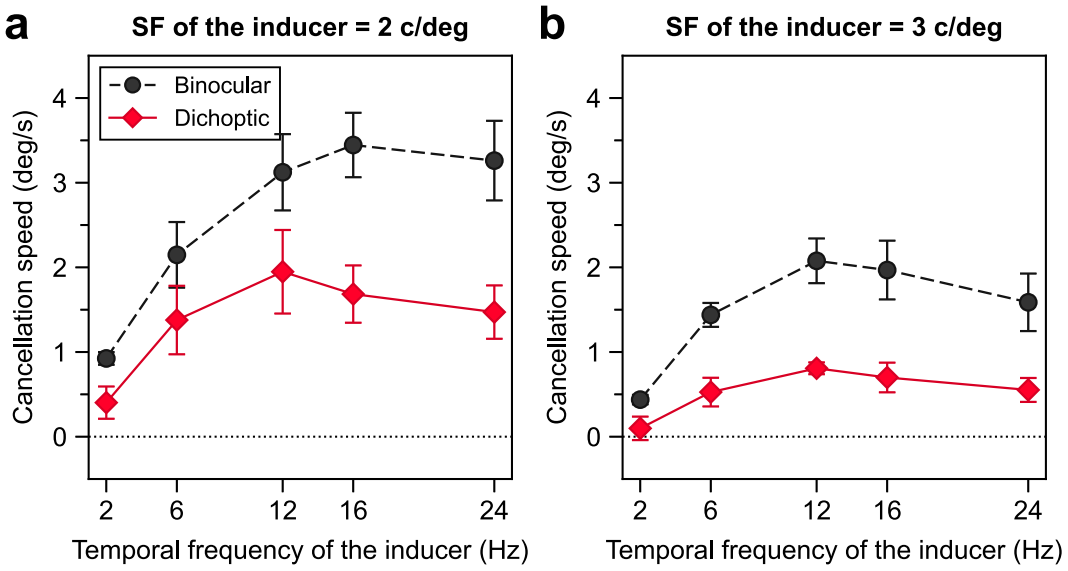

Figure 3. Results from experiment 2. The panels show the cancellation speed (mean ± standard error of the mean) as a function of the temporal frequency of the inducer for four observers. Black dots show the results for the binocular viewing condition. Red diamonds show the results for the dichoptic viewing condition. (**a**) Results for the spatial frequency of the inducer of 2 c/deg. (**b**) Results for the spatial frequency of the inducer of 3 c/deg.

The general pattern of the results shows an initial increase in cancellation speed with temporal frequency, followed by a stabilization or a slight decrease in the effect. Comparing the results of both inducers, we can see that the inducer of 2 c/deg (see Figure 3a) produces the motion illusion on the test component more intensely than the one with 3 c/deg (see Figure 3b), as shown by the higher cancellation speed in all the conditions tested. This confirms the results from experiment 1, where 2 c/deg produced the most intense illusory motion in the test stimulus.

Focusing on the inducer of 2 c/deg (see Figure 3a), the binocular results increase with the temporal frequency up to 16 Hz, from which they stabilize. Dichoptic presentations produce a similar pattern, but the cut-off temporal frequency is lower this time, specifically 12 Hz. A further increase in the temporal frequency of this inducer gradually reduces the cancellation speed. Furthermore, in agreement with experiment 1, the binocular viewing condition needed a higher speed to cancel the induced motion than the dichoptic one for all temporal frequencies tested.

Moving on to the 3 c/deg inducer (see Figure 3b), there is a similar pattern to the 2 c/deg inducer. The cancellation speed for binocular presentations increases until the inducer drifts at 12 Hz and then decreases. Dichoptic viewing produces a similar shape, with a saturation temporal frequency of 12 Hz. As it happened with the 2 c/deg inducer, the motion illusion is very weak with a temporal frequency of 2 Hz. Additionally, when the 3 c/deg inducer drifts at 12 Hz (speed of 4 deg/s), the cancellation speed results are consistent with what we observed in experiment 1 (see Figure 2a).

Experiment 3: Effect of the contrast level of the center and surround

In this experiment, we studied the effect of the contrast of both inducer and test stimulus, on the cancellation speed of the test, using 2 c/deg as the spatial frequency for the inducer, and two viewing conditions (binocular and dichoptic) for 50 ms. The spatial frequency of the test was always 1 c/deg. Figure 4 shows the cancellation speed as a function of the contrast of the inducer for binocular (black dots) and dichoptic (red diamonds) viewing conditions.

When the test has a Michelson's contrast of 0.1 (see Figure 4a), the cancellation speed for the binocular viewing condition increases very fast for low contrasts (0.1–0.3) of the inducer, and then, it decreases slightly until the inducer has high contrast (0.9). In contrast, in the dichoptic viewing condition, the cancellation speed increases with increasing contrast until 0.3 and then remains stable as the contrast of the inducer increases. Comparing both viewing conditions, the binocular condition needs higher cancellation speeds for all contrasts, but the difference between the cancellation speeds for both types of viewing conditions decreases as contrast of the inducer increases. Dichoptic condition also needs a higher inducer contrast to stabilize the cancellation speed.

