"Non-retinotopic processing" in Ternus motion displays modeled by spatiotemporal filters

Arezoo Pooresmaeili   Stella Maris Scientific Institute, Pisa, Italy

Guido Marco Cicchini   Università degli Studi di Firenze, Firenze, Italy

Maria Concetta Morrone   Stella Maris Scientific Institute, Pisa, Italy, Department of Physiological Sciences, Università di Pisa, Pisa, Italy, & Department of Robotic, Brain and Cognitive System, Istituto Italiano di Tecnologia, Genova, Italy

David Burr   Università degli Studi di Firenze, Firenze, Italy, & Institute of Neurosciences, CNR, Pisa, Italy

Recently, M. Boi, H. Ogmen, J. Krummenacher, T. U. Otto, & M. H. Herzog (2009) reported a fascinating visual effect, where the direction of apparent motion was disambiguated by cues along the path of apparent motion, the Ternus–Pikler group motion, even though no actual movement occurs in this stimulus. They referred to their study as a "litmus test" to distinguish "non-retinotopic" (motion-based) from "retinotopic" (retina-based) image processing. We adapted the test to one with simple grating stimuli that could be more readily modeled and replicated their psychophysical results quantitatively with this stimulus. We then modeled our experiments in 3D (*x*, *y*, *t*) Fourier space and demonstrated that the observed perceptual effects are readily accounted for by integration of information within a detector that is oriented in space and time, in a similar way to previous explanations of other motion illusions. This demonstration brings the study of Boi et al. into the more general context of perception of moving objects.

Keywords: Ternus motion display, retinotopic processing, spatiotemporal filters

Introduction

In his elegant essay entitled "reconstructing the visual image in space and time," Barlow (1979) pointed out that a major problem facing the visual system is "that the image is almost constantly moving over the retina, and consequently the copy being transmitted to the cortex also moves; why then is it seen as sharp, unsmeared and still?" (p. 189). He concluded that the visual system is designed to perform *spatiotemporal interpolation*. "Some process of integration must occur, because one experiences a single moving object [in apparent motion], not a succession of separate stationary ones; but if this is integration in the mathematical sense, it must be done along a trajectory in space and time matching that of the moving object, and not separately and sequentially in the two domains." That is, vision cannot be considered as a successive sequence of frames on the retina, but space and time must be analyzed together, conjointly.

Barlow's conclusions were based largely on experiments performed around that time, demonstrating spatiotemporal interpolation for stimuli in motion (Burr, 1979; Fahle & Poggio, 1981; Morgan & Thompson, 1975). In the simplest form—illustrated in Figure 1—a two-bar vernier sequence moves in stroboscopic motion, as if behind a picket fence, so the top and bottom bars are always displayed to the same spatial position but at different times. Although the two bars are never displayed together, they are perceived as a single object with a clear physical horizontal offset (Burr & Ross, 2004). A "photographic" or *retinotopic* integration predicts that no offset will be seen, whereas integration along the motion trajectory will result in an apparent offset (also see Movie 1).

This experiment and others like it show that the visual system does not consider only the retinal position of stimuli but interpolates between the physical samples and does so with high accuracy (Burr & Ross, 2004; Burr, 1979; see also Nishida, 2004). Burr, Ross, and Morrone (1986) later suggested that these effects, and others such as lack of motion smear (Burr, 1980), could be explained by the spatiotemporal tuning properties of motion detector units in early visual cortex. Their tuning in space–time can account both for the interchangeability of space on time in
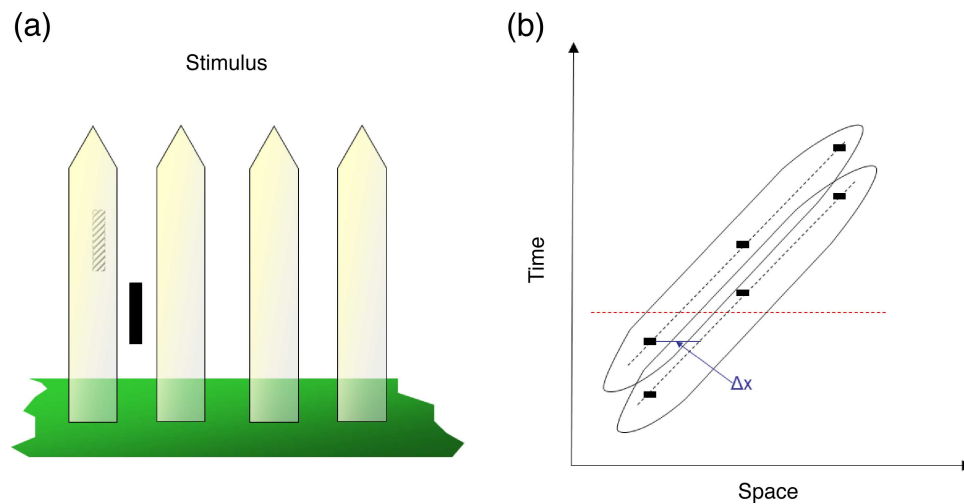
Figure 1. The concept of spatiotemporal integration. (a) A vernier stimulus moves behind a virtual picket fence. Although at any point of time only one of the bars is visible while the other is occluded by the fence, the stimulus is perceived as two bars with a horizontal offset between them (see Movie 1). (b) The stimulus is illustrated in space and time. The horizontal offset between the two bars can only be seen if motion is integrated along the trajectory illustrated by the elongated ellipses, whereas at any point of time (such as that illustrated by the red dashed line), *retinotopic* integration will not produce any horizontal offset.

causing spatial displacements of moving objects (hence, spatiotemporal interpolation) and the reduction in motion smear.

Recently, Herzog et al. have pursued a similar line of research, showing how motion can influence not only the form and position of stimuli but also motion itself. In their experiment, like the previous research on spatiotemporal interpolation, they show that perception not only depends on the retinal stimulation of stimuli but can also be strongly influenced by a motion trajectory, in their case apparent motion. In a clever series of experiments (Boi, Ogmen, & Herzog, 2011; Boi, Ogmen, Krummenacher, Otto, & Herzog, 2009; Otto, Ogmen, & Herzog, 2006; Otto, Ogmen, & Herzog, 2009), they have used the Ternus–Pikler stimulus (Ternus, 1926) to disentangle "retinotopic" from "motion-based" positional information. This stimulus is particularly interesting because, as illustrated in Figure 2, at appropriate timing, motion is induced in stimulus elements that do not actually move.

The Ternus–Pikler display used by Boi et al. (2009) comprises three disks, with a single dot inserted inside each (Figure 2). In the outer disks, the dot was in the center, while in the central disk, at each frame, the dot was presented along the trajectory of a clockwise or anti-clockwise rotation. In both clockwise and anti-clockwise rotations, the location of the dot at the third frame was 180° shifted relative to the first frame. Therefore, the location of the dot in the second frame was crucial to disambiguate the direction of rotation since +90° phases indicated clockwise rotation (phase progression from 0° to +90° to 180° to 270°) and 270° indicated anti-clockwise rotation (phase progression from 0° to 270° to 180° to 90°). The disks were shifted horizontally back and forth across the frames. The rotation is perceived only if the dot

inside the central disk at each frame is matched to the dot at a different retinotopic location (but one on the motion trajectory) in the next frame. The authors found that when the display contained three disks the direction of rotation was judged correctly, whereas displays with only two disks were associated with near chance level performance. Since the three-disk displays were perceived to move coherently as a group while the two disks lacked all sense of motion, the authors suggested that the non-retinotopic motion of the dot becomes apparent only after group motion is established. In such a framework, the motion processing occurs in two computational steps: First, motion correspondence between elements is established; and second, the pattern of correspondence provides the non-retinotopic reference frame, against which local motion is computed. In other words, their effect is considered to be contingent on computation of corresponding perceptual groups.

