Temporal frequency modulates the strength of the inhibitory interaction between motion sensors tuned to coarse and fine scales

Raúl Luna
Universidad Complutense de Madrid, Madrid, Spain

Ignacio Serrano-Pedraza
Universidad Complutense de Madrid, Madrid, Spain
Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK

Introduction

The basic idea about motion perception is that in the early stages of visual processing, motion is extracted by localized motion sensors tuned to spatial frequency, temporal frequency, and orientation, which work in parallel (Levinson & Sekuler, 1975; Adelson & Movshon, 1982; Anderson & Burr, 1987, 1989, 1991; Anderson, Burr, & Morrone, 1991). Classical energy models of human visual motion sensing have successfully implemented this basic structure to explain different motion phenomena like apparent motion, the missing-fundamental illusion, reverse Phi, etc., (Adelson & Bergen, 1985; Watson & Ahumada, 1985; van Santen & Sperling, 1985). However, there is a lot of psychophysical evidence suggesting the existence of an inhibitory mechanism that produces an interaction between motion sensors tuned to different scales in later stages of motion processing. In this line, human observers have been reported to make systematic errors in motion direction discrimination of very briefly presented motion stimuli containing features designed to activate motion sensors tuned to high and low spatial frequencies (Derrington & Henning, 1987; Henning & Derrington, 1988; Derrington, Fine, & Henning, 1993; Nishida, Yanagi, & Sato, 1995; Serrano-Pedraza et al., 2007; Serrano-Pedraza & Derrington, 2010; Serrano-Pedraza, Gamonoso-Cruz, Sierra-Vazquel, & Derrington, 2013; see also the “Interaction across different spatial scales” section in Nishida, 2011). In particular, at short durations, when a moving high spatial frequency pattern is added to a static low spatial frequency pattern, humans make...
systematic motion discrimination errors (Derrington & Henning, 1987).

Previous research has revealed different characteristics of this inhibitory mechanism. The interaction between coarse and fine scales is stronger at short durations (lower than 100 ms), although there is a strong effect for longer durations too (Serrano-Pedraza et al., 2007). This interaction is also very strong when the low spatial frequency component is lower than 1 c/° and the high spatial frequency component is higher than 3 c/° (Derrington & Henning, 1987; Henning & Derrington, 1988). On the other hand, the aforementioned interaction is not affected by the relative phase of the compound stimuli (Henning & Derrington, 1988), is stronger for large stimuli (Serrano-Pedraza & Derrington, 2010), depends on the relative contrast of the low and high spatial frequency components (Henning & Derrington, 1988; Serrano-Pedraza & Derrington, 2010), and has a larger effect on motion discrimination than on motion surround suppression (Serrano-Pedraza et al., 2013).

A simple model based on the energy models for motion perception that includes an inhibitory stage where sensors tuned to high and low spatial frequencies interact has successfully explained many of these psychophysical results (Serrano-Pedraza et al., 2007). The model predicts motion direction discrimination errors when the sensors tuned to low spatial frequencies (coarse sensors) have a strong signal for opposite directions of motion (briefly presented static stimuli) and when the sensors tuned to high spatial frequencies (fine sensors) have a strong signal for only one direction. This means that modulations of the strength of the signal or modulations of the strength of the output of the motion sensors tuned to high and low spatial frequencies should have an effect on motion direction discrimination. For example, for a complex stimulus composed of a static low-spatial frequency component added to a moving high-spatial frequency component, the interaction is stronger (e.g., motion perception is reversed) for a particular range of contrasts of the moving component and disappears for low and high contrasts (Serrano-Pedraza & Derrington, 2010). Classical models of motion sensing compute the direction of motion by simulating direction selective simple cells (Adelson & Bergen, 1985; Watson & Ahumada, 1985). Moreover, neurophysiological data show that these simple cells have either space-time inseparable receptive fields or a separable tuning for spatial and temporal frequency (DeAngelis, Ohzawa, & Freeman, 1993; Priebe, Lisberger, & Movshon, 2006). Thus, temporal frequency modulates the strength of the responses of direction selective simple cells (Priebe et al., 2006). It is interesting to note that although these direction selective simple cells may show a separable tuning for spatial frequency and temporal frequency (i.e., they are not speed tuned like complex cells or MT cells, e.g., Priebe et al., 2006), the speed of the stimulus can also modulate the strength of the cell’s response. According to Serrano-Pedraza et al.’s (2007) model, the interaction between scales takes place after motion sensors tuned to coarse and fine scales have produced their output, therefore, it is expected that the temporal frequency (or speed) of the moving component will have an effect on the strength of the interaction between scales. Because there are no previous studies that have explored in depth this prediction, the main objective of the present study is to explore the effect of different stimulus speeds (or temporal frequencies) on motion direction discrimination for complex stimuli composed of different combinations of fine and coarse scales.

In the present study we have performed three experiments measuring duration thresholds and proportion of correct discriminations for different combinations of low and high spatial frequencies drifting at different speeds (or temporal frequencies). Our results from Experiments 1 and 2 show that the strength of the interaction between motion sensors tuned to coarse and fine-scale patterns is modulated by the speed (or the temporal frequency) of the moving component. Moreover, this modulation is different depending on the complex stimulus type (i.e., both coarse and fine scales moving at the same time and with the same direction and speed, a static fine scale pattern added to a moving coarse scale pattern or a static coarse scale pattern added to a moving fine scale pattern). Our results from Experiment 3 show that the strength of the interaction is specifically modulated by the temporal frequency of the moving component. Furthermore, the inhibitory interaction shows a band-pass tuning function with a peak at 6 Hz when measuring duration thresholds and around 6 Hz and 12 Hz when measuring the proportion of correct responses. The presence of this band-pass tuning function depends on the particular combination of spatial frequencies for the static and the moving components.

