Current Biology, Vol. 14, 830–839, May 25, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.cub.2004.04.044

Motion-Based Analysis of Spatial Patterns by the Human Visual System

Shin'ya Nishida*
NTT Communication Science Laboratories
Nippon Telegraph and Telephone Corporation
3-1 Morinosato Wakamiya
Atsugi, Kanagawa, 243-0198
Japan

Summary

Background: It is known that the visibility of patterns presented through stationary multiple slits is significantly improved by pattern movements. This study investigated whether this spatiotemporal pattern interpolation is supported by motion mechanisms, as opposed to the general belief that the human visual cortex initially analyses spatial patterns independent of their movements.

Results: Psychophysical experiments showed that multislit viewing could not be ascribed to such motion-irrelevant factors as retinal painting by tracking eye movements or an increase in the number of views by pattern movements. Pattern perception was more strongly impaired by the masking noise moving in the same direction than by the noise moving in the opposite direction, which indicates the direction selectivity of the pattern interpolation mechanism. A direction-selective impairment of pattern perception by motion adaptation also indicates the direction selectivity of the interpolation mechanism. Finally, the map of effective spatial frequencies, estimated by a reverse-correlation technique, indicates observers' perception of higher spatial frequencies, the recovery of which is theoretically impossible without the aid of motion information.

Conclusions: These results provide clear evidence against the notion of separate analysis of pattern and motion. The visual system uses motion mechanisms to integrate spatial pattern information along the trajectory of pattern movement in order to obtain clear perception of moving patterns. The pattern integration mechanism is likely to be direction-selective filtering by V1 simple cells, but the integration of the local pattern information into a global figure should be guided by a higher-order motion mechanism such as MT pattern cells.

Introduction

According to a currently prevailing hypothesis, our brain processes spatial patterns and movements by separate subsystems, each starting with nondirection-selective and direction-selective neural sensors. This hypothesis is motivated by anatomical segregation of ventral and dorsal pathways, which respectively contain neural mechanisms suitable for the analysis of pattern and motion [1–3]. It is also supported by perceptual phenomena such as the perception of motion without the sense of pattern [4, 5] or the perception of pattern without the sense of motion [5, 6]. It is debatable, however, whether this parallel processing structure fully accounts for our successful perception of clear-moving patterns.

As opposed to the notion of separate processing, several forms of perceptual interaction between pattern and motion are known. Although most of them could be interpreted as occurring after separate analysis of the two attributes in different visual areas [7–12], a few suggest the existence of pattern-motion interactions at early processing stages. One is the influence of pattern orientation on motion perception, which suggests that the motion system may use motion streaks (speedlines) left by moving targets [13–17]. Another case, which this paper will address and which possibly indicates the complementary influence of motion on pattern perception, is spatiotemporal pattern interpolation [18, 19]. Consider an array of vertical slits, each having one pixel width. The slit separation is so wide that a stationary pattern shown through the slit array is hard to recognize. However, when a pattern is horizontally moving behind the stationary slit array, it can be effortlessly recognized in spite of only a fraction of the image being actually presented at any instant of time (see also [20–24]).

The spatiotemporal interpolation suggests that pattern perception may be facilitated by motion information. Specifically, it suggests that the direction-selective neural mechanisms may selectively integrate spatial pattern information that moves in the retinal image at the same speed [24–29]. Such a mechanism is generally useful to attain unblurred perception for moving patterns [26, 27, 30]. In spite of its theoretical importance, however, supporting evidence of this hypothesis is still sparse and conflicting [31–35].

In both electrophysiology and psychophysics, direction selectivity has been recognized as direct evidence of motion processing [36, 37]. To test the involvement of motion mechanisms in spatiotemporal pattern interpolation, I investigated the direction selectivity of the underlying mechanism by using both classical psychophysical techniques of masking and adaptation, which have been proven to be useful in assessing stimulus selectivity [38], and a modern psychophysical reverse correlation technique [39, 40], which is a powerful method for estimating the pattern representation in the observer's brain. The results altogether provide strong evidence of the involvement of motion mechanisms in pattern perception. In addition, they suggest the contribution of both the lower and higher neural mechanisms in the hierarchy of visual motion processing. In contrast to the general belief, our successful perception of dynamic visual world appears to be supported by tight and global interplay of pattern and motion processing.

Results

Retinal Painting by Tracking Eye Movements

Before directly examining the notion of motion-based pattern interpolation, I tested a few alternative accounts.