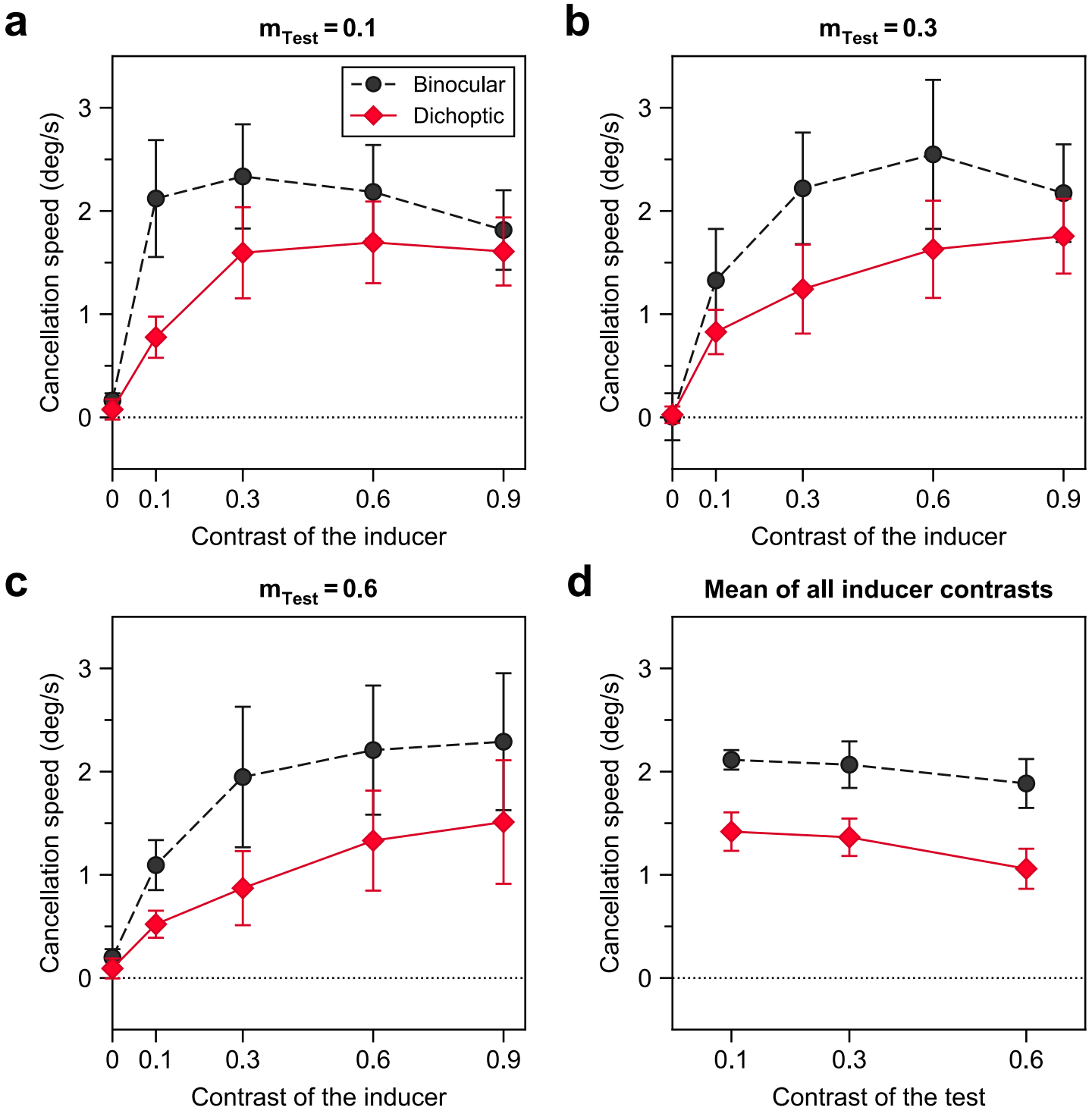

Figure 4. Results from experiment 3. The panels show the cancellation speed (mean ± standard error of the mean) as a function of the contrast level of the inducer for four observers. Black dots show the results for the binocular viewing condition. Red diamonds show the results for the dichoptic viewing condition. (**a**) The test stimulus has 0.1 contrast. (**b**) The test has 0.3 contrast. (**c**) The test has 0.6 contrast. (**d**) Average of the cancellation speeds of all inducer contrasts (excluding the control condition with zero contrast) as a function of the contrast of the test.

Continuing with 0.3 as the contrast of the test (see Figure 4b), the cancellation speed for the binocular condition increases until the contrast of the inducer is 0.3 and then approximately stabilizes. The shape of the results for the dichoptic condition is similar to the observations in panel a of Figure 4, but the cancellation speed stabilizes at 0.6, a higher contrast than before. This time, the difference between dichoptic and binocular conditions is greater for contrasts of the inducer between 0.3 and 0.6.

When the contrast of the test was 0.6 (see Figure 4c), the cancellation speed for the binocular condition increases until 0.3/0.6 contrast of the inducer and then stabilizes. The dichoptic condition, however, shows a continuous increase in the cancellation speed as the contrast of the inducer increases. The difference between both types of presentations is small for the lowest contrast and remains stable for the higher ones. Once again, the contrast from which the results stabilize is higher in the dichoptic conditions.

Finally, Figure 4d shows the mean of the cancellation speeds of all inducer contrasts (except the control condition with zero contrast) as a function of the contrast of the test. The results show that the cancellation speeds are similar within the binocular and dichoptic conditions for contrasts of 0.1 and 0.3 and slightly lower for 0.6, but differ between the binocular and dichoptic conditions, being lower for the dichoptic condition.

Model simulations

In this section, we aim to explain the reduction in the cross-scale interaction that occurs when the center and surround are presented dichoptically, compared with the binocular condition. To do this, we use the results of cancellation speed for the duration of 50 ms and the spatial frequency combination of 3 c/deg for the surround and 1 c/deg for the center (see Figure 2a). These results show a cancellation speed difference close to 1 deg/s (cancellation speeds: binocular = 1.965 deg/s, dichoptic = 0.896 deg/s). We chose this condition because it is consistent with the motion sensor interaction model of Serrano-Pedraza et al. (2007), which we use in the analysis. The original model is designed for overlapping low- and high-frequency stimuli. In our case, both stimuli are located in different spatial regions. To use this model, we assume that, owing to the spatial proximity of both stimuli, the connections between sensors tuned to low and high spatial frequencies remain active.

The original model was designed for binocular stimuli and, therefore, does not account for dichoptic presentations. To address this issue, we adapted the model to include an interocular gain control between monocular inputs, followed by a binocular summation (Ding & Sperling, 2006; Meese, Georgeson, & Baker, 2006; Meese & Hess, 2004). In particular, we use the late summation model (Meese & Hess, 2004; Meese et al., 2006) with monocular oriented motion energies, similar to Maehara, Hess, and Georgeson (2017). After binocular summation, the model includes the cross-scale interaction stage proposed by Serrano-Pedraza et al., (2007).

This new model includes several stages, starting with the extraction of oriented energy following Adelson and Bergen (1985). For this, we use 2D Gabor functions as spatial weighting functions (Watson & Ahumada, 1985). Specifically, the model contains two types: one tuned to 1 c/deg and the other to 3 c/deg (see examples in Figures 5b and 5c, respectively). Both types of sensors are located in the center and surround (see white circles in Figure 5a), and each type has a gain (see details of the model in the Appendix B). Next, the energy from the low spatial frequency (LF) sensors is pooled across the nine locations, and the same is done for the HF sensors. After extracting the oriented energy for LF and HF sensors in both motion directions, leftward and rightward, and for the left eye (LE) and right eye (RE), we obtained four motion energy outputs per eye. For simplicity, we only describe the outputs of the sensors tuned to leftward motion (L_{LE_LF}, L_{LE_HF}, L_{RE_LF}, and L_{RE_HF}). These outputs go through a stage of interocular gain control, followed by a binocular summation stage, that combines the motion energies that correspond to sensors tuned to the same scale and direction of motion on both eyes. For example, the leftward binocular response of the LF sensors (including interocular gain control and binocular summation) is $\mathbf{L_{B\_LF}} = (L^{p}_{LE\_LF} + L^{p}_{RE\_LF})/(z + L^{q}_{LE\_LF} + L^{q}_{RE\_LF})$, where p, q, and z are parameters. Based on previous studies (Meese et al., 2006; Maheara, et al., 2017), we used values within the fitted ranges. In particular, we set $p = 2.75$, $q = 3$, and $z = 0.1$ (note that different combinations of these parameter values would give similar results). Responses R_{B_LF}, L_{B_HF}, and R_{B_HF} are calculated in a similar way. Finally, these binocular responses reach the cross-scale interaction stage. This interaction consists of a subtraction followed by half-wave rectification between the LF and HF outputs. The interaction stage produces responses for both LF and HF sensors. Because we are simulating a cancellation speed task, which in this case is for a low-frequency test stimulus (1 c/deg), we only use the responses of the LF sensors after the interaction stage, where the energy from the HF sensors is subtracted from the LF sensors. For leftward motion, the response of the LF sensors can be computed as follows: $\mathbf{L_{B\_LF\_I}} = \lfloor L_{B\_LF} - L_{B\_HF} \rfloor$. The response to rightward motion after the interaction stage (not commented for simplicity) is calculated in an analogous manner. After the cross-scale interaction, we calculate the direction index, $DI = (R_{B\_LF\_I} - L_{B\_LF\_I})/(R_{B\_LF\_I} + L_{B\_LF\_I})$;