Finally, we have run simulations using the model developed by Serrano-Pedraza et al., 2007, but this time including a few modifications (i.e., the temporal impulse response). The model is based on the classical motion energy models (Adelson & Bergen, 1985; Watson & Ahumada, 1985; van Santen & Sperling, 1985), but it additionally includes a stage where the outputs of motion sensors tuned to different scales interact. Although the model is too simple to be realistic, it reproduces most of our psychophysical results. In particular, it shows that the speed (or the temporal frequency) of the moving component modulates the strength of the interaction between motion sensors tuned to different spatial scales.
Methods

Subjects

Six human subjects (aged 21–42), three males and three females (DBW, GCM, ISP, RLV, SAP, and ZPR), took part in Experiment 1. Subjects DBW, GCM, SAP, and ZPR were not aware of the purpose of the study. In Experiment 2, we tested four human subjects: two males and two females (DBW, PGG, RLV, and SAP); just like in Experiment 3 (OBG, PGG, RLV, and SAP; two males and two females). This time only RLV was aware of the purpose of the study. All subjects had normal or corrected-to-normal vision. The participant’s normal visual acuity was checked using the visual test Sloan ETDRS 2000 letter series (Precision Vision, Lasalle, IL) (tested at 40 cm and 3 m). We also measured the participant’s stereoscopic visual acuity by using the Frisby (Stereotest Ltd, Sheffield, UK) stereotest (60 cm distance). Only subjects with visual acuity lower than logMAR = 0.5 (in both eyes) and participants with stereovision (stereoaucuity lower than 500 arcsec) took part in the experiments. The experiments were carried out in a dark room, using a chin rest (UHCOTech HeadSpot, Houston, TX) to stabilize the subject’s head and to control the observation distance. To minimize tracking eye movements, the subjects were instructed to maintain fixation on a small cross (0.258° × 0.258°) in the center of the screen before the stimuli were presented. All subjects signed an informed consent before taking the experiment. Experimental procedures were approved by the Universidad Complutense de Madrid Ethics Committee.

Equipment

The stimuli were presented on a gamma-corrected 17-in. Eizo Flex Scan TS65 monitor (Eizo Corp., Hakusan, Japan) under the control of a Mac Pro 3.7 GHz Quad Core Intel Xeon E5 (graphics card AMD FirePro D300 2048 MB) running MATLAB (MathWorks, Natick, MA) using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, & Pelli, 2007; www.psychtoolbox.org). The output was processed by a DataPixx Lite box (Vpixx Technologies Inc., Montreal, Canada) that gave us a minimum of 14 bits of gray-scale resolution. The monitor was gamma corrected using a Minolta LS-100 photometer (Konica Minolta Optics, Inc., Osaka, Japan) with routines programed in MATLAB to automatize the calibration process. The screen had a resolution of 800 × 600 pixels (horizontal × vertical), a vertical frame rate of 148 Hz, and a mean luminance of 49.1 cd/m².

Stimuli

All stimuli were created in MATLAB. The stimuli images had 512 × 512 pixels, were presented in the center of the monitor in a square of 19.5 × 19.5 cm at a distance set to 100 cm, and subtended a visual angle of 11.3° × 11.3°. The display had a spatial resolution of 45.9 pixels/°. The stimuli used in all experiments were Gabor patches (vertical gratings spatially windowed by a 2D Gaussian function). The standard deviations of the 2D spatial Gaussian window were σx = σy = 2° (see Equation 1). Two types of stimuli were used: simple drifting vertical Gabor patches comprised of only one spatial frequency and complex vertical Gabor patches resulting from the addition of the two simple Gabor patches. The equation of the simple drifting Gabor patch is described as follows:

\[
L(x, y, t) = L_0 \left[ 1 + m(t) \exp \left\{ -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right\} \right] \times \cos(2\pi \rho_1 (\hat{x} - vt) + \phi_1) \] (1)

The equation of a complex moving stimulus is as follows:

\[
L(x, y, t) = L_0 \left[ 1 + m(t) \exp \left\{ -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right\} \right] \times \left[ \cos(2\pi \rho_1 (\hat{x} - vt) + \phi_1) + \cos(2\pi \rho_2 \hat{y} + \phi_2) \right] \] (2)

where \( \hat{x} = x \cos(\theta) + y \sin(\theta) \) and \( \hat{y} = -x \sin(\theta) + y \cos(\theta) \); x and y are the on-screen positions; \( \theta \) is the orientation, in degrees (all stimuli had a vertical orientation); \( \theta = 0^\circ \); \( L_0 \) is the mean luminance (\( L_0 = 49.1 \text{ cd/m}^2 \)); \( \rho_1 \) and \( \rho_2 \) are the spatial frequencies in cycles per degree (c/°); \( \phi_1 \) and \( \phi_2 \) are the phases of the Gabor patches, in radians (in all experiments the phases of the spatial components were randomized); \( \sigma_x \) and \( \sigma_y \) are the spatial standard deviations of the Gaussian window, in degrees of visual angle (°); \( v \) is the speed of the moving pattern, in degrees per second (°/s); and \( m \) is the Michelson contrast as a function of time given by

\[
m(t) = M \times \exp \left\{ -t^2/(2\sigma_m^2) \right\} \] (3)

where \( t \) is time, in milliseconds (ms); \( M \) is the peak contrast, and \( \sigma_m \) is the temporal standard deviation. Simple Gabor patches had a peak contrast of 0.275. On the other hand, the final contrast of the complex Gabor patches resulted from the addition of two simple
Gabor of 0.275 contrast. In the Procedure section we will describe the Bayesian staircase method that controls the parameter $r_t$ in order to obtain the duration thresholds ($2 \times r_t$). Figure 1 shows examples of the stimuli used in the experiments.

In Experiments 1 and 2, we tested five different conditions: (a) Drifting Gabor patch of 1 c/8 spatial frequency (1m); (b) Drifting Gabor patch of 3 c/8 spatial frequency (3m); (c) Complex stimulus composed of two drifting Gabor patches of 1 and 3 c/8 (1m + 3m); (d) Complex stimulus composed of two Gabor patches, 1 c/8 static added to 3 c/8 moving (1s + 3m); and (e) Complex stimulus composed of two Gabor patches, 1 c/8 static added to 3 c/8 static (1s + 3s). In all these examples, all moving components drift at a speed of 2c/s.

**Procedure**

In Experiment 1 we measured duration thresholds; that is, the minimum presentation time needed for a subject to discriminate the correct direction of motion of a stimulus, resulting in a performance of 82% correct responses during a motion direction discrimination task. Duration thresholds were defined as twice the standard deviation of the Gaussian temporal function, $2 \times r_t$ (see Equation 3). To measure these thresholds, we used adaptive Bayesian staircases (Treutwein, 1995) in a forced-choice direction discrimination task. The characteristics of the Bayesian staircases were as follows: (a) The prior probability-density function was uniform (Emerson, 1986) with a starting duration of 200 msec. (b) A logistic function was used as the model likelihood function (adapted from García-Pérez, 1998, his appendix A), with a spread value of 1; a delta parameter of 0.01; a lapse rate of 0.01; and a guess rate of 0.5. (c) The temporal presentation of the stimuli was determined by a Gaussian temporal window that controlled the stimulus contrast as a function of its time (see Equation 3). The value of the temporal standard deviation (in logarithmic values) for each trial was obtained from the mean of the posterior probability distribution (Emerson, 1986; King-Smith, Grigsby,
Vingrys, Benes, & Supowit, 1994). (d) The staircase stopped after 40 trials (Madigan & Williams, 1987; Anderson, 2003). (e) The final threshold was estimated from the mean of the final probability-density function. Three threshold estimations per condition were obtained for each subject.