In this study, we ask whether the effects Herzog et al. describe could be explained by the spatiotemporal properties of a motion energy model (Adelson & Bergen, 1985; Burr et al., 1986; van Santen & Sperling, 1985; Watson & Ahumada, 1985; for a review, see Burr & Thompson, 2011), which proved to be so successful in explaining other spatiotemporal effects. To this end, we simplified the Boi et al. stimulus and, in two experiments, showed that similar effects could be obtained with more basic stimuli, sinusoidal gratings, which permitted a more quantitative measure of the magnitude of the effects. We then carried out a 3D Fourier analysis of the stimulus display and modeled the results with a simple model, consisting of a linear filter operation (low-pass in time and space), followed by a non-linearity, analogous to previous "local energy" models (Morrone & Burr, 1988). The success of
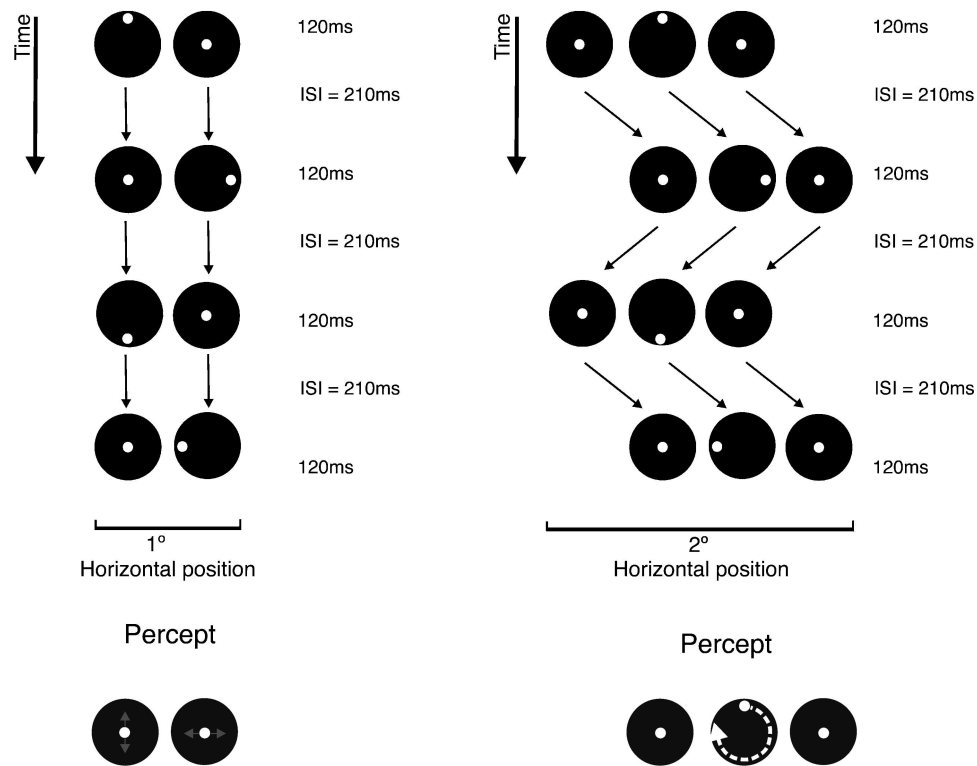
Figure 2. The visual stimuli used in Boi et al.'s (2009) study.

this method indicates that the "non-retinotopic" or motion contingent effects reported by Boi et al. may rely on the spatiotemporal properties of basic motion analysis.

Experimental methods

Six subjects participated in Experiments 1 and 2 (one of the authors and five laboratory members who were naive to the purpose of experiment). All subjects had normal or corrected-to-normal vision.

Subjects viewed the stimuli binocularly from a distance of 75 cm. The stimuli of Experiment 1 were generated with a VSG 2/3 graphics card (Cambridge Graphics) and displayed on a Barco calibrated CRT monitor. The stimuli of Experiment 2 were generated in Psychophysics Toolbox (Brainard, 1997) and were displayed on a Barco calibrated CRT monitor driven by an Acer Veriton M670G workstation with an ATI 4350 graphics card. In both experiments, the resolution of the monitor was 1280 * 1024 with a refresh rate of 85 HZ.

Experiment 1: Contrast threshold of the non-retinotopic motion integration

The stimuli of Experiment 1 were based on the logic that was used by Boi et al. (2009) as depicted in Figure 3a.

The phase of a grating is sequentially shifted across the frames and this progressive phase shift produces the percept of apparent motion in a certain direction. Importantly, the phase of the grating in the second frame (δ) is crucial to disambiguate the direction of motion. In Experiment 1, this phase δ was either $+90°$ or $-90°$. When δ is $+90°$, the progression of phase from $0°$ to $+90°$ to $180°$ to $-90°$ is perceived as upward drift, while at $\delta = -90°$, phase shift occurs from $0°$ to $-90°$ to $180°$ to $90°$ and, therefore, a downward drift is perceived. An example of target bar drifting upward is shown in Movie 2. Note that the position of the target bar is horizontally shifted back and forth across the frames. Therefore, in order to perceive the vertical drift, integration of the phase shifts has to occur across retinotopically non-contiguous locations as indicated by the path along the dashed arrows of Figure 3. In such a display, the strength of the motion signal can be manipulated by varying the luminance contrast of the sinusoidal grating.

The target grating was embedded in two types of stimulus displays that consisted of either two or three bars (Figure 3b, Movies 3 and 4) that were presented in separate blocks. In the two-bar displays, on the first frame, the target grating was on the fixation point and the other flanking bar was positioned either to the left or to the right of the target (randomized across trials). On subsequent frames, the target bar exchanged its horizontal position with the flanking bar. Perceptually, in two-bar displays, no apparent group motion was perceived, and instead, the target grating appeared to swap its position with the
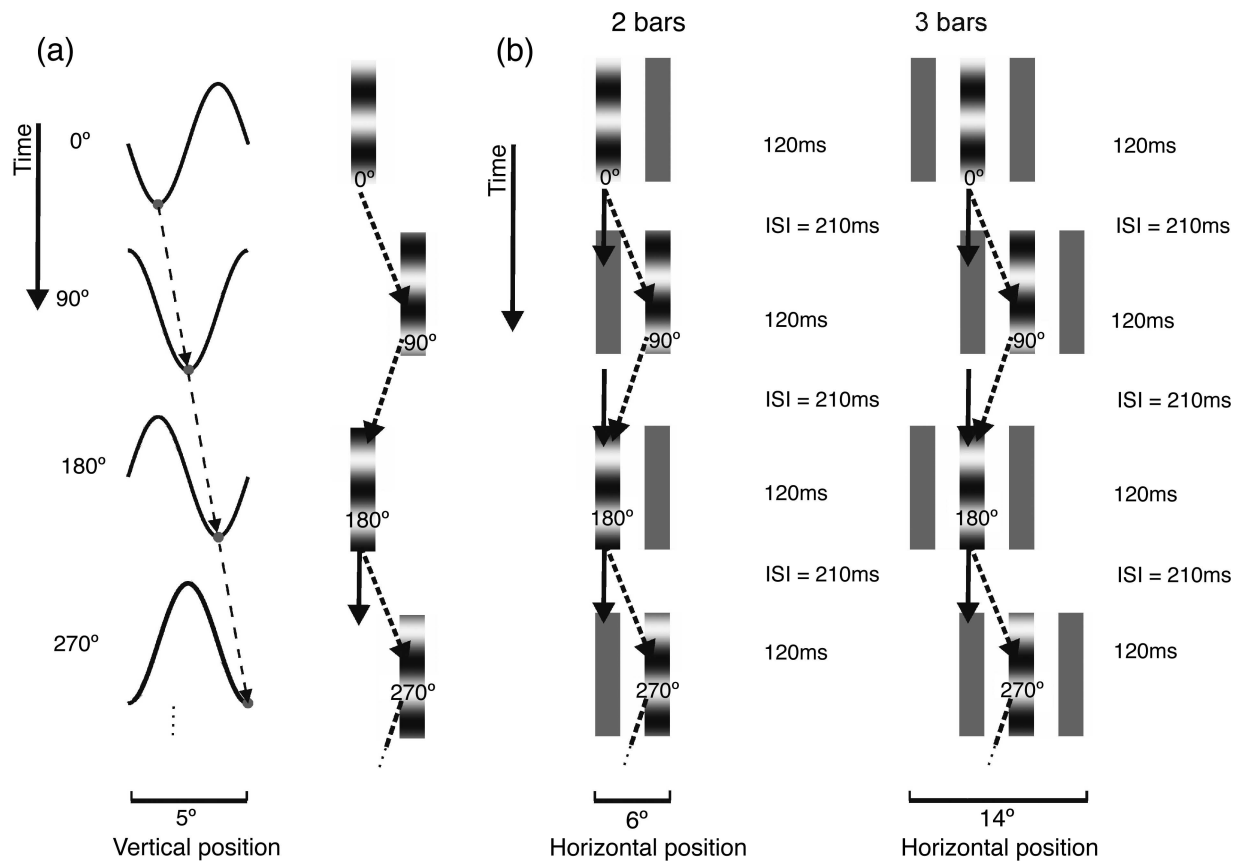
Figure 3. The stimuli of Experiment 1. (a) Illustration of the target grating across the frames. The traces at the left show the luminance profiles of the target grating, progressively shifting in phase. If we trace a given point on the sinusoid (marked here in gray), we can appreciate that as the phase of the sinusoid is shifted the location of this point changes in space, and thus, apparent motion is perceived. It is important to note that since the phase of the 1st and 3rd frames is always at 180°, the phase of the second frame (δ) is crucial in disambiguation of the motion. In this experiment, the phase δ in odd frames was either +90 ° or −90° (producing upward and downward motion, respectively). Note that the position of the target grating was horizontally shifted back and forth across the frames, so to perceive the vertical drift, information had to be integrated along a non-retinotopic path as indicated by the dashed arrows. (b) Stimulus displays. Subjects had to judge the drift direction of the target bar when it was presented in either 2- or 3-bar displays. The strength of the motion signal was manipulated by modulating the contrast of the grating above the mean luminance level.

adjacent bar (Movie 3). The three-bar displays were like the two-bar displays, with an additional bar. This third bar was displaced from a location to the left of the target bar in one frame to the right of the target bar in the next frame. In these displays, the bars were perceived to move horizontally as a group at the ISI (120 ms) that we used (Movie 4).