*Correspondence: nishida@brl.ntt.co.jp

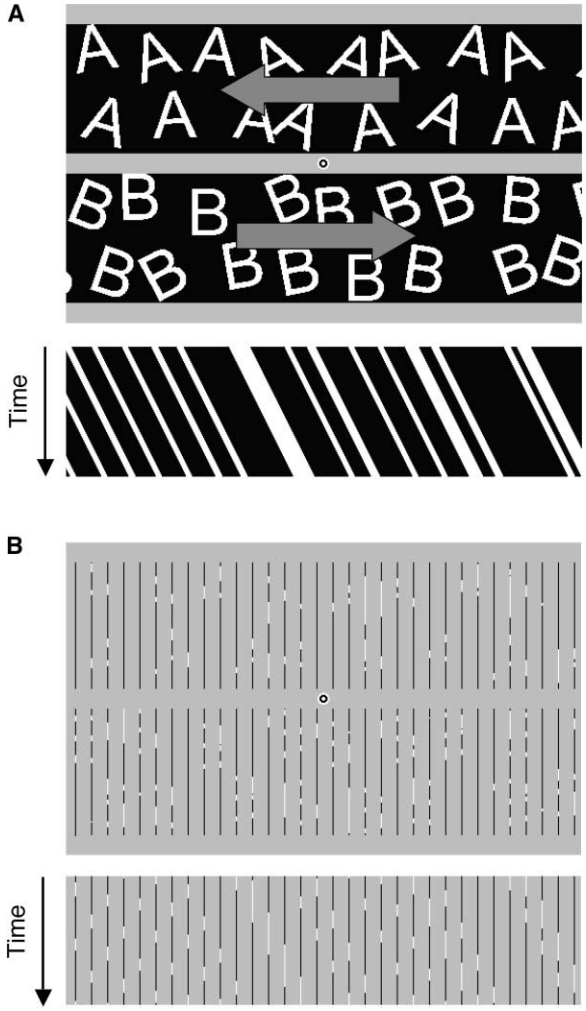

Figure 1. Stimulus Configuration

(A) The letter pattern presented behind the slits. Two sets of letters move in the opposite directions (as indicated by arrows). A panel with diagonal stripes, presented below, is a space-time (X-T) plot of a line of the lower-letter set.

(B) The pattern presented with stationary multiple slits. The condition of interslit separation was 32′. In this stationary figure, letter identification is almost impossible because of the presence of slits. However, letter movement significantly improves the visibility of the letters. A demonstration movie is available at http://www.brl.ntt.co.jp/people/nishida/demo.

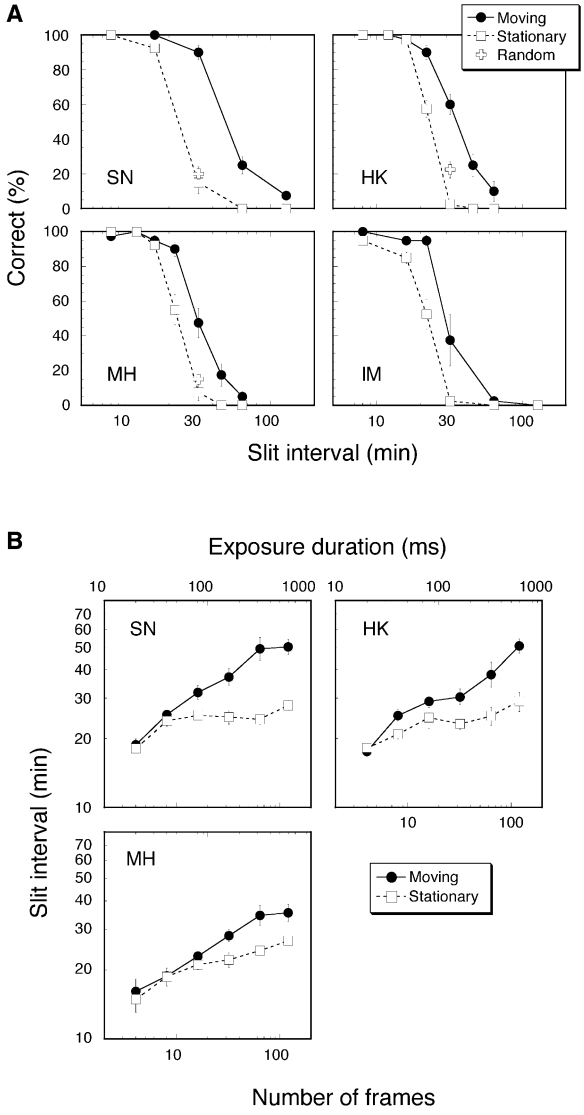

Figure 2. Letter Identification Was Better When the Letters Moved Behind the Slits Than When They Were Stationary