In Experiment 2, we measured the proportion of correct responses (i.e., motion perceived in the direction of the moving stimulus) in a forced-choice discrimination motion task. Proportions were calculated across 40 trials for each of the stimulus conditions and stimulus durations. The stimulus presentation durations were 25 and 50 ms (2 × στ, στ = 12.5 and 25 ms).

In Experiment 3, we measured both duration thresholds and proportion correct responses as in Experiments 1 and 2.

Results

Experiment 1: Effect of speed on motion direction discrimination of complex stimuli

The main objective of Experiment 1 is to determine the effect of speed on the interaction between motion sensors tuned to different spatial scales for different types of complex stimuli. In this experiment and in Experiment 2 we will be talking about speed (and not temporal frequency) because we want to use this same dimension used in previous research (Serrano-Pedraza et al., 2007; Serrano-Pedraza et al., 2010; Serrano-Pedraza et al., 2013) to compare conditions where only one component is moving (e.g., 1m, 3m, 1s + 3m, 1m + 3s) or the two components of the complex stimulus are moving (e.g., 1m + 3m). For example, for the condition 1m + 3m, the speed of both simple components, which is the same, corresponds to different temporal frequencies (e.g., a speed of 4°/s corresponds to 4 Hz and 12 Hz for 1 c/° and 3 c/° respectively).

We therefore measured duration thresholds for six subjects in a motion direction discrimination task using simple and complex stimuli for five different drifting speeds: 0.5°/s, 1°/s, 2°/s, 4°/s, and 8°/s. Figure 2 shows the duration thresholds (in log10 ms) as a function of speed for the five stimulus conditions tested. The condition with the lowest averaged thresholds for all speeds was 3m (green squares), the next conditions with similar duration thresholds were 1m (red dots) and 1m + 3m (blue triangles), then, with higher thresholds, we had the condition 1m + 3s (magenta triangles), and finally, the condition with the highest thresholds was 1s + 3m (black upright triangles).

From the averaged results, we can conclude that in all conditions but 1s + 3m, duration thresholds decrease with increasing speed. That is, our ability to discriminate motion direction improves with increasing speed. Similar results have been found using high contrast gratings (Lappin, Tadin, Nyquist, & Corn, 2009; see their Figure 2B). However, in the condition where the low spatial frequency is static, and the high spatial frequency component is moving (1s + 3m; black upright triangles), the averaged results show that duration thresholds increase with increasing speed up to 2°/s and then decrease. This result, for this particular condition, reveals that the speed (or the temporal frequency) of the moving component modulates the strength of the suggested inhibitory mechanism responsible for the interaction between motion sensors tuned to coarse and fine spatial scales in a band-pass fashion. Provided these results, one might be tempted to suggest that the interaction mechanism under study is either speed or temporal frequency tuned. However, further verifications still need to be performed to clarify this point. These will be addressed in Experiment 3. In any case, taking a look at each subject, individual differences regarding the 1s + 3m compound stimulus can be appreciated. Participants RLV, SAP, and ZPR do show a band-pass function. However, ISP shows a low-pass tuning function, DBW shows no tuning and GCM’s results for this stimulus are ill defined.

To estimate the inhibitory strength of the interaction between motion scales, we computed an inhibition index using the logarithmic ratio of the duration thresholds for complex to simple stimuli (log10(Dcomplex/Dsimple)). This can also be obtained by subtracting the logarithmic values of the duration thresholds, log10(Dcomplex)−log10(Dsimple); where Dcomplex and Dsimple are the duration thresholds, in ms, for the complex and the corresponding simple moving stimuli respectively. This index had been used before to estimate the strength of motion surround suppression (Tadin et al., 2006).

Figure 3 shows the average logarithmic ratios of the duration thresholds for complex to simple stimuli as a function of the speed of the moving components. Ratios larger than 0 indicate that the complex stimuli have a higher duration threshold than the simple ones. Figure 3A shows that for condition 1m + 3s, its ratio to 1m is greater than 0 for speeds lower than 1°/s. Conversely, for speeds higher than 2°/s there is no such effect. However, for the condition 1s + 3m, the ratios increase with increasing speed up to 2°/s, and then the strength of the interaction becomes almost constant. Thus, although duration thresholds (see Figure 2) suggest a band pass tuning function for speed, ratios show that the strength of the interaction increases with increasing speed, reaching an asymptotic value at about 2°/s. Figure 3B shows the logarithmic ratios of the complex stimulus 1m + 3m to the simple stimuli 1m and 3m as a function of speed. The results indicate that duration thresholds for both components moving together are higher (for all speeds tested) than the
thresholds for the simple stimulus 3m. On the other hand, duration thresholds for both components moving coherently at the same speed are lower than the thresholds for the simple stimulus 1m (for the speed of 2°/s and lower speeds). These results reveal that when the task is to discriminate the motion direction of a complex stimulus (1m + 3m), the visual system performs worse than when detecting only one of the components of the complex stimulus (in this case, 3m). Interestingly, duration thresholds for this complex stimulus are similar to the thresholds for the low spatial frequency component, in agreement with previous

Figure 2. Results from Experiment 1. Duration thresholds (in log_{10} ms) for simple (1 and 3 c/°) and complex stimuli (1m + 3m; 1m + 3s; 1s + 3m) as a function of speed. The top panels show the individual data (mean ± SEM). The bottom panel shows the average of six subjects. Each symbol represents the mean ± SEM. Note that the x-axis is represented on a logarithmic scale.
results (Hayashi, Sugita, Nishida, & Kawano, 2010). This suggests that subjects judge motion direction of a complex stimulus composed of coarse and fine features moving coherently at the same speed, using predominantly motion signals from the low spatial frequency component.

**Experiment 2: The effect of speed on motion discrimination for short durations**

In Experiment 1, the most interesting result was that, for the stimulus condition 1s + 3m, the averaged results showed that duration thresholds were modulated by speed (or temporal frequency) in a band-pass fashion, with a maximum at about 2°/s (6 Hz) (see Figure 2, upright black triangles). Nevertheless, as already said, no claims about a pure speed or a temporal frequency tuning of the inhibitory mechanism (independent of spatial frequency) can be made. Experiment 3 (will be explained) was in fact performed to shed light on this issue. Bear in mind, though, the individual differences reported in the previous section, which did not show a band-pass tuning function for every subject.