The bars subtended 2° horizontally and 10° vertically, separated from each other by 2°. The horizontal shift of the bars across the frames was equal to one inter-bar distance (i.e., 2°) across the frames. The stimuli were displayed on a white background (luminance of 60 cd/m^2). The mean luminance of the bars was about 28 cd/m^2. The luminance of the target grating (0.4 cpd) was modulated around the pedestal luminance (luminance of 28 cd/m^2) to produce different contrast levels. The flanking bar was identical to the target grating except that it did not contain a spatial luminance modulation and was simply gray with a luminance of 28 cd/m^2.

In Experiment 2, the bars subtended 1.2° horizontally and 3.2° vertically and were spaced 1.2° apart from each other (equal to the width of a bar). The stimuli were displayed on a gray background (luminance of 12 cd/m^2). Each bar had the same luminance as the background and contained a horizontal sinusoidal grating (0.6 c/deg, 50% Michelson contrast) inside it. The phase δ of the target grating was varied with a QUEST staircase method and subjects' performance was measured as a function of δ. All other details are the same as for Experiment 1.

A trial began with the appearance of the stimulus at the center of the screen. The stimulus sequence comprised 8 stimulus frames: two cycles of the motion sequence. Each stimulus frame was presented for 120 ms and was followed by a blank of 210 ms, corresponding to a drift frequency of 0.7 Hz. Subjects were instructed to maintain fixation on a central fixation point (black square, 0.5°) throughout a trial and indicate whether the target grating drifted upward or downward. When the presentation of

the stimulus frames was over, subjects indicated the direction of the drift by pressing page-up or page-down keys to indicate upward or downward directions, respectively. No feedback was provided to the subjects. The performance of the subjects was measured as a function of the luminance contrast of the target grating that was varied across the trials by using a QUEST staircase method. The psychometric function of each subject was fitted by a cumulative Gaussian function whose mean determined the threshold (the minimum contrast to bias performance in the upward (or downward) direction by 81%), standard deviation, and the steepness of the psychometric function (determining the noisiness of the judgment).

Results

In Experiment 1, our main question was whether the strength of motion signals and, hence, the contrast thresholds to detect the vertical drift were different in two- and three-bar displays. Figure 4a displays psychometric functions for three typical subjects. It can be seen that in the two-bar display subjects performed poorly with rather high-contrast thresholds (29%, 49%, and 57% for AP, MT, and MC), whereas in three-bar displays their contrast threshold was greatly decreased (11%, 13%, and 20% for AP, MT, and MC). This finding was consistent across all subjects (Figure 4b): In two-bar displays, the contrast thresholds were twice as high than the three-bar displays, clearly significantly different (mean thresholds: 61% and 34% with two and three bars, respectively, one-tailed paired t-test, $t = 4.57$, $p = 0.003$). The difference between these displays cannot be accounted by a difference in task difficulty, since the space constant (steepness of the psychometric functions) was similar in all cases and similar to what is normally observed (mean sigma: 0.15 and 0.16 log units in two- and three-bar displays, not significantly different: $t = 0.92$, $p = 0.20$).
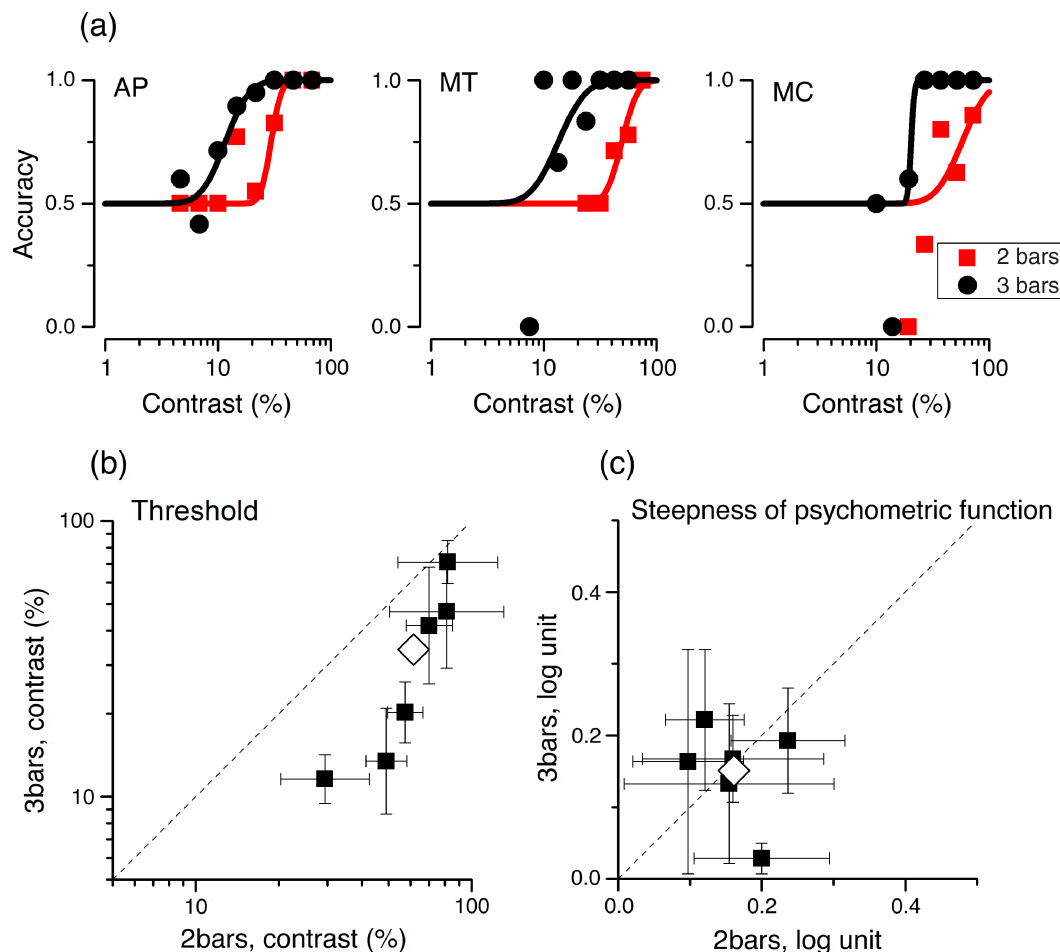

Figure 4. Psychophysical results of Experiment 1. (a) Psychometric functions of 3 subjects in two- and three-bar displays (red and black, respectively). In three-bar displays, the psychometric curves are shifted to the left indicating lower contrast thresholds to detect motion. (b) The thresholds and (c) steepness of psychometric functions (space constant of fitting Gaussian) of three-bar displays plotted against those of two-bar displays (errors show ±1 standard error of the mean, obtained by bootstrap). Every single subject showed a lower threshold in the 3- than the 2-bar condition, but the curve steepnesses were comparable in both conditions.

Experiment 2: The influence of the "non-retinotopic" signals on "retinotopic" motion integration