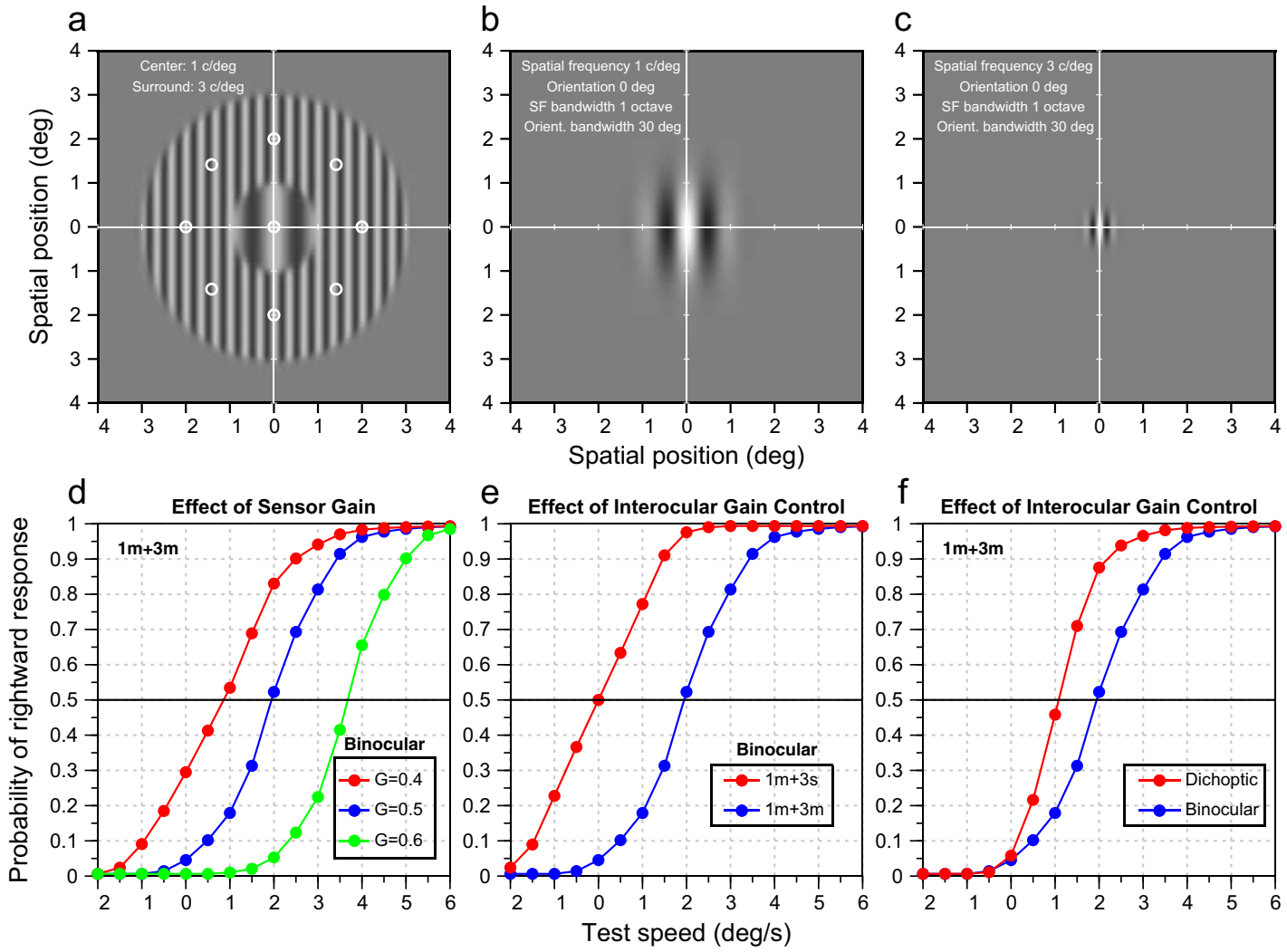

Figure 5. Simulation results. (**a**) Example of the stimulus used in the simulations and the locations of the spatial weighting functions (white circles). The center region (test stimulus) has a spatial frequency of 1 c/deg. The surround has a spatial frequency of 3 c/deg and can either move rightward at 4 deg/s or remain static. (**b**) Example of the spatial weighting function located in the center, with a spatial frequency of 1 c/deg. (**c**) Same as (b) but with a spatial frequency of 3 c/deg. (**d**) Effect of sensor gain for binocular viewing and condition 1m+3m. The gain of the low frequency sensor was always G = 1, the gain of the HF was G = 0.4 (red dots), G = 0.5 (blue dots), and G = 0.6 (green dots). (**e**) Effect of the interocular gain control. Simulation results for the binocular condition showing the probability of rightward response as a function of the speed of the test stimulus for a condition with a moving surround (1m+3m, blue dots) and a static surround (1m+3s, red dots). (**f**) Effect of the interocular gain control. Simulation results for 1m+3m, comparing binocular (blue dots) and dichoptic viewing (red dots) conditions. Note that positive test speed values indicate rightward motion, while negative values indicate leftward motion.

last, we transform this index into the probability of responding that the central stimulus (i.e., test) is moving to the right, using a normal cumulative distribution function (mean = 0, sd = 0.4).