Previous results have shown strong perceptual reversals for the condition 1s + 3m at very short presentation durations (19 and 68 ms) and for different temporal frequencies (Derrington & Henning, 1987). Those results reveal that perceptual reversals also depend on the speed of the moving spectral components, with the strongest effect happening at 8 Hz (2.6°/s), which is similar to the maximum we have found using duration thresholds in Experiment 1. Nevertheless, Derrington and Henning (1987) only tested the condition 1s + 3m, so the effect of speed on motion perception at short presentation times for other complex stimuli composed of different combinations of simple stimuli (1m + 3s and 1m + 3m) still has not been explored. Therefore, in Experiment 2 we have extended the results from Derrington and Henning (1987) by measuring five stimulus conditions (the same ones tested in Experiment 1). In doing that, we have measured the proportion of correct responses in a motion direction discrimination task for two short stimulus presentation durations (25 and 50 ms) and five different drifting speeds (0.5°/s, 1°/s, 2°/s, 4°/s, & 8°/s). For each condition, presentation time, and speed, we ran 40 trials.

Figure 4 sums up the performance of the observers in Experiment 2. Figures 4AC and 4BD show the performance for presentation times of 50 ms and 25 ms, respectively. In all panels, symbols represent the average proportion of correct responses (± 95% score confidence interval) for simple and complex stimuli as a function of the speed of the moving components.

As it can be seen from Figure 4C, the proportion of correct responses for the simple stimuli (red dots and green squares) decreases with decreasing speed, in agreement with the results from Experiment 1, where duration thresholds decreased with increasing speed (the lower the proportion of correct responses, the higher the duration thresholds). For speeds lower than
Figure 4. Results from Experiment 2 for four subjects. Each symbol represents the proportion of correct responses (mean ± 95% score confidence interval) as a function of the speed of the moving stimuli (in the case of simple stimuli) or the moving spectral component or components (in the case of complex stimuli). (A) Individual results for 50 ms presentation time. (B) Individual results for 25 ms presentation time. (C) Average results for 50 ms presentation time (duration of the Gaussian temporal window, $2 \times \sigma_t$, where $\sigma_t = 25$ ms). (D) Average results for 25 ms presentation time ($\sigma_t = 12.5$ ms). Red dots, 1 c/deg; green squares, 3 c/deg; blue triangles, 1m + 3m; magenta triangles, 1m + 3s; and black triangles, 1s + 3m. For each subject, the proportion of correct responses was obtained from 40 trials per condition and speed.
8°/s, the proportion of correct responses for the low spatial frequency stimulus (1m) is always lower than that for the high spatial frequency stimulus (3m), again in agreement with Experiment 1, where duration thresholds for the simple stimulus, 3m, were lower than those for 1m. For the complex stimuli, 1m + 3m (blue triangles) and 1m + 3s (magenta triangles), the proportion of correct responses also decreases with decreasing speed in agreement with the results from Experiment 1. Finally, the results for the condition 1s + 3m show a U-shape pattern with stronger reversals (perception of motion in the opposite direction) for speeds between 2°/s and 4°/s (temporal frequencies between 6 and 12 Hz). Once again, these results are in agreement with the results from Experiment 1, where performance was modulated by the speed (or temporal frequency) of the high spatial frequency moving component, and the highest duration threshold was obtained for 2°/s (i.e., 6 Hz), thus depicting a band-pass shape.

Figure 4D shows a lower proportion of correct responses for all stimulus conditions. Regarding the condition 1s + 3m, performance keeps stable at the chance level for speeds of 0.5°/s, 1°/s, and 2°/s, and then decreases with increasing speed. It is now for speed 8°/s that performance is maximally impaired. Reversals can in fact be appreciated for this speed, and to a lesser extent, for the speed of 4°/s.

**Experiment 3: Testing whether the interaction is speed or temporal frequency tuned**

Results from Experiments 1 and 2 reveal that for the condition 1s + 3m, the speed (or temporal frequency) of the moving component modulates the strength of the interaction. Given that we have tested only one spatial frequency for the moving component (i.e., 3 c/°), we are not in a position to make any assumptions on the band-pass function evidenced in duration thresholds or in proportion correct responses being speed tuned or temporal frequency tuned. Therefore, we have performed a third experiment testing four human subjects in seven different stimulus conditions: three simple stimuli, 1.5m c/°, 4m c/°, 6m c/°, and four complex stimuli, 0.5s + 1.5m c/°, 1s + 4m c/°, 1s + 6m c/°, and 2s + 6m c/°. The rationale behind this is that by having more spatial frequency pairs than only 1s + 3m, if a band-pass tuning is still found that has its peak at a given speed or temporal frequency (common for the different pairs of complex stimuli with different spectral components), then a specific tuning does exist. Moreover, by checking whether that peak is common to the different stimuli either when representing the results against speed or representing them against temporal frequency, one should be able to specify the tuning nature of the interaction mechanism, either speed or temporal frequency tuned. Bear in mind that for different spatial frequencies, the same drifting speed corresponds to different temporal frequencies, so a common tuning peak for temporal frequency (temporal frequency tuning) is translated into different peaks for speed (therefore, no speed tuning) and vice versa. To put this idea into practice, we measured both duration thresholds and proportion of correct responses for the different stimulus conditions. In the case of the proportion of correct responses, we used a stimulus presentation duration of 50 ms, which is the presentation time for which a band-pass modulation was evidenced in Experiment 2 (see Figure 4A and 4C). In the present experiment, the speeds of the moving components were calculated to have the same temporal frequencies as in Experiment 1 (i.e., 1.5, 3, 6, 12, and 24 Hz).

In any case, it is worth saying that previous research has shown that when the spatial frequency of the static component is higher than 2 c/° and the spatial frequency of the moving component differs significantly from the spatial frequency of the static component, then the interaction is reduced (Derrington & Henning, 1987). Thus, we expect a larger interaction for the conditions 0.5s + 1.5m and 1s + 3m than for 1s + 4m, 1s + 6m, or 2s + 6m.

Figures 5 and 6 show the results from Experiment 3 (we have split the results into two figures for clarity). The results for the conditions 3m and 1s + 3m, represented in both figures, are taken from Experiments 1 and 2. Figure 5 shows the results for simple stimuli (3m, 4m, and 6m) and complex stimuli (1s + 3m, 1s + 4m, 1s + 6m, and 2s + 6m). Figures 5A and 5B show individual data for the proportion of correct responses and duration thresholds, respectively. Figures 5B and 5C show the averaged results.