In Experiment 2, we tested directly the so-called "non-retinotopic" processing, by pitting retinotopic against non-retinotopic signal integration. To this end, we added motion signals along the retinotopic path (solid arrows in Figure 5), in addition to those in the non-retinotopic motion integration path (dashed arrows in Figure 5). The retinotopic motion signals were produced by inserting in frames 1 and 3 a grating of specific phase at the location corresponding to the retinotopic position of the gratings in frames 2 and 4. The retinotopic grating was placed behind the fixation point and the subjects were instructed to fixate throughout the trial and report the direction of the drift that occurred at the fixation spot. The strength of the retinotopic motion was manipulated by varying the phase of the grating (δ) on odd frames (Figure 5a). The absolute value of the phase determines the motion strength. Movies 5 and 6 illustrate stimuli with a $\delta = 90°$ where a clear upward motion can be seen; as $|\delta|$ becomes smaller than 90°, the strength of motion (upward or downward) is reduced. If the "non-retinotopic" signal is integrated with the "retinotopic" signal even when the phase of the retinotopic grating is zero ($\delta = 0$) and the grating is just flickering, due to the progression of phase along the non-retinotopic path (0°, 90°, 180°, 270°), an upward bias could be produced. This condition is portrayed in Movies 7 and 8 where $\delta = 0$ and a non-retinotopic bias is present. Whereas with three-bar stimuli (Movie 8) the upward bias is readily visible, with the two-bar stimuli (Movie 7) motion is
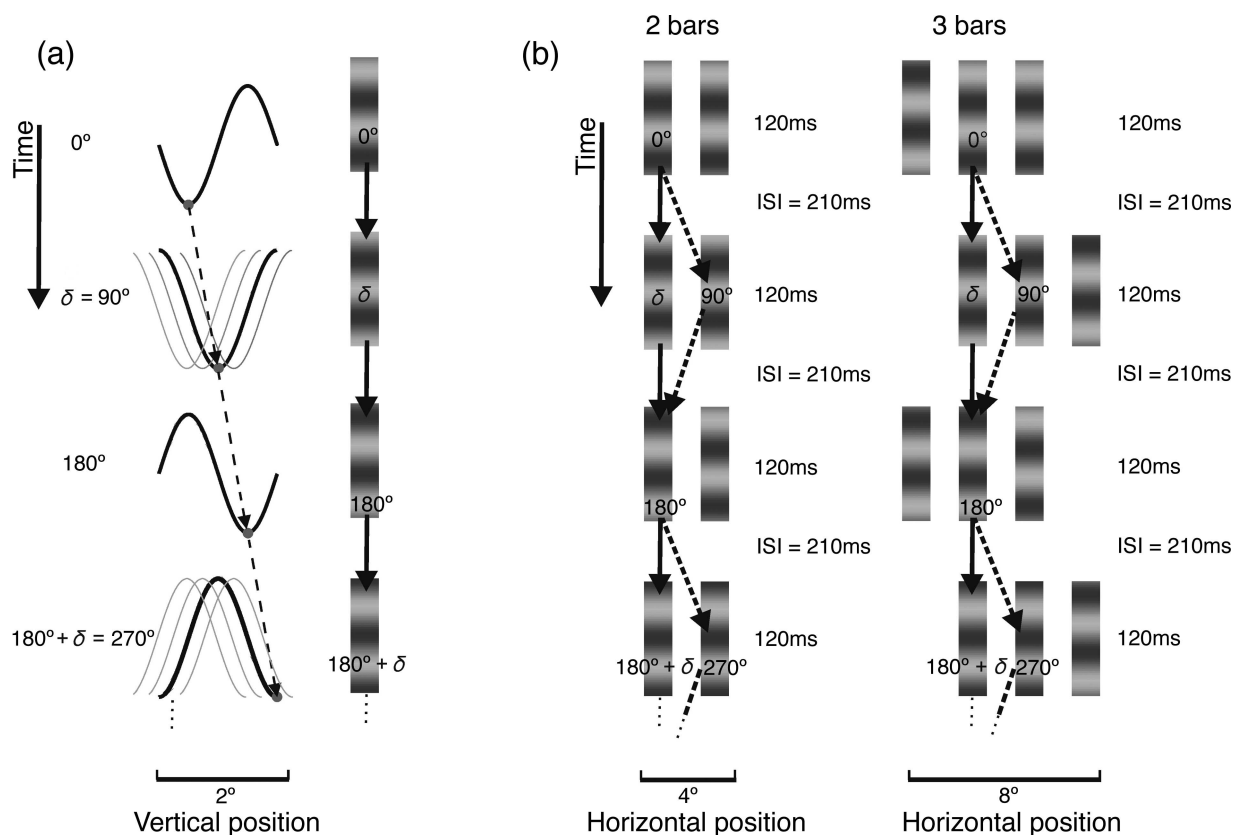

Figure 5. The stimuli of Experiment 2. (a) Illustration of the target grating across frames. Traces at the left show luminance profiles of the target grating, showing how the phases are displaced each frame. The dashed line shows the progress of a given point on the sinusoid. Since the phase of the 1st and 3rd frames is always 180°, the phase of the second frame (δ) is crucial for disambiguation of the motion. Smooth drift was perceived when the phase δ in odd frames is either +90° or −90° (producing upward or downward motion, respectively), whereas smaller values of δ produce a less clear perception of upward or downward motion. Thus, by varying the phase of the grating at this frame (as shown by the thin, light traces), the direction and the strength of the motion can be manipulated. (b) Stimulus displays. Subjects had to judge the drift direction of the target bar when it was presented in either 2- or 3-bar displays. The target grating was presented at the same retinotopic location across frames and completed a cycle in 4 frames. On each trial, 8 frames (2 drift cycles) were displayed. The phase of the grating adjacent to the target was set to 90° and 270° in odd frames. If motion integration occurs along the non-retinotopic path (dashed arrows), the subject would be biased to perceive upward motion even when the phase δ of the target bar is 0. In the control variant of Experiment 2, the phase of the non-retinotopic grating was set to 0, and therefore, no bias in perception was expected.
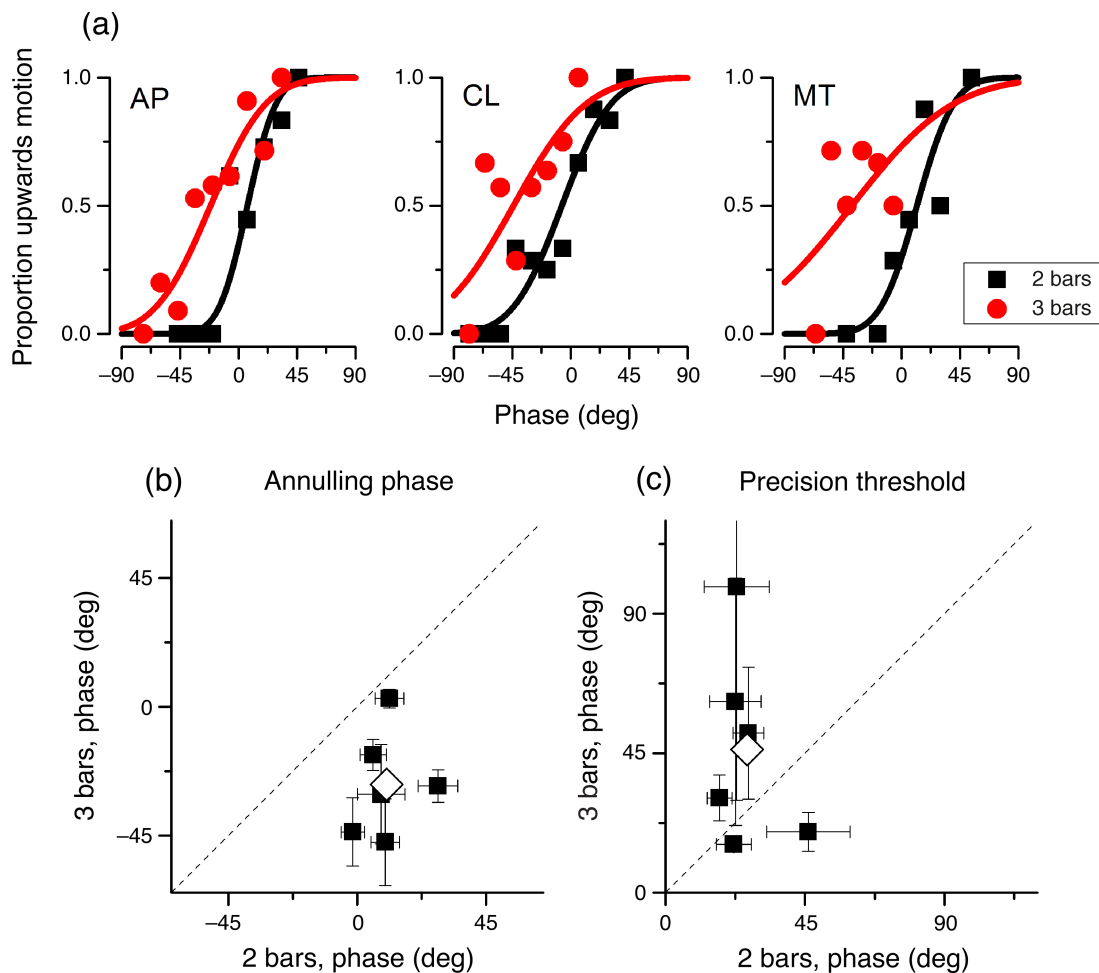
Figure 6. Psychophysical results of Experiment 2. (a) Psychometric functions of 3 subjects in two- and three-bar displays (red and black, respectively). Whereas in two-bar displays the curves are centered on 0, in three-bar displays the annulling points are shifted to negative values. (b) Annulling points of all subjects in two- and three-bar displays. (c) Precision thresholds for the judgments with two- and three-bar displays.

rather ambiguous. We measured this bias by manipulating δ and finding the phase that annulled the bias introduced by the non-retinotopic drift. In a control variant of Experiment 2, we tested the stimuli in which this pattern of non-retinotopic phase integration (along the dashed arrows) could not occur since the phase of the flanking grating always remained at 0°. The stimuli of Experiment 2 in which a grating (i.e., the equivalent of the dot of Figure 1) was present inside all the bars more closely resembled the stimuli of Boi et al.