(A) The percent of trials in which the observer could correctly identify the two letters is plotted as a function of slit separation. The exposure duration was 320 ms (64 frame). The results of four observers are shown. The error bars indicate the standard errors across blocks. An open cross symbol indicates the data obtained when the order of motion sequence was randomized (for S.N., H.K., and M.H., only).

(B) The letter identification performance (in terms the threshold slit interval for which the proportion of correct trials was 50%) is plotted as a function of the number of frames (bottom abscissa)/exposure duration (top abscissa). The effects of letter movements were evident even when the exposure duration was less than 100 ms.

Retinal painting of a slit image by tracking eye movements has been proposed as a possible explanation of slit viewing (including both single- and multiple-slit cases) [33, 34, 41]. Tracking eye movements indeed improved pattern recognition in multislit viewing, but the question is whether it is always necessary. In the first experiment, two letters were presented for 320 ms through separate slit arrays. The moving directions of the two letters were opposite each other [21] and unpredictable to the observer (Figure 1). The observer's task was to report both of the letters while maintaining fixation at the display center. Figure 2A shows the percent of trials in which the observer could correctly report the two letters, as a function of slit separation. Although the contribution of tracking eye movements was expected to be excluded, letter-identification performance was significantly better when the letters were moving than when they were stationary.

I also examined the effect of exposure duration (by changing the number of movie frames). The letter identification performance was evaluated in terms of the threshold slit interval for which the proportion of correct trials was 50%. The larger threshold interval implies the

better letter identification. The result (Figure 2B) indicates that the letter movement improved the letter identification even when the exposure duration was less than 100 ms, which was obviously too brief for the observer to track the two letters moving in the opposite directions. These results strongly suggest that the spatiotemporal interpolation does not require the retinal painting by tracking movements, at least in the case of multislit viewing.

Number of Samples

Another factor that potentially improves the letter identification by pattern movements is an increase in the number of images differently sampled through the slits. It is possible that critical features of some characters may be visible only in particular sampling phases. To reduce the disadvantage of the stationary letter with regard to this aspect, all the experiments reported in this paper used a display in which each slit array contained 16 letters, all the same alphabet character at irregular positions and orientations, to show at least some of the letters in their "good" sampling phases. Nevertheless, the letter identification was greatly improved by letter movements, as described above. Additionally, when the same 64 frames (320 ms) of letter motion sequence were presented in a random order [34], the letter identification was not greatly improved as compared to the stationary condition (Figure 2A, open cross). This suggests that the increase in the number of samples is not sufficient as an explanation of multislit viewing. Further support for this conclusion was provided by the results of the following experiments.

Spatiotemporal Structure of Multislit Display

For further discussion of the neural mechanisms underlying multislit viewing, I should explain spatiotemporal structure of the multislit display, and a computational problem that the visual system has to solve. Figure 3A shows a three-dimensional (3D) space-time illustration of a moving letter "A" presented without slits and its Fourier spectrum. Since the letter is moving at a constant speed, all the frequency components lie on a single plane in the 3D Fourier space (Figure 3B) and on a single line in a two-dimensional (2D) temporal frequency versus horizontal spatial frequency space. The plane/line passes through the origin, and the slant is linearly related to the pattern speed. When the letter is shown through slits (Figure 3C), a sampling aliasing noise is added to the spectrum (Figure 3D). The interval between the noise and signal is inversely proportional to the interval of the slits. If the temporal change in the stimulus is not taken into account, the aliasing noise makes it impossible to recover the original pattern signal (Nyquist-Shannon sampling theorem [42]). However, since the signal and noise are separated in the 3D Fourier space (also in the 2D temporal-frequency versus horizontal spatial-frequency space), it is theoretically possible to recover the original pattern with an infinite accuracy by the analysis of the spatiotemporal structure of the stimulus [25, 43, 44].