Figures 5A and 5C show that the results for the simple stimuli (3m, 4m, and 6m) show similar patterns. This is, the proportion of correct responses becomes reduced for temporal frequencies lower than 3 Hz. On the other hand, for the conditions 1s + 4m (gray triangles), 1s + 6m (white up-triangles), and 2s + 6m (white down-triangles), although the effect of the interaction is still very strong, results do not show a clear band-pass tuning for temporal frequency. Only condition 1s + 4m shows a slight band-pass tuning with a small trough in the proportion of correct responses at 12 Hz. Thus, from these conditions still we cannot conclude whether the interaction between scales is speed or temporal frequency tuned.

Figures 5B and 5D show duration thresholds (in log10 ms) as a function of temporal frequency. Results for simple stimuli show, once more, similar patterns: duration thresholds are very alike, and they become reduced with increasing temporal frequency. In the case of the complex
Figure 5. Results from Experiment 3. (A) Individual data. Proportion of correct responses as a function of the temporal frequency of the moving component. Each symbol represents the proportion of correct responses (mean ± 95% score confidence interval) as a function of the temporal frequency of the moving component for simple (4m and 6m c/°) and complex stimuli (1s + 4m; 1s + 6m; and 2s + 6m c/°). For each subject, the proportion of correct responses was obtained from 40 trials per stimulus condition and speed. (B) Individual data. Duration thresholds (in log_{10} ms) as a function of temporal frequency. (C) Averaged data from the panels presented in (A) (N = 4 subjects) and from Figure 4C (N = 4 subjects). Each symbol represents the proportion of correct responses (mean ± 95% score confidence interval) as a function of the temporal frequency of the moving component for simple (3m, 4m, and 6m c/°) and complex stimuli (1s + 3m; 1s + 4m; 1s + 6m; and 2s + 6m c/°). (D) Averaged data from the panels presented in (B) (N = 4 subjects) and from Figure 2 (N = 6 subjects). Each symbol represents the mean ± SEM of four subjects. (E) Strength of the inhibitory interaction (log ratio of complex to simple stimuli) obtained from the results of Experiment 1 (black triangles) and Experiment 3 (gray triangles, white up-triangles, and white down-triangles).
stimuli, we have found, as expected, that the interaction is stronger (i.e., higher duration thresholds) for 1s + 3m than for the other two conditions. Additionally, the condition 1s + 4m shows a stronger interaction than 1s + 6m or 2s + 6m. Moreover, the shape of the data shows that the tuning curve is evident in conditions 1s + 3m and in 1s + 4m, and disappears for 1s + 6m and 2s + 6m. Furthermore, in both conditions (1s + 3m and in 1s + 4m)
the maximum duration threshold appears at 6 Hz. Thus, these results suggest that the interaction between coarse and fine scales is tuned to temporal frequency, but only for a particular combination of spatial frequencies.

Figure 5E shows the log ratios of the duration thresholds for the complex stimuli to the simple stimuli. Results show that the strength of the interaction is reduced when the spatial frequency of the moving component increases. The four conditions show a similar pattern: the strength increases up to 6 Hz and then keeps constant. Thus, the ratios do not show selectivity for temporal frequency.

Figure 6 shows the results for the conditions 1.5m and 0.5s + 1.5m. We have added the conditions 3m and 1s + 3m for comparison. Figures 6A and 6B show the individual data whereas Figures 6C and 6D show the averaged data.

Figures 6A and 6C show that the proportion of correct responses for the simple stimulus (1.5m) become reduced with decreasing temporal frequency. For the complex stimulus (0.5s + 1.5m), the proportion of correct responses shows a clear band-pass tuning for temporal frequency with a minimum at 12 Hz both for the individual and the averaged results. Only subject PGG shows a minimum between 6 and 12 Hz.

Figures 6B and 6D show that duration thresholds for the simple stimulus (1.5m) become reduced with increasing temporal frequency (same pattern as that found for 3m, 4m and 6m). For the complex stimulus (0.5s + 1.5m), duration thresholds show a peak at 6Hz for everyone but RLV, whose peak appears at a temporal frequency of 3Hz. Finally, the averaged data for duration thresholds suggests a temporal frequency band-pass tuning function with a peak at 6Hz.

Figure 6E shows the log ratios of the duration threshold for the complex stimulus (0.5s + 1.5m) to the simple stimulus (1.5m). Again, ratios increase up to 6 Hz and then keep constant.

Model simulations

Serrano-Pedraza et al. (2007) proposed a simple model to explain the perceptual reversals of briefly presented visual complex stimuli composed of a static low spatial frequency component and a moving high spatial frequency component (i.e., 1s + 3m). Here we wanted to test whether this same basic model could reproduce some of the results presented in Experiment 2. Our basic motion analyzer uses the computational approach from the Adelson & Bergen (1985) motion energy detector and the filter parameters from the Watson and Ahumada (1985) linear motion sensor. The model implemented here has two differences with respect to Serrano-Pedraza et al.’s model (see a full description in Appendix A). First, we are not using spatial sensors for many different positions and orientations (we are only using vertical sensors centered on the stimulus position), and second, we have implemented the temporal impulse response suggested by Watson & Ahumada, (1985) with parameters that fit the results from Robson (1966). Basically, the model contains two spatial sensors of spatial frequencies 1 and 3 c/deg. Each spatial sensor computes the inner product with the stimulus (multiplying the sensor by the stimulus across time) and then the inner product is convolved with the temporal impulse response. Next, we calculate the oriented energy integrating across time, following Adelson and Bergen (1985; see their Figure 18b). Later, the response of the high frequency sensor is subtracted from the response of the low frequency sensor (and vice versa) for the same direction of motion (right or left). Responses are then half-wave rectified and, finally, the definitive response of the model is taken from the spatial frequency channel that has the highest difference between the right and left responses. The highest difference is then converted to a direction index, which is translated into a performance score using a sigmoidal response function to obtain the probability of correct responses.

We run the simulations using the same stimulus conditions and presentation durations used in Experiment 2. Figure 7A and 7B show the simulation results for two presentation times, 50 and 25 ms, respectively. Figure 7A reproduces the main findings from Experiment 2 (see Figure 4A). It shows that for the simple stimuli (1m and 3m) and for two of the complex stimuli (1m + 3m and 1m + 3s), the proportion of correct responses decreases with decreasing speed. The model also predicts a worse performance for the condition 1m + 3m (blue triangles) than for the simple stimuli. This last prediction is not in agreement with the averaged data, although one subject, DBW (Figure 4A), shows results that are similar to the model predictions. In fact, in Serrano-Pedraza et al. (2007), three participants (two authors and one unexperienced observer) showed strong impairments for the condition 1m + 3m, using anisotropic noise. These results suggest that the interaction in the condition 1m + 3m is not present in all subjects. For the condition 1s + 3m (black triangles), the model predicts a U-shape similar to the one found in humans’ results, with a minimum between 2°/s and 4°/s (6 and 12 Hz). However, the model predicts a lower proportion of correct responses than those found in the empirical results. Nevertheless, it is important to note that the model has not been fitted to the data, so we are more interested in the shape of the predictions than in the particular values predicted by the model.