Results

The psychometric functions of the 3 subjects of Experiment 2 are shown in Figure 6a. In two-bar displays, the probability of detecting upward motion increased as δ was increased from −90° to 90°. Importantly, at $\delta = 0$, almost no vertical drift was perceived. This was quite different from the performance of the subjects in three-bar

displays where subjects tended to perceive the grating to drift upward even at $\delta = 0$. When the grating flickered, subjects showed a strong bias for perceiving upward motion as demonstrated by negative PSEs (annulling point). Therefore, when the phase of the non-retinotopic grating was set to the values that could produce an upward bias, the three-bar displays were affected by this bias whereas the two-bar displays were not influenced.

This pattern of results occurred consistently across all the subjects (Figure 6b). In three-bar displays, the annulling points were shifted toward negative values (mean = −27.2). On the other hand, the two-bar displays were not affected by this bias since the annulling point was positive (mean = 10.2). The difference between two- and three-bar displays was significant (one-tailed paired t-test, $t = 4.26$, $p = 0.004$).

To verify that the difference between the two- and three-bar displays was indeed due to the integration of the non-retinotopic grating with the main grating, we next
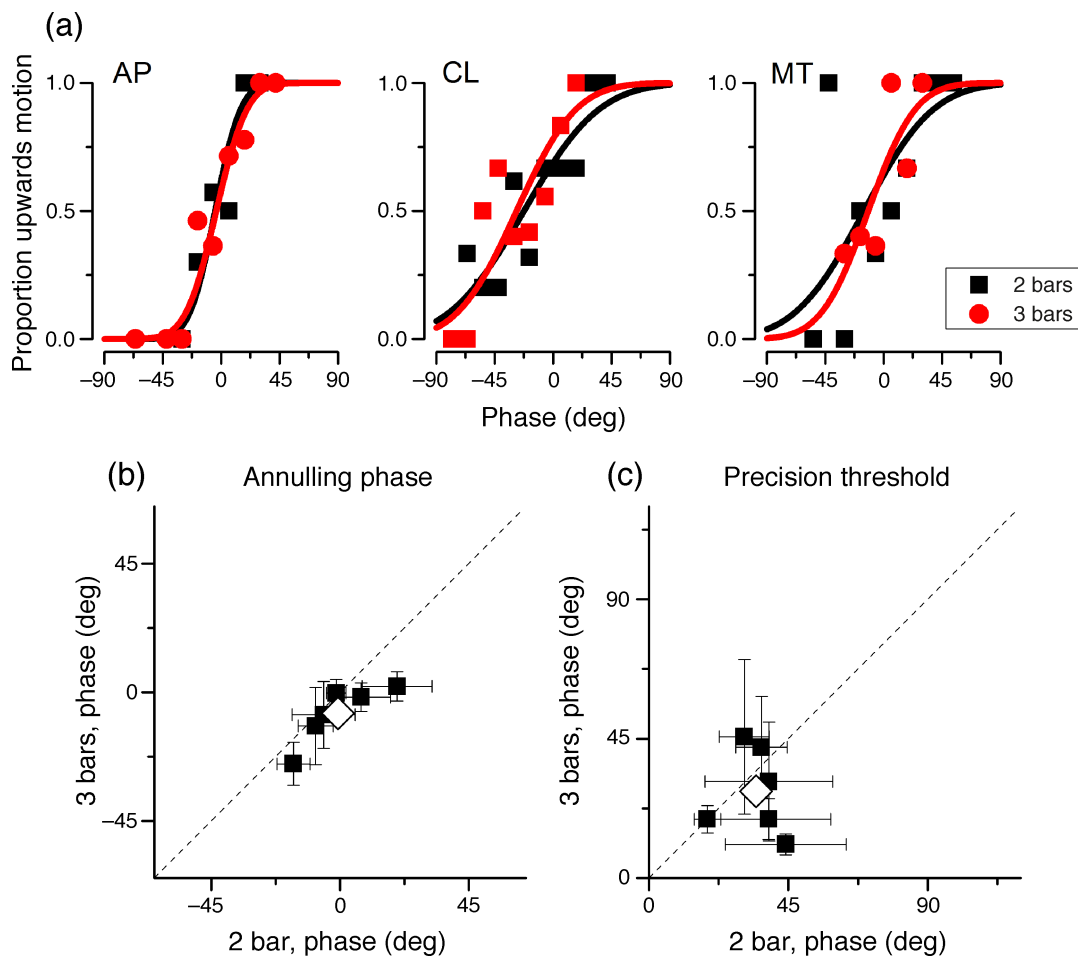
Figure 7. Psychophysical results of the control variant of Experiment 2. (a) Psychometric functions of 3 subjects in two- and three-bar displays (red and black, respectively). Unlike the main experiment, in both types of displays, the curves are centered on 0. The upward bias of the three-bar display is largely reduced. (b) Annulling points of all subjects in two- and three-bar displays. (c) JNDs of two- and three-bar displays.

tested the condition where the phase of the non-retinotopic grating was set to zero (Figure 7). In this control variant of Experiment 2, the difference between two- and three-bar displays completely vanished as can be seen by comparing the psychometric functions (Figure 7a). Across all subjects, the negative shift of the annulling point in three-bar displays was largely reduced (mean = −7.45), while the annulling point of the two-bar display remained around zero (mean = −0.79). The difference between the two- and three-bar displays was not significant in this condition (one-tailed paired *t*-test, $t = 1.87$, $p = 0.06$).

Comparison of precision thresholds provides a measure of the task difficulty (Figures 6c and 7c). The mean precision threshold of the two- and three-bar displays was 26° and 46° in the main version of Experiment 2 and 34° and 28° in the control variant. The lower precision thresholds in three-bar displays of the main experiment were perhaps due to the strong interference of the non-retinotopic with the target grating that rendered the

perceptual decision about drift direction difficult. Importantly, we did not observe a significant difference between the precision thresholds in two- and three-bar displays either in the main or in the control experiment (paired *t*-test, *P*s > 0.05). Therefore, we can be confident that a difference in the task difficulty cannot account for the difference between two- and three-bar displays in the main experiment.

Taken together, the results of our psychophysical experiments show that the non-retinotopic influences are quite different in two- and three-bar displays, presumably because the three-bar displays evoke a strong sense of apparent motion. We next examined whether a model based on extraction of motion energy by oriented spatiotemporal filters can account for the observed difference between these stimulus displays. We applied the local motion energy model to simulate the data of Experiment 2 that were more similar to the litmus stimuli of Boi et al. We did not attempt to model the contrast

thresholds of Experiment 1, which would necessarily require a non-linear contrast response and threshold stage, unnecessarily increasing the complexity of the model.

Computation of the motion signals from the 3D Fourier spectrum

We computed the 3D Fourier spectrum of the stimuli of Experiment 2. Each stimulus comprised 64 temporal frames simulating 2 full temporal cycles either for 2- or 3-bar conditions. Each frame containing the sinusoidal grating was followed by 7 blank frames mimicking the experimental timeline of 120-ms stimulus exposure, followed by 210 ms of blank. The phase of the test grating was shifted every 8 frames. The size and spatial relations of the stimuli were similar to the stimuli of Experiment 2.

Figure 8 illustrates how the direction of an upward drifting motion can be inferred from the 3D Fourier spectrum. The stimulus in the frequency domain is defined in the three-dimensional space ω_x, ω_y, and ω_t corresponding to spatial and temporal frequencies, respectively. The quadrant of this spectrum with $\omega_t > 0$ and $\omega_y > 0$ represents the upward drift while the quadrant with $\omega_t > 0$ and $\omega_x < 0$ represents downward motion. Given that two cycles of drifting grating have been simulated, the plane $\omega_t = 2$ represents the fundamental frequency of the motion. Figure 8a shows the spectrum of a three-bar display, where the target bar moves upward. It can be seen
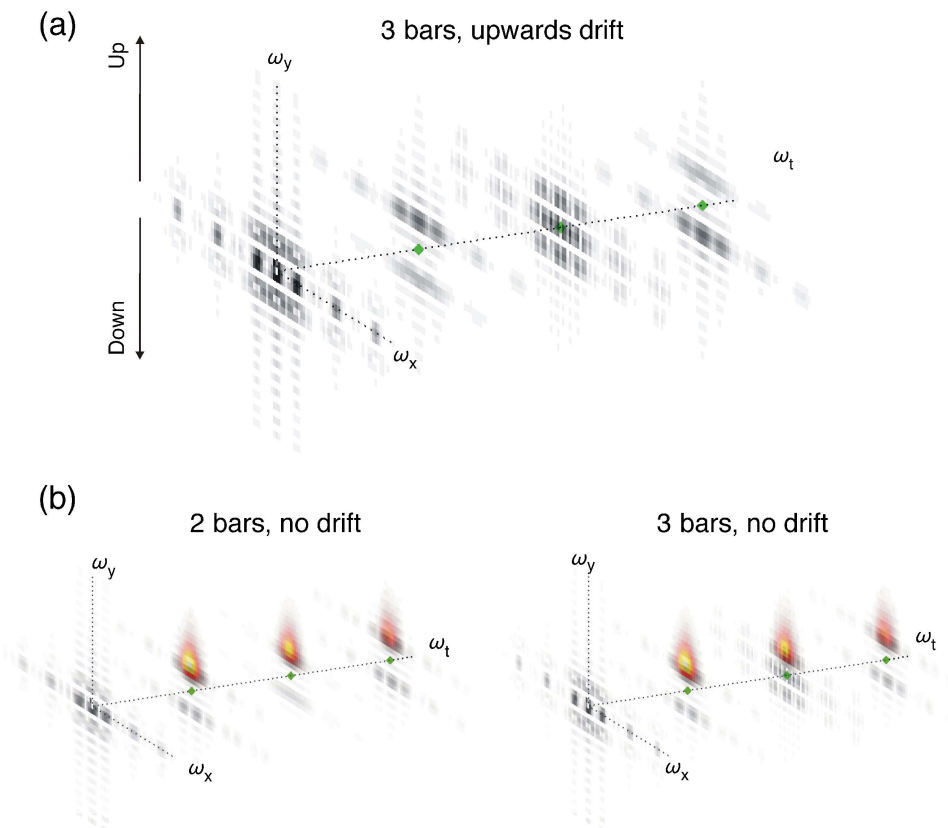