We tested the model for 17 test speeds, from −2.0 deg/s to 6.0 deg/s in steps of 0.5 deg/s, using the stimulus represented in Figure 5a with the same parameters as those used in experiment 1. Figure 5d shows the effect of changing the relative gains of the motion sensors for binocular viewing and with the experimental condition 1m+3m. The panel represents the probabilities of responding that the test stimulus moves to the right as a function of the speed of the test stimulus (positive values indicate the central stimulus moves to the right, and negative values indicate the stimulus moves to the left). If the gain of the high-frequency selective sensors increased (G = 0.6), then a stronger interaction would occur increasing the cancellation speed (green dots). In contrast, if the gain of these high-frequency sensors decreased (G = 0.4), the strength of the interaction would also decrease, resulting in a reduction of the cancellation speed. The results for G = 0.5 show that,

for a probability of 0.5, the test cancellation speed value is similar to the value shown in Figure 2a (e.g., 1.96 deg/s). In this case, the effect of the interaction can be observed. For example, when the test speed is 0, the model predicts movement to the left (i.e., a low probability of responding to the right). The simulation also show that for a test speed close to 2 deg/s, in the same direction as the background movement, the test is perceived as stationary. These predictions are also presented in Figures 5e and 5f. Note that changing the gain of the sensors has a similar effect on the model as changing the stimulus contrast. The greater the surround contrast relative to the center, the greater the effect. Therefore, the model predicts a greater cross-scale interaction as the contrast difference between the test and the surround increases, as shown in the results of Figure 4.

Figure 5e shows two simulation results for binocular viewing condition. The red dots show the model's result when the background remains static (1m+3s, 1 c/deg moving plus 3 c/deg stationary). As can be seen, when the test speed is 0 deg/s, the probability is 0.5. The blue dots, however, show the results when the high-frequency surround moves to the right at a speed of 4 deg/s as in Figure 5d.

Figure 5f shows the simulations for dichoptic (red dots) and binocular (blue dots) conditions with stimulus 1m+3m. As can be seen, for the dichoptic condition, the strength of the interaction decreases. That is, for a probability of 0.5, the difference in cancellation speeds between dichoptic and binocular conditions is approximately 1 deg/s, consistent with the results shown in Figure 2a.

Simulation results (not shown) comparing binocular and dichoptic conditions for 100-ms duration show lower cancellation speeds for both conditions compared with the 50-ms duration. This result is expected, given that the same model has previously been used to explain the decrease in the strength of cross-scale interaction for binocular presentations with increasing stimulus duration (Luna & Serrano-Pedraza, 2018; Luna & Serrano-Pedraza, 2020; Serrano-Pedraza & Derrington, 2010; Serrano-Pedraza et al., 2007).

Discussion

In this work, we have presented three experiments where we measured the speed needed to cancel the illusory motion that a moving stimulus induced on a static test stimulus under several conditions. With this cancellation speed, we determined indirectly the strength of the interaction between spatial scales in motion processing using stimuli with different spatial frequency components presented on different retinal positions. Results showed that the cancellation speed depends on the viewing condition (binocular vs. dichoptic), the stimulus duration, the spatial frequency, the temporal frequency, and the relative contrast of the components of the stimulus. We have also performed simulations with a model of motion sensing that includes an interocular gain control stage (Meese et al, 2006), and a cross-scale interaction stage (Serrano-Pedraza et al., 2007). This model successfully explains the effect of the surround motion on center motion when both have different spatial frequencies, as well as the differences in cancellation speed observed between the binocular and dichoptic viewing conditions.

Effect of dichoptic presentations

In general, our results show that the strength of the interaction between spatial scales is higher for binocular viewing than for dichoptic viewing on most of the conditions tested. These results agree with those obtained by Derrington et al., (1993). They showed that the cancelation speed for the binocular condition was almost twice as fast as that of the dichoptic condition. Despite the latter one producing weaker results, the motion illusion was still present. Our results replicate this pattern and extend those results for different temporal frequencies, relative contrasts, and with the components of the stimulus presented in different retinal locations. However, we obtained slightly slower cancellation speeds for both viewing conditions. The difference might be explained by the shorter duration that they used (36 ms), because reducing the stimulus presentation increases the strength of the interaction (Derrington & Henning, 1987; Henning & Derrington, 1988; Luna & Serrano-Pedraza, 2020; Serrano-Pedraza et al., 2007; Serrano-Pedraza & Derrington, 2010). The different spatial structure of the stimuli could also explain this discrepancy. Henning and Derrington (1988) compared the strength of this mechanism for both superimposed and spatially separated components and showed that the intensity of this interaction was weaker for the latter condition. Regardless of the weaker intensity, these data provide an insight about what might be occurring physiologically in this interaction between spatial scales. Getting the effect in dichoptic presentations shows that there is no need to present both scales in the same eye to activate this mechanism. It provides evidence that the interaction mechanism integrates motion signals from both eyes.

Effect of the stimulus duration

In experiment 1, the interaction between spatial scales was stronger for the duration of 50 ms when

compared with 100 ms, indicating that the intensity of this mechanism increases with decreasing presentation time. These results also agree with previous research that used superimposing components (Derrington & Henning, 1987; Henning & Derrington, 1988; Luna & Serrano-Pedraza, 2020; Serrano-Pedraza et al., 2007; Serrano-Pedraza & Derrington, 2010). Furthermore, for 50 ms and in both viewing conditions, the activity of the motion mechanism was the strongest when the inducer had 2 c/deg, a higher spatial frequency than that of the test (1 c/deg). The inducer of 3 c/deg also produced strong results on both viewing conditions. However, when the duration increased to 100 ms, the interaction between spatial frequencies weakened and the inducer with the spatial frequency of the test produced the strongest illusion.

There are two relevant results to consider here. First, the shortest duration showed the strongest motion induction, where the illusion was stronger for binocular than for dichoptic condition. Second, decreasing the duration of the stimuli shifted the inducer's spatial frequency that produces the maximum cancellation speed to spatial frequencies higher than the spatial frequency of the test. For the shortest duration tested (50 ms), we have found that 2 c/deg (inducer) added to 1 c/deg (test) is the combination of spatial frequencies that produces the strongest illusion. Previous research, with short durations, had found that the interaction also appeared for different combinations of spatial frequencies, such as 0.5+1.5 c/deg (Derrington & Henning, 1987; Henning & Derrington, 1988; Luna & Serrano-Pedraza, 2018; Luna & Serrano-Pedraza, 2020). For our longest duration (100 ms), the maximum cancellation speed was found when the spatial frequency of the test and inducer were the same (1 c/deg). However, previous results on motion induction, using very long presentations (3 s), found that, for the two lowest spatial frequencies tested (0.625 and 1.250 c/deg), the peak of the induction ratio did not match the inducing frequency, as it was higher than the test (Levi & Schor, 1984). The authors suggest that this lack of coincidence between the spatial frequency of the inducer and the test could be due to the use of a constant temporal frequency for all inducing stimulus. In our case, we are presenting all inducers with the same speed (i.e., different temporal frequencies). Here, the interesting result is that decreasing the duration of the stimulus shifts the most effective inducer's spatial frequency to higher spatial frequencies, showing a strong interaction between different spatial scales.