Pattern interpolation in the real space-time domain can be regarded as a filtering operation in the Fourier domain. Two types of filtering are capable of extracting the original pattern signal and excluding the sampling noise. One is nondirection-selective filtering (Figure 3E). A low-pass filtering both in spatial and temporal frequency domains can selectively extract low spatial-frequency components of the signal [35]. The other one is direction-selective filtering of the signal components (Figure 3F). In the real space-time domain, this possibility implies the integration of pattern information over the trajectory of the pattern movement (motion-based pattern processing). These two possibilities can be distinguished on the basis of two features. One is the direction selectivity per se, and the other is the range of spatial frequency components that can be recovered. The next two experiments examined the first point, and the last experiment examined the second point.

Noise Masking

The direction specificity of multislit viewing was tested by means of a masking technique. Interslit areas were filled with a random noise mask that coherently moved in the same direction as or the opposite direction to the letter movement (Figure 4A). The slits and the mask were not physically overlapped. The percent of correct trials was measured as a function of the noise contrast. If the underlying mechanism of multislit viewing is not sensitive to motion direction, the masking effect on letter identification should remain the same regardless of the direction of noise motion relative to letter motion. The results (Figure 4B), however, indicate that the masking effect (i.e., impairment of letter identification) was much larger for the mask moving in the same direction than that moving in the opposite direction.

In the above experiment, the speed of the noise movement was matched with that of the letter movement. To investigate the speed tuning of the masking effect, I measured the threshold slit interval while systematically varying the speed of the noise masker. As shown in Figure 4C, for the noise moving in the opposite direction, the masking effect was nearly the same as the stationary masker regardless of the noise speed. The masking effect however was significantly increased for the noise moving in the same direction, with the strongest masking effect being observed at around the condition where the speeds of the letter and noise were matched. The results thus suggest that the masking effect is broadly tuned to the stimulus speed.

The direction and speed selectivity of the masking indicates that letter recognition in multislit viewing is mediated by a mechanism that is sensitive to pattern motion, which integrates pattern signals moving in the same direction (and at a similar speed) more than those moving otherwise.

Motion Adaptation

The next experiment employed a motion-adaptation technique. Under adaptation to drifting random lines, the percent of correct trials was measured for the stimuli moving in the same direction as or the opposite direction to the adapted direction. The results (Figure 5) indicated that the letter identification was significantly worse for the letters moving in the adapted direction than those