Figure 7B shows a reduction in the proportion of correct responses which is in agreement with the results presented in Figure 4B. Conversely, for the condition 1s + 3m (black triangles), the prediction does not
reproduce our empirical results. Although the proportions of correct responses for this stimulus are higher than those of the prediction for 50 ms, the U-shape is still preserved, but is not present in the empirical results.

Finally, the model also predicts very similar results when using a different temporal impulse response (Adelson & Bergen, 1985) (results not shown).

Limitations of the model

The model proposed is too simple to be completely realistic, but it serves to demonstrate that the subtraction of sensor responses reproduces some of the basic failures and reversals evidenced in motion perception. However, as seen in Figure 7B, the model fails to reproduce the results for the condition 1s + 3m at very short durations (25 ms). It should also be noted that we also tried to replicate the results presented in Figure 2 by changing the presentation durations in the model until we got a probability of correct responses of 0.82 (the same used in the Bayesian staircases). However, in the condition 1s + 3m, for the speed of 0.5°/s, the model predicted a worse performance as we increased the presentation time. Also, for the speed of 8°/s, the model’s prediction never went higher than a probability of 0.5. Therefore, the model in its current state cannot account for the results of duration thresholds. The same model, but changing different parameters such as the spatial frequency of the sensors, their gains, the temporal impulse response functions, etc. could account for some of the results obtained from Experiment 3, in particular, the ones about the proportion of correct responses. Future work will focus on improving the current model to predict the results from duration thresholds and for different combinations of spatial frequencies for the static and moving components.

Discussion

Motion direction discrimination of a fine scale pattern is impaired when a static coarse scale pattern is added to it (Derrington & Henning, 1987). Previous studies have tried to characterize this phenomenon by manipulating different stimulus characteristics like the presentation time, the size, the relative contrast, the spatial frequency, etc. (Derrington & Henning, 1987; Henning & Derrington, 1988; Derrington et al., 1993; Serrano-Pedraza et al., 2007). All the results obtained suggest that an inhibitory interaction between motion sensors tuned to different fine and coarse scales underlies this phenomenon. Finally, a simple model implementing this idea has successfully explained some of the results (Serrano-Pedraza et al., 2007).

Here, we have performed three experiments to test the effect of speed (or temporal frequency) on motion direction discrimination for different types of complex stimuli, and we have also tested a basic motion model to explain some of the results obtained.
In Experiment 1, for the first time, we have measured duration thresholds for two simple and three complex stimuli as a function of the speed of the moving components. Our results show that for all conditions but 1s + 3m (1 c/° static added to a moving 3 c/°), duration thresholds decrease with increasing speed (see Figure 2), not showing any speed (or temporal frequency) band-pass tuning. A similar result was previously found by Lappin et al. (2009) using gratings with high contrasts. This result is not very surprising given that duration thresholds are the minimum time needed to detect the correct direction of motion, so when the stimulus moves slowly, more time is needed to correctly discriminate its direction. However, the averaged results for all subjects when a static pattern was added to a moving pattern (1s + 3m) show that duration thresholds increase with increasing speed up to 2°/s (6 Hz) and then decrease, showing a band-pass shape (see our Figure 2). It is important to notice that duration thresholds for 3c/° do not show any band-pass tuning for speed or temporal frequency. Therefore, speed seems to modulate the strength of the aforementioned inhibitory interaction, which in principle could suggest that the hypothetical interaction between motion sensors is speed tuned. However, given that we have only tested one single spatial frequency for the moving component (i.e., 1s + 3m), we cannot confirm with this experiment that the interaction is purely speed tuned. Temporal frequency tuning must be taken into consideration too (such consideration will be addressed later). In any case, one must consider the individual differences evidenced, which only show a band-pass tuning for three out of six subjects. Therefore, caution should be taken when generalizing these results: a band-pass tuning function might be evidenced for some subjects, but not for all of them; be it because the stimulus choice is not optimal for them to have such tuning or because no band-pass tuning actually exists for them. In a further analysis, we have calculated the logarithmic ratio of the duration thresholds of the complex stimulus (1s + 3m) to the simple stimulus (3m) in order to measure the strength of the interaction. Results show that the strength of the interaction increases with speed up to 2°/s (6 Hz) and then stabilizes (see Figure 3A). Therefore, the band-pass shape is only present when representing duration thresholds, and not logarithmic ratios, as a function of speed (or temporal frequency). Regarding the case when a static fine-scale pattern was added to a moving coarse-scale pattern (1m + 3s), previous results from Serrano-Pedraza et al. (2013) showed that a facilitation effect occurred at high contrasts when a 1 c/° pattern moving at a speed of 2°/s was added to a 3c/° static pattern. However, at low contrasts, duration thresholds were similar to the case when the moving pattern was presented alone. This, in fact, is reflected in the results from Experiment 1.

There is another relevant result that concerns the complex stimulus where both coarse and fine scales move together coherently at the same speed (1m + 3m). Performance for this stimulus is generally impaired with respect to the high spatial frequency stimulus (3m) at least for the slowest speeds tested. It is surprising that although this complex stimulus contains information from fine scales, for which our human visual system performs better in motion direction discrimination, it does not seem to take advantage of that information. In fact, performance for the complex stimulus generally resembles that for the simple low spatial frequency stimulus (1m), in agreement with previous results (Hayashi et al., 2010).