Figure 8. Illustration of the 3D Fourier spectra. (a) To illustrate how the direction of motion can be inferred from the 3D Fourier spectrum, the spectrum of an upward drifting stimulus ($\delta = 90°$) with three bars is depicted. The signal in the frequency domain is defined as ω_x, ω_y, and ω_t, corresponding to spatial frequencies in x, y, and t, respectively. The spectrum is truncated between the central plane $\omega_t = 0$ and $\omega_t = 6$ and only planes corresponding to $\omega_t = 0$, $\omega_t = 2$, $\omega_t = 4$, and $\omega_t = 6$ are shown. Since two cycles of the stimulus are simulated, all odd planes of ω_t are blank and are not shown in the figure. Thus, the 4 planes shown are the even planes, corresponding to the fundamental frequency and higher harmonics (0, 0.7, 1.4, and 2.8 Hz, respectively). The upward direction of motion corresponds to the volume for $\omega_y > 0$ while the downward motion for $\omega_y < 0$. For example, in the plane $\omega_t = 2$, corresponding to the fundamental frequency, two high amplitude peaks are present at the spatial frequency of the horizontal grating, but this signal is stronger for upward motion (i.e., the band in $\omega_y > 0$). (b) Sample 3D spectra of the two- and three-bar stimuli in the main experimental condition with δ set to 0 (no upward drift). An example of the directional filter for upward motion is overlaid on the spectra with hot colors. A similar filter computed downward motion, and the outputs of the two directional filters were then compared (see Figure 10).

that the $\omega_t = 2$ plane contains two peaks of highest amplitude: the peak for positive ω_ys (corresponding to the upward motion) is more intense than the peak for the negative ω_ys (corresponding to the downward motion). Hence, higher positive or negative peaks indicate upward or downward motion.

Figure 8b shows 3D Fourier spectra for two- and three-bar stimuli with phase $\delta = 0$. Compared to the spectrum of Figure 8a, here the spectra are more balanced between upward and downward directions. However, close inspection reveals that in the plane of the second temporal harmonic (i.e., the third plane shown in Figure 8b), the distribution for upward and downward motion is different for the two- and three-bar stimuli. We therefore employed a directional filter tuned to very low spatial frequencies along ω_x and broad in temporal frequency to embrace both the 1st and higher harmonics (and the beats between them). An example of such a filter is overlaid on the spectra of Figure 8b in red.

The directional filter is a band-pass filter for ω_y and ω_t and is given by the product of two band-pass lognormals and a Gaussian function:

$$f_x(\omega_x) = e^{-\frac{\omega_x^2}{\sigma_x^2}}$$

$$f_y(\omega_y) = e^{-\frac{\left(ln\left(\omega_y/\mu_y\right)\right)^2}{(ln(\sigma_y))^2}} \tag{1}$$

$$f_t(\omega_t) = e^{-\frac{(ln(\omega_t/\mu_t))^2}{(ln(\sigma_t))^2}},$$

where μ_y and μ_t represent the peak of the filter, σ_y and σ_t define the falloff factor where attenuation reaches 37%, and σ_x is the filter falloff in ω_x. For upward motion, the filters are defined only for positive ω_y and ω_t and are zero elsewhere. For downward motion, the filters are defined only for negative ω_y and ω_t and are zero elsewhere.

The filter falloff in ω_y and ω_t was centered on the region with highest energy ($\omega_y = 4$, $\omega_t = 2$) and falloff factors were 2 and 4, respectively. A crucial parameter is the spatial spread along x, and we tested a set of filters by varying σ of Equation 1.

The real and the imaginary parts of the inverse Fourier transform of the filters of Figure 8b is shown in Figure 9, corresponding to an even- and odd-symmetric quadrature filter pair. The motion selectivity is given by the orientation of the filter in y–t plane.

The direction of motion was calculated by first convolving the stimulus with two sets of even- and odd-symmetric operators (from Equation 1; Figure 9), tuned for upward and downward motion, and then computing the Pythagorean sum (motion energy) of their output for each direction. The difference in the amplitude of the

highest peaks for the upward and downward motion energy (over the entire time series) defines the bias in perceived direction. Operatively, as shown in Figure 10, the spectra were multiplied by the two pairs of filters, setting the negative quadrant to zero. The signals were then back transformed into the space–time domain (inverse Fourier transform) and calculating the complex absolute values.

Figure 11 shows the output of the model for 12 filters of width (wd) varying between 0.5 and 6.11 degrees. For each filter, we computed the motion bias for phases ($-90° < \delta < 90°$) of the grating and normalized this to the maximum bias, where positive values indicate upward motion and negative values indicate downward motion. For filters larger than 1 degree, there was a clear difference in the output of two- and three-bar stimuli: The annulling phase, where no net motion was observed, was always more negative for the three-bar condition, mimicking the psychophysical results. Importantly, the difference between the two- and three-bar displays was not detectable in the average rather than the global maxima of the upward and downward energies (as shown in the inset of filters with wd = 1.22 to 0.76 in Figure 11a). This indicates that a non-linear operation on the energy output is necessary to discern the difference between the two- and three-bar display outputs.

Finally, we directly compared the output of the model with the psychophysical data (Figure 11c) by reconstructing the psychometric function from the mean values of the PSE (annulling point) and precision thresholds across subjects. The psychometric curve of the three-bar display is shifted to the left, reflecting a strong upward bias, while this was not the case in two-bar displays. To simulate the results for the three-bar stimuli, a filter of width greater than 1 is needed, while for the two-bar displays (red curves of Figure 11c) a filter of width of about 0.67 is more appropriate. This seems a reasonable approach, as the 3-bar displays were 1.5 times larger than the two-bar displays.