Effect of temporal frequency and contrast

In experiment 2, we showed that the interaction between low and high spatial frequencies depends on the temporal frequency of the fine-scale component. Regardless of the spatial frequency of the inducer, the intensity of this mechanism increased with the temporal frequency of the inducer up to 12 or 16 Hz and stabilized from that point on, or decreased showing a slight bandpass shape. Luna and Serrano-Pedraza (2018) tested the effect of the temporal frequency using superimposing components and measuring duration thresholds and the proportion of correct responses. They also found a bandpass tuning of this interaction to the temporal frequency of the moving component, with the peak occurring at 6 to 12 Hz.

In experiment 3, we showed that the interaction between low and high spatial frequencies depends on the relative contrasts of both components. The shape of the results depends on the viewing condition and contrast of the test stimulus, but, in general, the strength of the motion illusion tends to stabilize for the higher contrast levels of the inducer. Serrano-Pedraza and Derrington (2010), using superimposed components, also showed a dependency of this motion mechanism on the contrast of the components. The motion illusion only arose when the contrast ratio between the fine and coarse gratings was in the range of 0.8 to 4.0, for a stimulus size of 2.8 deg, a duration of 25 ms, and a drifting velocity of 4 deg/s. They concluded that participants perceived the correct direction of motion for contrast ratios greater than 4 because the signal of the HF component was too strong, and the motion sensors tuned to fine scales dominated perception in that case. In our study, the interaction between spatial scales was present over a wider range of contrast ratios, from 0.17 to 9.00. This finding can be explained given that the subject's task used in this work was different. In our study, the subject had to respond to the perceived motion of the test ignoring the motion of the inducer, and in their study, they had to respond to the component with the strongest motion signal. There are other differences between both studies like the combination of spatial frequencies used, the viewing conditions, and the stimulus structure. For example, they superimposed a grating of 1 c/deg with another one of 3 c/deg, while we presented a grating of 1 c/deg in the center surrounded by another one of 2 c/deg.

Model simulations

One of the most interesting aspects of this study is the difference in the strength of the illusion observed between binocular and dichoptic presentation conditions at short durations, when a HF signal is moving in the surround and a LF signal is present in the center. In an attempt to explain this difference, we used a motion-sensing model that includes a cross-scale interaction stage, which has been used successfully in previous studies to explain the perceived

Spatial frequency (c/deg)	Binocular cancellation speed (deg/s)	Dichoptic cancellation speed (deg/s)	Difference (binocular–dichoptic)	Ratio (binocular/dichoptic)
2	2.495	1.665	0.83	1.49
3	1.965	0.896	1.07	2.19
4	1.238	0.261	0.98	4.74

Table 1. Results and comparisons between binocular and dichoptic presentations for the duration of 50 ms from experiment 1.

motion direction of moving complex patterns comprising superimposed coarse and fine scales (Luna & Serrano, 2018, Luna & Serrano, 2020; Serrano-Pedraza & Derrington, 2010; Serrano-Pedraza, Gamonoso-Cruz, Sierra-Vaźquez, & Derrington, 2013; Serrano-Pedraza et al., 2007). However, the original model was developed for binocular viewing conditions and does not account for dichoptic presentations. To address this limitation, we modified the model using interocular gain control followed by a binocular summation stage (Meese & Hess, 2004; Meese et al., 2006), before the cross-scale interaction.

The simulation results show that the model successfully reproduces the main finding of the study. In particular, it predicts the reduction in the cancellation speed observed in the dichoptic condition compared with the binocular condition (e.g., difference in cancellation speed of approximately 1 deg/s between binocular and dichoptic condition). Regarding the different combinations of spatial frequencies tested in the experiments, this study presents simulations for one experimental condition, 1 c/deg in the center and 3 c/deg in the periphery (1 m + 3 m), with a duration of 50 ms. When analyzing the binocular-dichoptic relationship in terms of cancellation speed across different spatial frequency combinations, our results suggest that the cross-scale interaction is a complex mechanism. In fact, results for the duration of 50 ms, shown in Table 1 (based on the data in Figure 2a), indicate that the binocular-dichoptic difference in cancellation speeds is approximately constant (∼1 deg/s). In contrast, the ratio between binocular and dichoptic cancellation speeds increases nonlinearly as the spatial frequency increases. This finding suggests that dichoptic cancellation speeds decrease more steeply than binocular cancellation speeds with increasing spatial frequencies. In contrast, the effect at 100 ms duration (see Figure 2b) is much smaller than at 50 ms, and the spatial frequencies of the surround that most influence the test stimulus differ from those observed in the 50 ms condition. These findings indicate that the cross-scale motion interaction is a complex and non-linear process, suggesting that the model in its current form is too simple to accurately reflect the underlying mechanisms.

In any case, a model composed of three stages—interocular gain control, binocular summation, and cross-scale interaction—seems to be the most plausible explanation. It is important to note that our model predicts that cancellation speed will be similar in monocular and dichoptic conditions (results not shown). Regarding motion perception, a binocular advantage has also been found (i.e., lower global motion thresholds) compared with monocular and dichoptic conditions for contrasts of less than 0.1 (Hess, Hutchinson, Ledgeway, & Mansouri, 2007). Additionally, greater binocular sensitivity compared with monocular sensitivity has been observed when measuring contrast thresholds of phase-reversing gratings (Rose, 1978; Rose, 1980). Binocular summation or binocular combination models attempt to explain why two eyes perform better than one at threshold. However, this advantage is less clear for suprathreshold stimuli. For instance, there is very little binocular advantage in suprathreshold contrast discrimination (Legge, 1984). Interestingly, a recent study measuring duration thresholds in a motion task with suprathreshold stimuli found that the strength of the interaction between fine and coarse scales was similar for both monocular and binocular presentations (Arranz-Paraíso, Prados-Rodríguez, & Serrano-Pedraza, 2022). Therefore, future studies should include a monocular condition to test this prediction of the model.

Overall, the results presented here show that a model that includes an interocular gain control predicts the results for both binocular and dichoptic viewing conditions. Our findings also indicate that the results obtained when the spatial scales are presented at different retinal locations resemble those observed with superimposed scales. In any case, the strong interaction between spatial scales found under dichoptic viewing suggests that this mechanism integrates motion signals from both eyes, even when the fine and coarse scales are presented on different retinal locations.