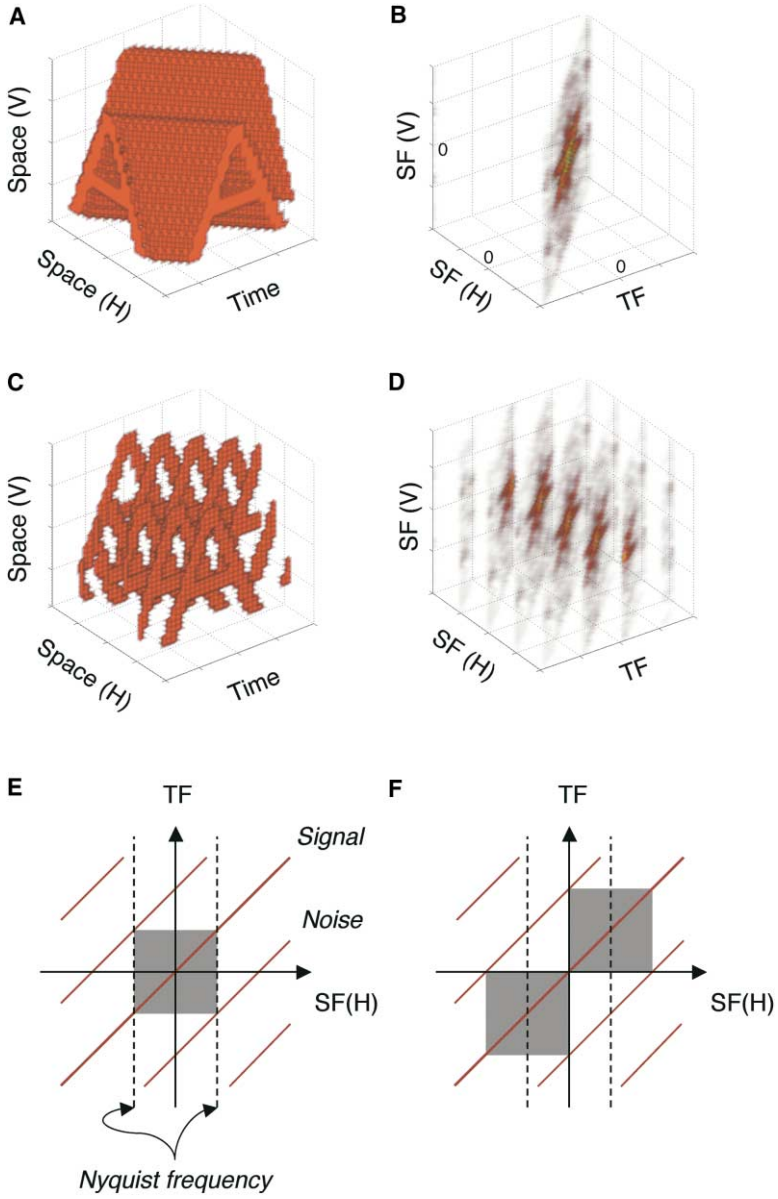

Figure 3. The Spatiotemporal Structure of Multislit Display

(A) A 3D space-time (X-Y-T) plot of an original unsampled moving letter "A."

(B) The Fourier spectrum of (A), which is lying on a single plane passing through the origin.

(C) A 3D space-time plot of a moving letter sampled by a vertical slit array.

(D) The Fourier spectrum of (C). Due to the slit sampling, aliasing noise components appear on the planes parallel to the signal plane.

(E) A 2D Fourier spectrum of the sampled image (top view of [D]). Dashed lines indicate the spatial Nyquit frequency (1/2*w*, where *w* is the slit interval). Gray area indicates the passband of the optimal nondirection-selective spatiotemporal low-pass filter that segregates signal from noise. The signal components higher than the Nyquist frequency cannot be recovered.

(F) The passband of a direction-selective filter that segregates signal from noise. The signal components higher than the Nyquist frequency can be recovered.

moving in the opposite direction. Motion adaptation impaired letter identification for the adapted direction, or improved letter identification for the opposite direction. These performance changes appear to correlate with the impairment of the motion perception in the adapted direction and/or the facilitation of the motion perception in the nonadapted direction, for the conditions with large slit intervals. The direction-selective adaptation effect also indicates that the motion information plays a critical role in pattern perception in multislit viewing.

Spatial-Frequency Map

The final experiment investigated the degree to which pattern information is recovered in multislit viewing in terms of the range of spatial frequency and orientation. As noted above, if spatiotemporal interpolation in multislit viewing is mediated by nondirectional low-pass filtering, there is a theoretical upper limit of the signal spatial frequency that can be recovered. Specifically, as is evident in Figure 3E, given that the slit-sampling interval is *w*, the upper horizontal spatial frequency should be 1/2*w*, which corresponds to the Nyquist spatial frequency of the slit sampling. In the case of direction-selective filtering, on the other hand, there is no such theoretical limit (Figure 3F). Therefore, the recovery of the horizontal spatial-frequency components higher than the Nyquist frequency can be regarded as evidence that the underlying mechanism of multislit viewing is direction selective, and more importantly, it indeed makes use of motion information for effective interpolation of moving spatial patterns.

A variant of the psychophysical reverse-correlation technique was used to estimate the spatial-frequency map of the recovered pattern [39, 40]. In brief, I divided image components into numerous, small spatial-frequency subbands, presented various letter images while