Discussion

We used a modified version of Boi et al.'s stimulus to investigate "non-retinotopic" motion effects. In Experiment 1, we found that contrast thresholds to detect the vertical drift of the target bar were twice as low in three-bar Ternus display compared with two-bar, non-motion displays. In Experiment 2, we showed that the vertical drift of a grating presented to the same retinotopic location across the frames is influenced by the signal from non-retinotopic locations. Importantly, this bias occurred only if a clear motion signal (in this case, upward bias) was present at the non-retinotopic locations.
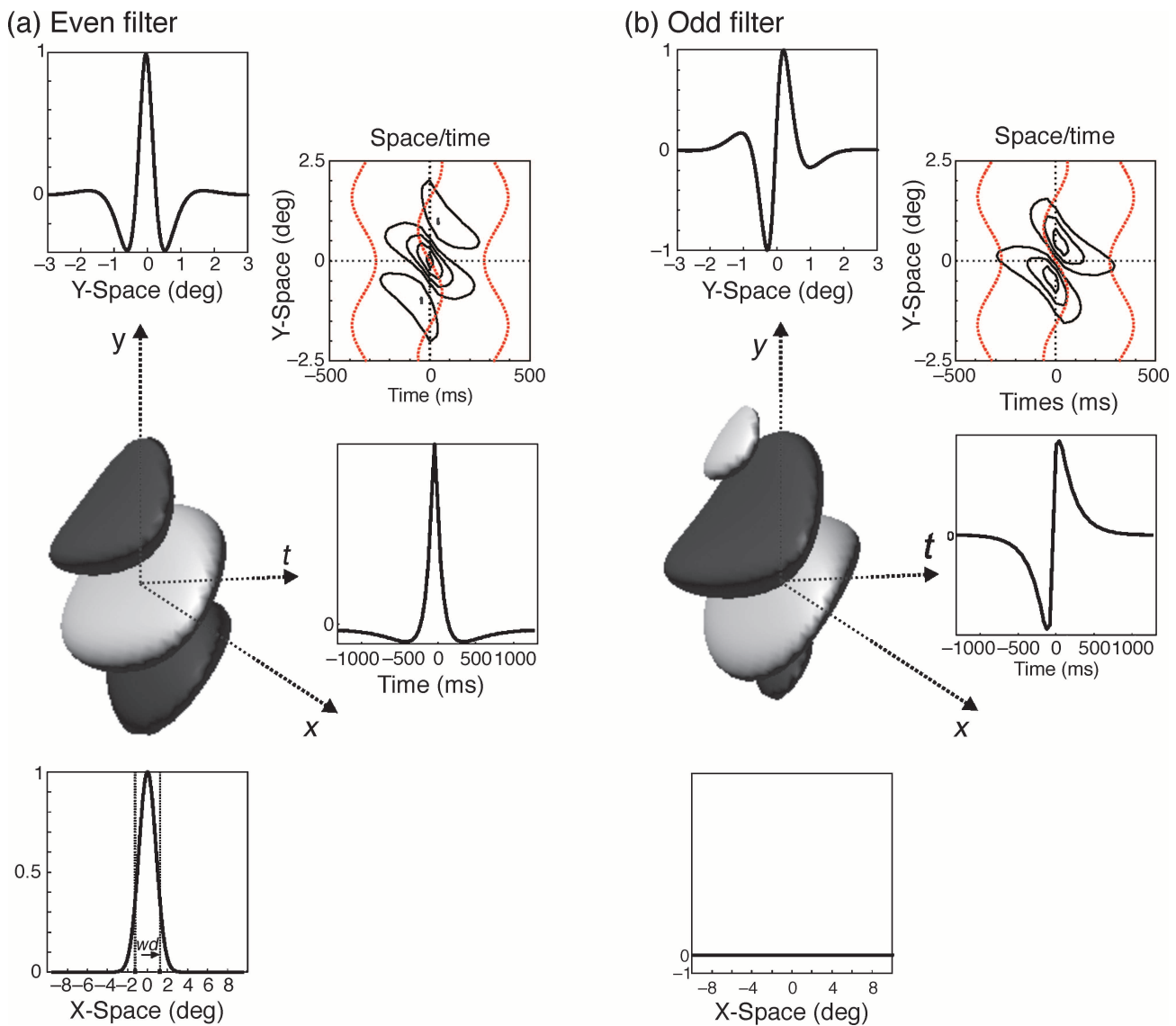
Figure 9. Space–time representation of the filters. The (a) even- and (b) odd-symmetric impulse response function of the filters of Figure 8b (red surface), represented as a 3D surface and as a cross section along the axes and for the plane *x* = 0. The 2D plot illustrates the orientation in space–time of the detectors. To illustrate the relative size of the filter with respect to the stimulus, we have overlaid the stimulus (red sinusoid) in space–time on the surface plots. The width of the filter along *X*, wd, is a crucial parameter to simulate the data.

The influence of the non-retinotopic signal on retinotopic processing was stronger in three-bar displays compared with two-bar displays. As such, our psychophysical results are in line with the observations of Boi et al. and provide further evidence for an interaction between retinotopic and non-retinotopic processing mechanisms.
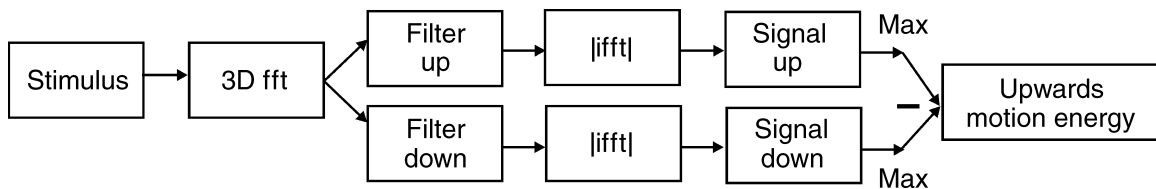

Figure 10. Computational steps to extract the motion direction: First, the 3D Fourier spectrum of the stimulus in *x*, *y*, and *t* is constructed. The spectrum is then filtered by directional filters, and upward or downward motion energy was computed. The local maxima over space and time in each direction are measured. The net motion direction is considered proportional to the difference between the strongest peak in the two directions over all frames.
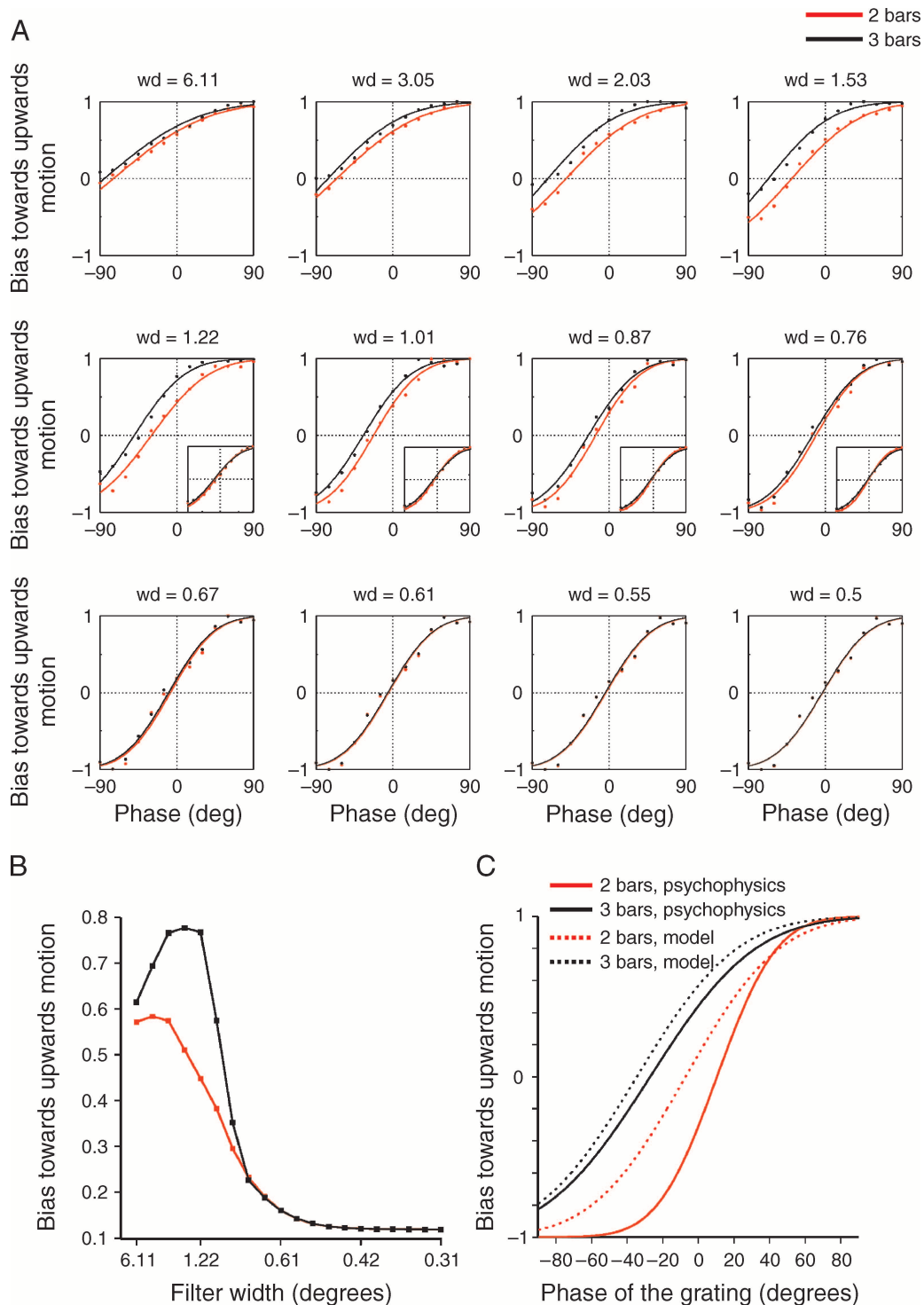
Figure 11. The output of the model for filters that vary in size along the horizontal dimension. (a) Bias toward upward motion, computed as the difference in global maximum upward and downward energies, for filters of different horizontal width. Note that very large filters (wd > 1) show a difference between the three- and two-bar displays. The insets illustrate the performance evaluated from average rather than peak energy: This strategy fails to reveal any difference at any filter width. (b) The bias toward upward motion at phase δ = 0. For large filter widths, there is a far stronger bias for three-bar stimulus. (c) Model simulation of the behavioral results. The output of the model is shown for filters with size scaled to the ratio between the size of two- and three-bar displays (i.e., wd = 0.67 and 1 degrees, respectively).

We point out that we observed a small bias in behavioral responses (Figure 6) in Experiment 2 that may correspond to a small bias present in the stimulus display (Figure 5). Although in our stimuli the phases at each spatial location were counterbalanced, a broad low-pass filter would pick up a small phase progression caused by the blending of the neighboring sinusoids, even when the phase of target grating was zero ($\delta = 0$). We believe that this small baseline bias may explain why, even in two-bar display at phase zeros, some subjects have a small bias toward upward motion. Interestingly also, our simulations seem to pick up this residual motion energy, as shown by an upward bias at phase zero in larger filters (see Figures 11a and 11b). Nevertheless, it is unlikely that this tiny bias is the source of the far stronger upward bias that was observed in the three-bar displays.