This model proposes that the cross-scale interaction occurs after binocular summation. This ordering is important because, if the cross-scale interaction is implemented after the interocular gain control but before binocular summation, the model fails to reproduce the effect observed in the dichoptic condition. The fact that cross-scale interaction emerges

when the fine and coarse scales are presented in different retinal locations, and therefore, in different V1 regions, suggests that this interaction could arise from the processing in the extrastriate cortex. Previous findings suggest also this hypothesis. For example, it is known that 80% of MT neurons have non-homogeneity of the center-surround spatial organization (Born & Bradley, 2005; Raiguel, Van Hulle, Xiao, Marcar, & Orban, 1995; Xiao, Raiguel, Marcar, Koenderink, & Orban, 1995; Xiao, Raiguel, Marcar, & Orban, 1997), and there are psychophysical results consistent with this spatial organization. Serrano-Pedraza, Hogg, and Read (2011) found stronger motion surround suppression when the stimulus was elongated in the direction of motion. In a different study, using elongated spatial windows, the motion interaction between spatial scales was only present when the window was elongated in the direction of motion (Serrano-Pedraza & Derrington, 2010). However, our results do not rule out the possibility that the interaction between fine and coarse scales could also arise from the lateral interactions between spatially separated V1 motion sensors tuned to different spatial scales. Further research will be needed to identify the cortical region responsible for the cross-scale interaction in motion processing.

Keywords: motion discrimination, binocular interaction, dichoptic viewing, cross-scale motion interaction, center–surround motion interactions

Acknowledgments

The authors thank Andrew M. Derrington for his help designing the experiments.

Supported by grant PID2021-122245NB-I00 from Ministerio de Ciencia e Innovación (Spain) to I.S-P. (https://www.ucm.es/serranopedrazalab/).

Author contributions: CRediT authorship contribution statement, O.B.: Formal Analysis, Data Curation, Investigation, Software, Validation, Visualization, & Writing – Original Draft; I.S.-P.: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Formal Analysis, Software, Supervision, Validation, Visualization, & Writing – Review & Editing.

Data availability: The datasets generated during and/or analysed during the current study are available here: https://hdl.handle.net/20.500.14352/108051.

Commercial relationships: None
Corresponding author: Ignacio Serrano-Pedraza.
Email: iserrano@ucm.es.
Address: Department of Experimental Psychology, Universidad Complutense de Madrid, Madrid 28223, Spain.

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, https://doi.org/10.1364/josaa.2.000284.

Anderson, S. J., & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research, 25*(8), 1147–1154, https://doi.org/10.1016/0042-6989(85)90104-x.

Anderson, S. J., & Burr, D. C. (1987). Receptive field size of human motion detection units. *Vision Research, 27*(4), 621–635, https://doi.org/10.1016/0042-6989(87)90047-2.

Anderson, S. J., & Burr, D. C. (1989). Receptive field properties of human motion detector units inferred from spatial frequency masking. *Vision Research, 29*(10), 1343–1358, https://doi.org/10.1016/0042-6989(89)90191-0.

Anderson, S. J., & Burr, D. C. (1991). Spatial summation properties of directionally selective mechanisms in human vision. *Journal of the Optical Society of America. A, Optics and Image Science, 8*(8), 1330–1339, https://doi.org/10.1364/josaa.8.001330.

Anderson, S. J., Burr, D. C., & Morrone, M. C. (1991). Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision. *Journal of the Optical Society of America. A, Optics and Image Science, 8*(8), 1340–1351, https://doi.org/10.1364/josaa.8.001340.

Arranz-Paraíso, S., Prados-Rodríguez, F., & Serrano-Pedraza, I. (2022). The strength of the interaction between fine and coarse scales is unaffected under monocular viewing. *9th Iberian Conference on Perception*. Barcelona, Spain: Universidad Autónoma de Barcelona. Retrieved from https://www.cip2022.org/.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience, 28* 157–189, https://doi.org/10.1146/annurev.neuro.26.041002.131052.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Derrington, A. M., Fine, I., & Henning, G. B. (1993). Errors in direction-of-motion discrimination with dichoptically viewed

stimuli. *Vision Research, 33*(11), 1491–1494, https://doi.org/10.1016/0042-6989(93)90142-j.

Derrington, A. M., & Henning, G. B. (1987). Errors in direction-of-motion discrimination with complex stimuli. *Vision Research, 27*(1), 61–75, https://doi.org/10.1016/0042-6989(87)90143-x.

Ding, J., & Sperling, G. (2006). A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America, 103*(4), 1141–1146, https://doi.org/10.1073/pnas.0509629103.

Emerson, P. L. (1986). Observations on maximum-likelihood and Bayesian methods of forced-choice sequential threshold estimation. *Perception & Psychophysics, 39*(2), 151–153, https://doi.org/10.3758/bf03211498.

Gonzalez, R. C., & Wintz, P. (1987). *Digital image processing*. Second edition. Reading, MA: Addison-Wesley.

Henning, G. B., & Derrington, A. M. (1988). Direction-of-motion discrimination with complex patterns: Further observations. *Journal of the Optical Society of America. A, Optics and Image Science, 5*(10), 1759–1766, https://doi.org/10.1364/josaa.5.001759.

Hess, R. F., Hutchinson, C. V., Ledgeway, T., & Mansouri, B. (2007). Binocular influences on global motion processing in the human visual system. *Vision Research, 47*, 1682–1692.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research, 34*(7), 885–912, https://doi.org/10.1016/0042-6989(94)90039-6.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception, 36*(ECVP Abstract Supplement), 1–16.

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

Levi, D. M., & Schor, C. M. (1984). Spatial and velocity tuning of processes underlying induced motion. *Vision Research, 24*(10), 1189–1195, https://doi.org/10.1016/0042-6989(84)90174-3.

Levinson, E., & Sekuler, R. (1975). The independence of channels in human vision selective for direction of movement. *Journal of Physiology, 250*(2), 347–366, https://doi.org/10.1113/jphysiol.1975.sp011058.

Luna, R., & Serrano-Pedraza, I. (2018). Temporal frequency modulates the strength of the inhibitory interaction between motion sensors tuned to coarse and fine scales. *Journal of Vision, 18*(13), 17, https://doi.org/10.1167/18.13.17.

Luna, R., & Serrano-Pedraza, I. (2020). Interaction between motion scales: When performance in motion discrimination is worse for a compound stimulus than for its integrating components. *Vision Research, 167*, 60–69, https://doi.org/10.1016/j.visres.2019.12.002.

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), 1–21. doi:10.1167/17.1.7.

Meese, T. S., & Hess, R. F. (2004). Low spatial frequencies are suppressively masked across spatial scale, orientation, field position, and eye of origin. *Journal of Vision, 4*(10), 843Y859, doi:10.1167/4.10.2.

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

Nishida, S. (2011). Advancement of motion psychophysics: Review 2001-2010. *Journal of Vision, 11*(5), 11, https://doi.org/10.1167/11.5.11.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Raiguel, S., Van Hulle, M. M., Xiao, D. K., Marcar, V. L., & Orban, G. A. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *European Journal of Neuroscience, 7*(10), 2064–2082, https://doi.org/10.1111/j.1460-9568.1995.tb00629.x.

Robson, J. G. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America. 56*, 1141–1142.

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

Rose, D. (1980). The binocular: Monocular sensitivity ratio for movement detection varies with temporal frequency. *Perception, 9*, 577–580.