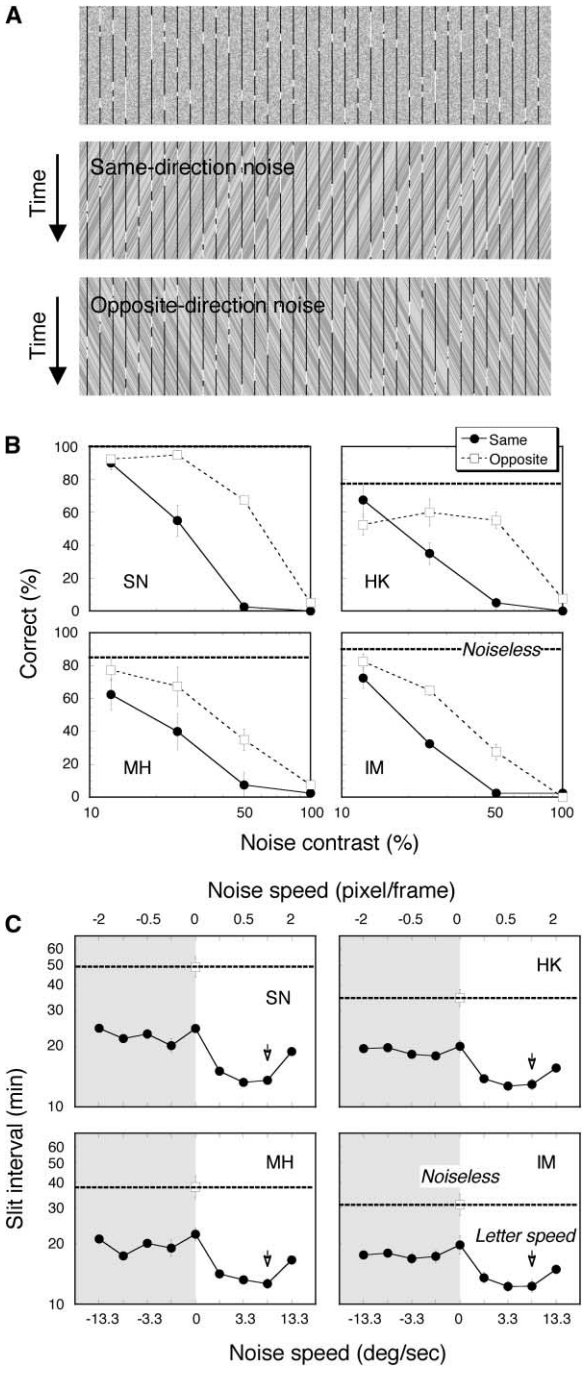

Figure 4. Direction-Selective Masking of Letter Identification by Random-Dot Noise

(A) A slit array with 50% contrast noise. A space-space (X-Y) plot, as well as space-time (X-T) plots for the noise moving in the same direction as or that opposite to the letter movement.

(B) The percent of correct trials as a function of noise contrast. The noise movement was in the same direction (filled circles) or in the opposite direction (open squares). A dashed line indicates the performance measured without noise. The speed of the noise movement was matched with the letter movement (6.7°/s). The exposure duration was 320 ms. The interslit interval was determined for each observer in a pilot test as the value that gave ~50% identification with the 25% contrast noise moving in the same direction (28′, 28′, 26′, and 22′ for S.N., H.K., M.H., and I.M., respectively). Letter identification was impaired more effectively when the noise moved in the same direction. Demonstration movies are available at http://www.brl.ntt.co.jp/people/nishida/demo.

(C) Speed tuning of the noise masking. The letter identification performance (in terms the threshold slit interval for which the proportion of correct trials was 50%) is plotted as a function of the noise speed. The speed of letter movement (6.7°/s) is indicated by an arrow. The results indicate that the noise masking is broadly tuned to speed.

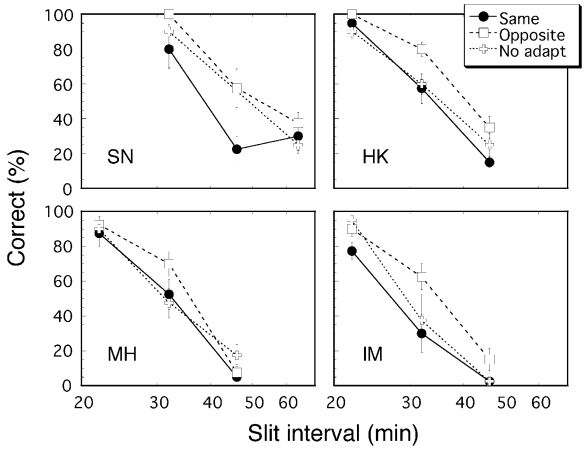

Figure 5. Direction-Selective Adaptation of Letter Identification

The percent of correct trials measured after adaptation to moving random lines is plotted as a function of slit interval. Letter identification was worse when the adapted direction was the same as the letter movement (filled circle) than when it was opposite (open square). The data measured without adaptation are shown for comparison (open cross, replotted from Figure 2A).

randomly including or excluding each subband component (Figures 6A–6C), and measured the letter identification performance. The stimulus was presented without slits in a control condition and with slits in the main condition. The letter image was filtered before, not after, slit sampling. The contribution of each subband to letter identification, or impact factor, was estimated by the correlation of the presentation states of the subband with observer's judgments. The impact factor was expected to be positive if the observer was able to correctly perceive that subband and use the obtained information for the letter identification.