What are the underlying mechanisms of these effects? Boi et al. proposed a two-stage model where first the grouping rules for the horizontal motion of the Ternus–Pikler displays are established and then a non-retinotopic framework is provided for the analysis of the apparent rotary motion of the dot. We asked whether this is necessary to account for the observed effects and tested the hypothesis that the non-retinotopic effects could be explained by basic motion computation models, where motion is analyzed by optimally tuned filters oriented in space–time. The model we proposed extracts the directional motion energy by directional filtering of the 3D Fourier spectrum of the stimuli (Figures 8–10). The output of this simple model reproduced the difference between the two- and three-bar displays in psychophysics (Figure 11). The three-bar displays were associated with higher upward signals that required more negative phases to annul. Obviously, the output of the model depended on the width of the filter. If the filters were chosen to be of comparable size as the displays, the model provided a very good match to the psychophysics.

The filters that we used are equivalent to spatiotemporal detectors that have been previously used to extract motion energy (Adelson & Bergen, 1985; Burr & Ross, 1986; Heeger, 1987; Simoncelli & Heeger, 1998; Watson & Ahumada, 1983). Using the psychophysical technique of masking, Burr et al. (1986) computed the spatiotemporal profile of a hypothetical motion detector: Receptive field slanted in space–time to integrate motion signals over space and time, with spatial extent ranging from 2 arc sec to 7 degrees (Anderson & Burr, 1989). Our modeling results are in close correspondence with these previous findings. The estimated size of the motion detector in our study was >0.61 degree, which falls in the reported range of Anderson and Burr (1989).

There are, of course, several differences between the spatiotemporal interpolation studies of Burr et al. (Burr & Ross, 2004; Burr, 1979; Burr & Ross, 1979; Burr et al., 1986) and the Ternus displays used here, similar to those of Boi et al. One obvious difference is whereas the apparent motion of Burr and Ross' study was smooth and

compelling (see Movie 1), that of the Ternus display is less compelling and occurs best with fairly long time constants, corresponding to "long-term" motion in Braddick's (1980) classic distinction. However, this does not affect the principle that spatiotemporally oriented detectors can be activated by these stimuli, and their activation can explain the results obtained. Our filters were tuned to quite low temporal frequencies, with a peak at 0.7 Hz, not outside the range of temporal tuning demonstrated by early masking studies (Anderson & Burr, 1985). Interestingly, in our modeling, we found it necessary to introduce a non-linearity to account for our results (Figure 11a). This accelerating non-linearity (consider only maxima) is similar to the algorithms previously proposed to explain the detection and tracking of the features in human visual system (Del Viva & Morrone, 1998, 2006; Morrone & Burr, 1988). We chose to consider only peaks, a "winner take all" approach, but in practice any accelerating non-linearity (such as summing the squared output or other form of Minkowski sum) would have yielded very similar results. We believe that the processing of a complex stimulus such as the Ternus displays in our study may require this non-linearity and a moment-by-moment integration of information in order to decode the direction of motion.

Motion of objects can be computed by neural mechanisms with large receptive fields, oriented in space–time. In this framework, features of a moving object are defined as peaks in local motion energy (Del Viva & Morrone, 1998), so the motion and analysis of individual features are not separate mechanisms but are tightly linked to each other. On the other hand, previous models proposed to account for non-retinotopic effects in Ternus–Pikler displays treat the computation of motion and the detection of the features and attribution or integration of features as separate entities (Ogmen, 2007). This necessitates multiple processing stages generically parceled in difficult-to-define terms such as "non-retinotopic memory" (Ogmen, 2007; Scharnowski, Hermens, Kammer, Ogmen, & Herzog, 2007) and optional attention (Otto, Ogmen, & Herzog, 2010). Our model, on the other hand, is based on simple mechanisms known to exist in early motion-processing areas. It should be stressed that the model proposed here is an *existence proof*. We are not suggesting that this model is the only one that could account for the results, nor do we wish to make precise predictions of where the neural substrate for these filters may lie. We have tried to model the effects with the simplest mechanism possible, a single spatiotemporal filter followed by a basic non-linearity. Obviously, with a more complex neural network considering a population of neurons, it would be possible to simulate more precisely the data observed here and, by extension, those using the more complex stimuli of Boi et al. Our goal was not to produce the best possible model but to demonstrate that very simple and well-studied neural mechanisms—filters tuned in space–time—predict these types of non-retinotopic results. These filters can act to

create *spatiotemporal interpolation*, along the lines suggested by Barlow 30 years ago, both allowing for a fine-grain analysis of objects in motions and causing many interesting motion illusions.

References

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

Anderson, S. J., & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research, 25,* 1147–1154.

Anderson, S. J., & Burr, D. C. (1989). Receptive field properties of human motion detector units inferred from spatial frequency masking. *Vision Research, 29,* 1343–1358.

Barlow, H. B. (1979). Reconstructing the visual image in space and time. *Nature, 279,* 189–190.

Boi, M., Ogmen, H., & Herzog, M. H. (2011). Motion and tilt aftereffects occur largely in retinal, not in object, coordinates in the Ternus–Pikler display. *Journal of Vision, 11*(3):7, 1–11, http://www.journalofvision.org/content/11/3/7, doi:10.1167/11.3.7. [PubMed] [Article]

Boi, M., Ogmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision, 9*(13):5, 5–11, http://www.journalofvision.org/content/9/13/5, doi:10.1167/9.13.5. [PubMed] [Article]

Braddick, O. J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 290,* 137–151.

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

Burr, D. (1980). Motion smear. *Nature, 284,* 164–165.

Burr, D., & Ross, J. (2004). Vision: The world through picket fences. *Current Biology, 14,* R381–R382.

Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985–2010. *Vision Research, 51,* 1431–1456.

Burr, D. C. (1979). Acuity for apparent vernier offset. *Vision Research, 19,* 835–837.

Burr, D. C., & Ross, J. (1979). How does binocular delay give information about depth? *Vision Research, 19,* 523–532.

Burr, D. C., & Ross, J. (1986). Visual processing of motion. *Trends in Neurosciences, 9,* 304–306.

Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 227,* 249–265.

Del Viva, M. M., & Morrone, M. C. (1998). Motion analysis by feature tracking. *Vision Research, 38,* 3633–3653.

Del Viva, M. M., & Morrone, M. C. (2006). A feature-tracking model simulates the motion direction bias induced by phase congruency. *Journal of Vision, 6*(3):1, 179–195, http://www.journalofvision.org/content/6/3/1, doi:10.1167/6.3.1. [PubMed] [Article]

Fahle, M., & Poggio, T. (1981). Visual hyperacuity: Spatiotemporal interpolation in human vision. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 213,* 451–477.

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A, 4,* 1455–1471.

Morgan, M. J., & Thompson, P. (1975). Apparent motion and the Pulfrich effect. *Perception, 4,* 3–18.

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London B: Biological Sciences, 235,* 221–245.

Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Current Biology, 14,* 830–839.

Ogmen, H. (2007). A theory of moving form perception: Synergy between masking, perceptual grouping, and motion computation in retinotopic and non-retinotopic representations. *Advances in Cognitive Psychology, 3,* 67–84.

Otto, T. U., Ogmen, H., & Herzog, M. H. (2006). The flight path of the phoenix—The visible trace of invisible elements in human vision. *Journal of Vision, 6*(10):7, 1079–1086, http://www.journalofvision.org/content/6/10/7, doi:10.1167/6.10.7. [PubMed] [Article]

Otto, T. U., Ogmen, H., & Herzog, M. H. (2009). Feature integration across space, time, and orientation.

Journal of Experimental Psychology: Human Perception and Performance, 35, 1670–1686.

Otto, T. U., Ogmen, H., & Herzog, M. H. (2010). Attention and non-retinotopic feature integration. *Journal of Vision, 10*(12):8, 1–13, http://www.journalofvision.org/content/10/12/8, doi:10.1167/10.12.8. [PubMed] [Article]

Scharnowski, F., Hermens, F., Kammer, T., Ogmen, H., & Herzog, M. H. (2007). Feature fusion reveals slow and fast visual memories. *Journal of Cognitive Neuroscience, 19,* 632–641.

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

Ternus, J. (1926). Experimentelle Untersuchungen über phänomenale Identität. *Psychologische Forschung, 7,* 81–136. Translated to English in W. D. Ellis (Ed.), *A Sourcebook of Gestalt Psychology.* New York: Humanities Press; 1950.

van Santen, J. P., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2,* 300–321.

Watson, A. B., & Ahumada, A. J. (1983). A look at the motion in the frequency domain. In J. K. Tsotsos (Ed.), *Motion: Representation and perception* (pp. 1–10). New York: Association for Computing Machinery.

Watson, A. B., & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, 2,* 322–341.