In Experiment 2, we have measured the proportion of correct responses for the same five stimulus conditions tested in Experiment 1, two presentation times (25 and 50 ms) and five different speeds. For all conditions except 1s + 3m and for both presentation times, the proportion of correct responses increases with increasing speed, in agreement with the results of Experiment 1. Also, performance is generally enhanced when the stimulus presentation time is increased from 25 to 50 ms. This can be easily explained through the spatiotemporal spectrum of the stimuli: shorter presentation times introduce more energy in the opposite direction of motion, making the task harder (see Figure 1A). However, for the condition 1s + 3m, presented for 50 ms (Figure 4A), our results show that the proportion of correct responses as a function of speed (or temporal frequency) depicts a U-shape, and for speeds of 2°/s and 4°/s (6 Hz and 12 Hz), it is below the chance level (perception of motion is reversed). The aforesaid U-shape disappears at very short durations (25 ms), where the proportion of correct responses for 0.5°/s, 1°/s, and 2°/s lies at the chance level. More experiments changing the contrast and increasing the size of the stimuli would clarify whether the tuning is still present at very short durations. Derrington and Henning (1987) measured the proportion of correct responses only for the condition 1s + 3m for different temporal frequencies and two durations of 18 and 68 ms. The results they obtained showed strong perceptual reversals, with the stronger effect happening at 8 Hz (2.6°/s). Their data suggests that speed modulates the strength of the interaction, but given the strength of the reversals, it is hard to conclude from their results that the interaction between scales is tuned to a particular temporal frequency or speed. Finally, the stronger interactions found by Derrington and Henning (1987) are probably related to the larger size of the stimuli they used, 7.5° × 6.25°. Bear in mind that in our study we have used 4° (i.e. 2σₓ), and in a previous study, we showed that
perceptual reversals increased with increasing stimulus size (Serrano-Pedraza & Derrington, 2010).

In Experiment 3, we have measured duration thresholds and proportion of correct responses for different combinations of spatial frequencies (1s + 4m, 1s + 6m, 0.5s + 1.5m, and 2s + 6m c/°). Our aim was to confirm whether the band-pass tuning function found for the condition 1s + 3m in Experiments 1 and 2 was speed or temporal frequency tuned by checking whether a band-pass tuning function still appeared with a common peak either for speed (evidence for a speed tuned mechanism) or for temporal frequency (temporal frequency tuned mechanism). A common tuning peak for temporal frequency would mean different peaks for speed and vice versa. Our results showing the proportion of correct responses are not conclusive for 1s + 6m and 2s + 6m. However, for the condition 1s + 4m, we have found a slight tuning around 12 Hz, and for 0.5s + 1.5m, we found a clear band-pass tuning with a minimum at 12 Hz (see Figure 6C) (remember the similar temporal frequency tuning for 1s + 3m when measuring the proportion of correct responses, around 6 Hz and 12 Hz). For duration thresholds we haven’t found a band-pass tuning for conditions 1s + 6m and 2s + 6m. However, we have found a clear band-pass tuning function for 1s + 4m and 0.5s + 1.5m with a peak at 6 Hz, which is the same we have found for 1s + 3m (see Figure 5C and 6C). These results suggest that the interaction between coarse and fine scales is temporal frequency tuned and is restricted to a particular combination of spatial frequencies. In addition, one may appreciate some gradation in the tuning strength when comparing the tuning peaks for the conditions 1s + 3m, 1s + 4m, and 1s + 6m (from strongest to weakest). This suggests some spatial frequency limit for the inhibitory mechanism for which the band-pass tuning starts to vanish. Our results indeed suggest that the combinations 1s + 3m and 0.5s + 1.5m, which produce the strongest interaction, are optimal in order to show the band-pass tuning. On the contrary, 1s + 4m is somewhat intermediate at showing the aforesaid tuning, whereas 1s + 6m and 2s + 6m tell us about the aforementioned limits for the inhibitory mechanism. Taking all these results together we can confirm that the mechanism underlying the interaction between scales is more complex than it was previously thought. Therefore, more research is needed to understand the properties of this inhibitory mechanism.

In summary, the results reported in the present study are novel in that Experiments 1, 2, and 3 show that the interaction between motion sensors evidenced in previous studies is temporal frequency dependent. The temporal frequency tuning of the interaction mechanism becomes obvious when a static coarse scale pattern is added to a moving fine scale pattern (for a particular combination of static and moving spatial frequencies). In this case, one can see that the tuning function has a band-pass nature, having its peak at a temporal frequency of 6 Hz when duration thresholds are measured. On the other hand, the interaction is stronger (larger number of reversals) at temporal frequencies between 6 Hz and 12 Hz when measuring the proportion of correct responses with a presentation time of 50 ms.

As a final check, we have simulated the results from Experiment 2 using a simple model of motion sensing (Serrano-Pedraza et al., 2007). The model contains two stages: in the first stage the model computes the oriented energy (Adelson & Bergen, 1985) using spatial sensors tuned to high and low spatial frequencies. In the second stage, the model computes an interaction between the outputs of the motion sensors tuned to different scales. This simple model reproduces most of the results from Experiment 2; in particular, the data obtained for stimulus presentation duration of 50 ms. The model predicts that for the condition 1s + 3m, the proportion of correct responses as a function of speed depicts a U-shape with the maximum number of reversals (minimum proportion of correct responses) happening between 2°/s and 4°/s. According to the model, changes in the speed (or temporal frequency) of the moving spectral components produce an imbalance between the energies of the motion sensors that is responsible for the evidenced temporal frequency tuning. The maximum interaction takes place at a temporal frequency of about 9 Hz, which is approximately where the temporal sensitivity function used in the simulations has its maximum. Consequently, the shape of the temporal frequency tuning curve of the interaction seems to be related to the shape of the temporal sensitivity function. Nevertheless, and as already said, individual differences do exist. In this sense, different parameters concerning the spatial and temporal impulse response functions can be adjusted in such a way that diverse individual responses could be reproduced. Perhaps no unitary model is valid for the present case, but anyhow, our model does reproduce a band-pass tuning function when certain conditions are met. Of course, not all subjects may meet these conditions, this being the reason why not all of them show a tuning. On the other hand, we are aware that the model we have used is too simple to be realistic, but given that it reproduces most of the psychophysical results presented in this study, the subtraction between the outputs of motion sensors tuned to fine and coarse scales seems to be a feasible underlying mechanism for the observed phenomenon. It is interesting to note that there are other nonlinear models that, without implementing any particular interaction between motion sensors, could explain part of the results obtained with complex stimuli (Hayashi, Watanabe, Yokoyama, & Nishida, 2017). With regard to our proposed mecha-
nism, there is other evidence that supports its existence. For example, the interaction between motion sensors could explain reversals in the perceived direction of motion of second-order modulations (Cropper, Kvansakul, & Johnston, 2009) and the asymmetric shape of the motion after-effect tuning functions (Ledgeway & Hutchinson, 2009). On the other hand, it could also explain why judgments on motion direction are based on motion signals from low spatial frequencies rather than high spatial frequencies (Hayashi et al., 2010).

Lastly, little is known about the ecological function of the reported antagonism between motion sensors. One may speculate that it could facilitate some aspects of image interpretation, like the detection of object rotations, where there exist differential motions between fine-scale features of the surface texture and coarse-scale features of the object body (Serrano-Pedraza et al., 2007). Likewise, the interaction mechanism could help signaling differences between figure and background when a high spatial frequency object (eye-fixated) moves over a low spatial frequency static background and vice-versa, as would be the case of motion parallax.