In the impact factor maps (correlograms of letter identification, Figure 6D), the subbands showing positive factors (indicated by green) clustered at a relatively low spatial-frequency range. This is because of the use of fairly large fonts. Without slits (Figure 6D, left), the upper band of vertical spatial frequency was approximately two cycles/32′. This corresponded to 5.75 cycles/letter height, which was slightly higher than the reported best spatial-frequency band for letter recognition estimated by other methods (e.g., approximately three cycles/character [45]). In comparison with the no-slit condition, the green area further shrinks for the slit condition (interval = 32′, Figure 6D, right). This means that the observer could not perfectly recover the original pattern signal when it was presented through the widely separated slits. (cf., At the interval of 32′, the letter identification was nearly impossible for stationary letters, and the

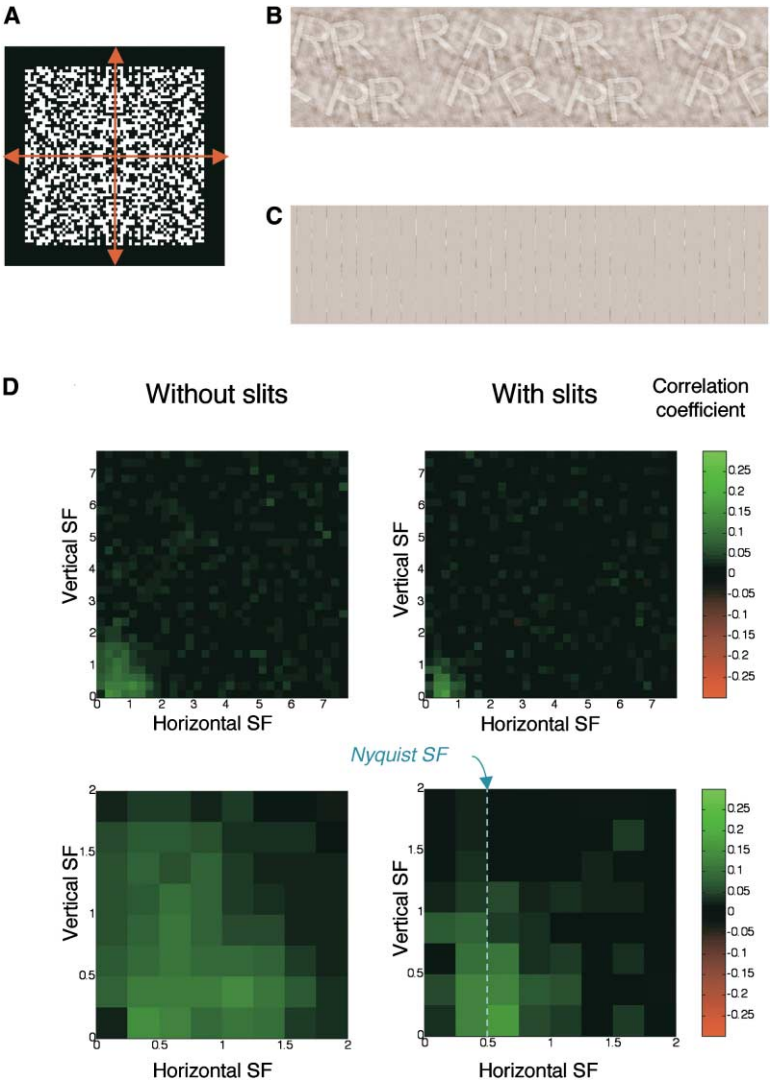

Figure 6. The Estimation of the Map of Perceived Spatial Frequencies

(A) An example of random-frequency mask (window open rate: 50%).

(B) An example of filtered image.

(C) An example of a filtered image shown through slits (interslit interval: 32′).

(D) The spatial frequency map (correlogram) of letter identification obtained without slits and with slits. Each map is based on 6000 trials (12,000 responses) by three observers. The percent correct was 60.7% for the no-slit condition and 51.1% for the slit condition. For each condition, the top panel is the frequency map for the full 32 × 32 subband windows (from zero to eight cycles, corresponding to the first quadrant of the mask shown in [A]). Spatial frequency is expressed in terms of slit sampling cycle (one cycle = 1/32′). The color indicates the correlation coefficient of the state of each subband and the observers' response. Green indicates that the pattern components in this subband had positive contribution to the letter identification (i.e., somehow visible). Since such subbands cluster in the lower spatial frequency range, the frequency map up to two cycles is magnified in the bottom panel. In the right panel (slit condition), a dashed vertical line indicates the Nyquist frequency of slit sampling (a half of the sampling frequency). The correlogram indicates that the frequency components higher than this Nyquist frequency were perceived by the observers. This theoretically requires direction-based pattern interpolation.