Serrano-Pedraza, I., & Derrington, A. M. (2010). Antagonism between fine and coarse motion sensors depends on stimulus size and contrast. *Journal of Vision, 10*(8), 18, https://doi.org/10.1167/10.8.18.

Serrano-Pedraza, I., Goddard, P., & Derrington, A. M. (2007). Evidence for reciprocal antagonism between motion sensors tuned to coarse and

fine features. *Journal of Vision, 7*(12), 1–14, https://doi.org/10.1167/7.12.8.

Serrano-Pedraza, I., Gamonoso-Cruz, M. J., Sierra-Vaźquez, V., & Derrington, A. M. (2013). Comparing the effect of the interaction between fine and coarse scales and surround suppression on motion discrimination. *Journal of Vision, 13*(11), 1–13, doi:10.1167/13.11.5.

Serrano-Pedraza, I., Hogg, E. L., & Read, J. C. A. (2011). Spatial non-homogeneity of the antagonistic surround in motion perception. *Journal of Vision, 11*(2), 3, https://doi.org/10.1167/11.2.1.

Serrano-Pedraza, I., Manjunath, V., Osunkunle, O., Clarke, M. P., & Read, J. C. A. (2011). Visual suppression in intermittent exotropia during binocular alignment. *Investigative Ophthalmology & Visual Science, 52*(5), 2352–2364, https://doi.org/10.1167/iovs.10-6144.

Serrano-Pedraza, I., Vancleef, K., Herbert, W., Goodship, N., Woodhouse, M., & Read, J. C. A. (2020). Efficient estimation of stereo thresholds: What slope should be assumed for the psychometric function? *PLoS One, 15*(1), e0226822, https://doi.org/10.1371/journal.pone.0226822.

Sierra-Vázquez, V., Serrano-Pedraza, I., & Luna, D. (2006). The effect of spatial-frequency filtering on the visual processing of global structure. *Perception, 35*, 1583–1609.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*(17), 2503–2522.

van Santen, J. P., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America. A, Optics and Image Science, 2*(2), 300–321, https://doi.org/10.1364/josaa.2.000300.

Watson, A. B. (1986). Temporal Sensitivity. In: J. Thomas (Ed.). *Handbook of perception and human performance*. New York: Wiley.

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America. A, Optics and Image Science, 2*(2), 322–341, https://doi.org/10.1364/josaa.2.000322.

Xiao, D. K., Raiguel, S., Marcar, V., Koenderink, J., & Orban, G. A. (1995). Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. *Proceedings of the National Academy of Sciences of the United States of America, 92*(24), 11303–11306, https://doi.org/10.1073/pnas.92.24.11303.

Xiao, D. K., Raiguel, S., Marcar, V., & Orban, G. A. (1997). The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cerebral Cortex (New York, N.Y.: 1991), 7*(7), 662–677, https://doi.org/10.1093/cercor/7.7.662.

Appendix A: Fitted data from experiment 1

In this section, we present the fit of the data obtained in experiment 1. The main objective is to identify the theoretical spatial frequency of the inducer at which the cancellation speed is maximal ($f_{\max}$). To achieve this, we fitted a lognormal function with three free parameters ($f_{\max}$, A, and s) using the Matlab's *lsqnonlin* function (see blue lines in Figure A1). The lognormal function fitted to the cancellation speeds (in deg/s), was:

$$CancelSpeed\,(f) = A \exp\left[-\frac{\ln^2\,(f/f_{max})}{2s^2}\right], \quad \text{(A1)}$$

where f is the spatial frequency of the inducer (in c/deg); $f_{\max}$ is the spatial frequency at which the function reaches its maximum (in c/deg); A the value of the maximum (in deg/s); and s is the spread of the function. We also calculated the bandwidth (full width at half maximum) in octaves of the fitted function using the following equation:

$$B_{oct} = s \times \left[\frac{2\sqrt{2}}{\sqrt{\ln\,(2)}}\right]. \quad \text{(A2)}$$

Figure A1 shows the fits for four conditions (blue lines), binocular 50 ms and 100 ms (black dots), and dichoptic 50 ms and 100 ms (red diamonds). The fitted parameters are located in the top-right part of each panel.

Appendix B: Model details

In this section, we will describe the details of the cross-scale interaction model up to the calculation of the oriented energy from the motion sensors. The remainder of the model has been described in the main text, in the section "Model Simulations." The first stage of the model is based on the energy model outlined by Adelson and Bergen (1985, specifically their Figure 18b). It includes spatial weighting functions tuned to low and high spatial frequencies, and temporal impulse response functions. Each spatial weighting function is represented by a 2D Gabor function (Watson & Ahumada, 1985):

$$f\,(x, y) = G\,(\rho_0) \times \exp\left\{-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right\} \times \cos\,(2\pi\,\rho_0 x + \phi)\,. \quad \text{(B1)}$$