Keywords: motion perception, speed, temporal frequency, interaction between fine and coarse scales, motion sensors, inhibition

Acknowledgments

Supported by grant PSI2014-51960 from Ministerio de Economía y Competitividad (Spain) to ISP. RLV is supported by the predoctoral fellowship BES-2015-074077 from Ministerio de Economía y Competitividad (Spain), co-funded by the European Social Fund.

Commercial relationships: none.
Corresponding authors: Raúl Luna; Ignacio Serrano-Pedraza.
Email: ralu001@ucm.es; iserrano@ucm.es.
Address: Faculty of Psychology, Universidad Complutense de Madrid, Madrid, Spain.

References


### Appendix A

#### Model

The model used in the simulations is based on the energy model described by Adelson and Bergen (1985, their figure 18b) and includes a stage that combines the outputs of the motion sensors tuned to low and high spatial frequencies. The model contains spatial weighting functions and temporal impulse response functions; the spatial weighting function was a 2D Gabor function (Watson & Ahumada, 1985):

\[
f(x, y) = \gamma(\rho_0) \times \exp\left\{ -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right\}
\]

\[
\times \cos(2\pi\rho_0 x + \phi),
\]

where
\[ \hat{x} = (x - x') \cos(\theta_0) + (y - y') \sin(\theta_0) \] 
\[ \hat{y} = -(x - x') \sin(\theta_0) + (y - y') \cos(\theta_0). \]

The spreads of the Gaussian function \( \sigma_x \) and \( \sigma_y \) were obtained using the following equations:
\[ \sigma_x = \frac{\sqrt{\log(2)} (1 + 2^B)}{\rho_0 \sqrt{2\pi} (2^B - 1)} \]
\[ \sigma_y = \frac{\sqrt{\log(2)}}{\rho_0 \sqrt{2\pi} \tan(\pi_0/2)} \]

where \( B = 1 \) octave (bandwidth in spatial frequency, full width at half-height), \( \pi = 30^\circ \) was the orientation bandwidth of the sensors in degrees (full width at half-height), and the spatial frequencies of the sensors were \( \rho_0 \in \{1, 3\} \) c/\( \pi \). The function \( \gamma \) was the gain of the sensor, where \( \gamma(1 c/\pi) = 0.2 \) and \( \gamma(3 c/\pi) = 1 \). The orientation of the sensors was \( \theta_0 = 0^\circ \) (vertical). All simulations were completed with the sensors located at the position 0 degrees (\( x' = 0, y' = 0 \)). For each spatial frequency (1 or 3 c/\( \pi \)), the model used a quadrature pair of spatial sensors \( f_1 \) and \( f_2 \) (Adelson & Bergen, 1985; Watson & Ahumada, 1985; van Santen & Sperling, 1985). For the sensor \( f_1(x, y) \), the phase was \( \phi = 0 \) rad and for \( f_2(x, y) \), the phase was \( \phi = \pi/2 \) rad.

For the temporal impulse response functions \( h_1(t) \) and \( h_2(t) \), we used the equation from Watson and Ahumada, (1985, their equations 12 and 13):
\[ h_2(t) = \xi[h_21(t) - \zeta h_22(t)], \]
\[ h_21(t) = u(t) \times \left( \frac{(t/\tau_1)^{10} e^{-t/\tau_1}}{\tau_1(n_1 - 1)!} \right), \]
\[ h_22(t) = \hbar(t) \times \left( \frac{(t/\tau_2)^{10} e^{-t/\tau_2}}{\tau_2(n_2 - 1)!} \right), \]

where \( u(t) \) is the unit step function. The parameters used in the simulations were: \( \xi = 214, \zeta = 0.9 \), \( \tau_1 = 6.22, \tau_2 = 8.27, n_1 = 9, n_2 = 10 \). These are the parameters that fit the temporal contrast sensitivity function (obtained from the temporal impulse response function) to the data of Robson (1966) for the spatial frequency of 0.5 c/\( \pi \) (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).

In the first stage, the model calculated the responses of the motion sensors to the stimuli. The response of a motion sensor was calculated from the inner product of the stimulus with the spatial weighting function of the sensor, and the convolution of the inner product with the temporal impulse response function:
\[ A(t) = h_1(t) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y, t) \times f_1(x, y) \, dx \, dy \]
\[ A'(t) = h_2(t) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y, t) \times f_1(x, y) \, dx \, dy \]
\[ B(t) = h_1(t) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y, t) \times f_2(x, y) \, dx \, dy \]
\[ B'(t) = h_2(t) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y, t) \times f_2(x, y) \, dx \, dy \]

Following Adelson and Bergen (1985; see their Figure 18b) we calculated the oriented energy by integrating across time:
\[ L = \int (A(i) - B'(i))^2 + (A'(i) + B(i))^2 \, dt \]
\[ R = \int (A(i) + B'(i))^2 + (A'(i) - B(i))^2 \, dt \]

For both motion sensors tuned to low (LF) and high (HF) spatial frequencies, we calculated the oriented energy: \( L_{LF}, R_{LF}, L_{HF}, \) and \( R_{HF} \). In the second stage, we implemented the interaction between the outputs of the sensors tuned to different spatial frequencies. The interaction consisted of a subtraction and half-wave rectification between sensors with low and high spatial frequency:
\[ L_{LF} = |L_{LF} - L_{HF}|, R_{LF} = |R_{LF} - R_{HF}|, \]
\[ L_{HF} = |L_{HF} - L_{LF}|, R_{HF} = |R_{HF} - R_{LF}|. \]

After the subtraction, the psychophysical response was calculated using the sensor, LF or HF, that had the greatest difference between left and right oriented energy:
\[ \max(|L_{LF} - R_{LF}|, |L_{HF} - R_{HF}|). \]

Next, the direction index (DI) was calculated using the following function
\[ DI = \frac{R - L}{R + L}, \quad -1 \leq DI \leq 1. \]

Finally, the DI was transformed into proportion of correct responses using a normal cumulative distribution function:
\[ P(R | R) = 0.5 \times [1 + erf(DI/0.5)] \],
where
\[ \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp[-t^2] \, dt. \]

The stimuli used in the simulations were the same ones used in the experiments. However, for the simulations, the stimuli always moved rightwards. We also used different combinations of spatial phases (0, \( \pi/2 \), \( \pi \), and \( 3\pi/2\) rad) for the simple and complex stimuli, and we calculated the average of the model predictions for each stimulus.