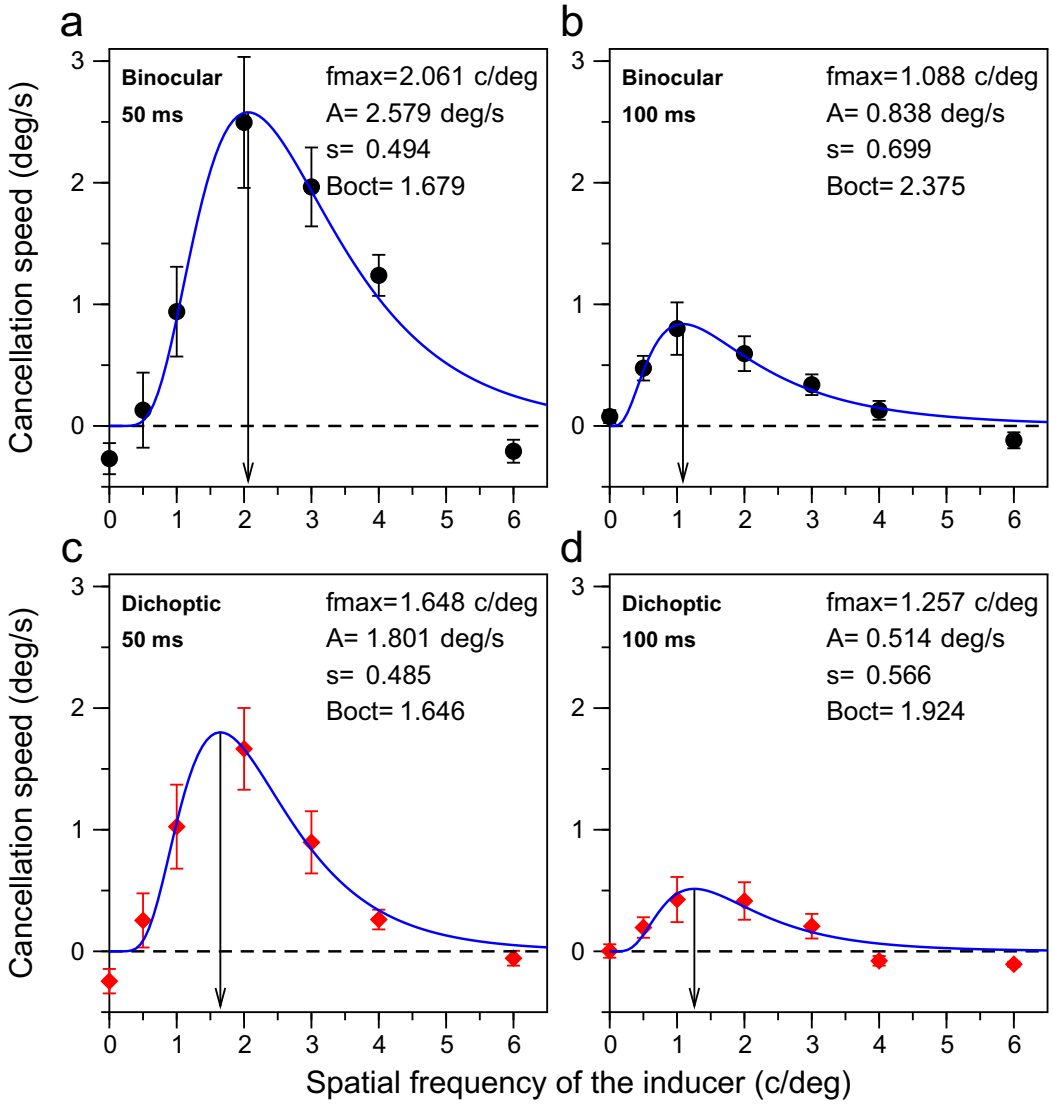

Figure A1. Fitted lognormal function to the data from experiment 1. The panels show the cancellation speed (mean ± standard error of the mean) as a function of the spatial frequency of the inducer for four observers. Black dots show the results for the binocular viewing condition. Red diamonds show the results for the dichoptic viewing condition. Blue line show the fitted lognormal function. The fitted parameters f_{max}, A, and s are presented on the top right part of each panel. We also include the bandwidth (full width at full maximum) of the fitted function (B_{oct}, in octaves). The arrow in each panel signals the location of the maximum of the lognormal function (f_{max}). (**a**) Results for binocular condition and duration of 50 ms. (**b**) Results for binocular condition and duration of 100 ms. (**c**) Results for dichoptic condition and duration of 50 ms. (**d**) Results for dichoptic condition and duration of 100 ms.

The spreads of the Gaussian window σ_x and σ_y were obtained using the following equations:

$$\sigma_x = \frac{\sqrt{\ln(2)}(1+2^B)}{\rho_0\sqrt{2}\pi(2^B-1)}, \quad \text{(B2)}$$

$$\sigma_y = \frac{\sqrt{\ln(2)}}{\rho_0\sqrt{2}\pi\tan(\alpha_0/2)}, \quad \text{(B3)}$$

where $B = 1$ octave (bandwidth in spatial frequency, full width at half-maximum), the orientation bandwidth of the sensors in degrees (full width at half-height) was $\alpha_0 = 30$ deg, and the spatial frequencies of the sensors were $\rho_0 \in \{1, 3\}$ c/deg. The gain of the sensor (G) was $G(1\text{c/deg}) = 1$ for the low-frequency sensors and $G(3\text{ c/deg}) = 0.5$ for the high-frequency sensors (the same for all locations). We also tested gains of $G(3\text{ c/deg}) = 0.4$ and $G(3\text{ c/deg}) = 0.6$ (see Figure 5). All sensors were vertically oriented. The simulations were performed with the sensors located at 9 different positions (see Figure 5a). The model used a quadrature pair of spatial sensors f_1 and f_2 for each spatial frequency (1 or 3 c/deg). The phase was $\phi = 0$ rad for the sensor $f_1(x, y)$, and $\phi = \pi/2$ rad for $f_2(x, y)$.

For the temporal impulse response functions $h_1(t)$, and $h_2(t)$, we used the equation from Watson & Ahumada, (1985, their equations 12 and 13):

$$h_2(t) = \xi\left[h_{21}(t) - \zeta h_{22}(t)\right], \quad \text{(B4)}$$

$$h_{2i}(t) = u(t) \times \left[\frac{(t/\tau_i)^{n_i-1} e^{-t/\tau_i}}{\tau_i (n_i - 1)!}\right], \quad \text{(B5)}$$

where $u(t)$, is the unit step function. The parameters used in the simulations were: $\xi = 214$, $\zeta = 0.9$, $\tau_1 = 6.22$, $\tau_2 = 8.27$, $n_1 = 9$, $n_2 = 10$. The temporal contrast sensitivity function obtained from this temporal impulse response function fit to the data of Robson (1966) for the spatial frequency of 0.5 c/deg (Watson, 1986). The fastest function, $h_1(t)$, was the quadrature pair of $h_2(t)$, calculated in the frequency domain using the Hilbert transform of $h_2(t)$ (Watson & Ahumada, 1985).

To obtain the response of a motion sensor located in the position i, we calculate the inner product of the stimulus ($I(x, y, t)$) with the spatial weighting function of the sensor and convolve the output with the temporal impulse response function:

$$A_i(t) = h_1(t) * \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x, y, t) \times f_{1i}(x, y)\, dxdy \quad \text{(B6)}$$

$$A'_i(t) = h_2(t) * \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x, y, t) \times f_{1i}(x, y)\, dxdy \quad \text{(B7)}$$

$$B_i(t) = h_1(t) * \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x, y, t) \times f_{2i}(x, y)\, dxdy \quad \text{(B8)}$$

$$B'_i(t) = h_2(t) * \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x, y, t) \times f_{2i}(x, y)\, dxdy. \quad \text{(B9)}$$

Following (Adelson & Bergen 1985; see their Figure 18b), we then calculate the oriented energy for leftward and rightward motion by integrating across time:

$$L_i = \int \left(A_i(t) - B'_i(t)\right)^2 + \left(A'_i(t) + B_i(t)\right)^2 dt \quad \text{(B10)}$$

$$R_i = \int \left(A_i(t) + B'_i(t)\right)^2 + \left(A'_i(t) - B_i(t)\right)^2 dt \quad \text{(B11)}$$

Finally, the energy was pooled across n locations (e.g., we used 9 locations): $L = \sum_{i=1}^{n} L_i$; $R = \sum_{i=1}^{n} R_i$. This way, we compute the energy for low- and high-spatial frequency tuned sensors (L_{LF}, R_{LF}, L_{HF}, and R_{HF}). These are the values that we use in later stages of the model, described in the main text.