single-letter identification for unfiltered moving letters was ~90%.) Nevertheless, the green area extends beyond the Nyquist horizontal spatial frequency (0.5 cycles/slit interval), with the best subband (that showing the highest correlation coefficient) being present in the over-Nyquist range. As noted above, this result theoretically implies that the visual system does use the motion information for spatiotemporal pattern interpolation. In addition to the spatial frequency, one can read from the 2D correlograms how different orientation components contributed to the pattern identification. The ratio of the horizontal frequency relative to the vertical frequency determines orientation—as the vertical frequency increases, the image component becomes the more slanted from the vertical slits. The distribution of the green area indicates contribution of a wide range of orientations.

Discussion

By analyzing the spatiotemporal pattern interpolation for a multislit display, this study provided several lines of clear evidence of the involvement of motion mechanisms in pattern perception. First, simultaneous perception of two briefly presented and oppositely moving letters (Figure 2) indicates that the spatiotemporal interpolation does not require the retinal painting by tracking eye movements. Second, the effect of random ordering of motion sequence (and the effects of noise masking and motion adaptation as well) indicates that the increase in the number of samples alone is unable to account for the spatiotemporal interpolation effect. Third, the direction-selective impairment by noise making (Figure 4) indicates the direction selectivity of the interpolation mechanism. Fourth, the same conclusion is supported by the direction-selective impairment by motion adaptation (Figure 5). Finally, the motion-induced interpolation makes it possible to perceive the over-Nyquist spatial-frequency components, which are, theoretically, impossible to recover without the aid of motion information (Figure 6). These results all together indicate that the spatiotemporal interpolation is based on the brain's integration of spatial pattern information along the motion trajectory, as opposed to the currently pre-

vailing notion of separate processing of pattern and motion.

It has been shown that the performances of spatiotemporal pattern interpolation are positively correlated with the quality of visual motion [19, 25, 29]. For instance, reduction of the number of motion frames deteriorated both the smooth motion perception and the spatiotemporal interpolation [19]. Although such results can be interpreted as indicting some relevance of motion mechanisms to spatiotemporal interpolation, they do not directly demonstrate the direction-selective integration of pattern information.

Several forms of perceptual interaction between pattern and motion are known. They include the perception of shape defined by motion discontinuities [7, 8], the distortion of apparent position and shape by motion signals [10–12], and the effects of spatial patterns on motion integration [9]. These interactions could be interpreted as occurring after separate analysis of the two attributes in different visual areas. On the other hand, the spatiotemporal interpolation in multislit viewing (as well as the effects of pattern orientation on motion perception [13–17]) suggests an interaction at the earlier stages. Unlike the perception of motion-defined shapes [7, 8], the perceived motion is spatially coherent, so that motion per se gives no shape information. Instead, what is seen is a luminance-defined pattern, which is supposed to be encoded in early visual processing. It is known that human visual-motion processing consists of multiple stages. According to the standard model [46–51], local motion signal is first extracted by direction-selective spatiotemporal filtering. Then, local motion energy is computed by summing the squared outputs of a pair of quadrature phase filters. In a higher stage, the true 2D motion vector is recovered through selective integration of local motion energies across orientations, across spatial frequencies, and over space. It has been suggested that these functional stages may correspond to direction-selective V1 simple cells, complex cells, and MT pattern cells, respectively [49, 50, 52–57]. Which stages of motion processing are involved in motion-based pattern interpolation? A good candidate is the initial direction-selective filters, since they are sensitive not only to motion direction, but also to pattern orientation and spatial frequency, thus being able to integrate pattern information over motion trajectory [25–27]. Indeed, when a linear spatiotemporal filter tuned to pattern motion direction is applied to a space-time plot of a multislit display, a continuous spatiotemporal pattern, similar to the original pattern, can be obtained (Figures 7B and 7C). In the next stage, however, since the motion-energy extraction smoothes out the spatial phase-dependent modulation in motion signals over space, fine pattern information is no longer available (Figure 7D). This analysis suggests that the initial direction-selective filters (or V1 simple cells) may be used for the analysis of moving patterns, and if this is the case, their outputs should feed to further pattern analysis mechanisms through bifurcated pathways before undergoing further motion processing. According to this hypothesis, the direction-selective neurons have dual purposes, and the motion processing and the pattern processing are not segregated at its earliest stage.

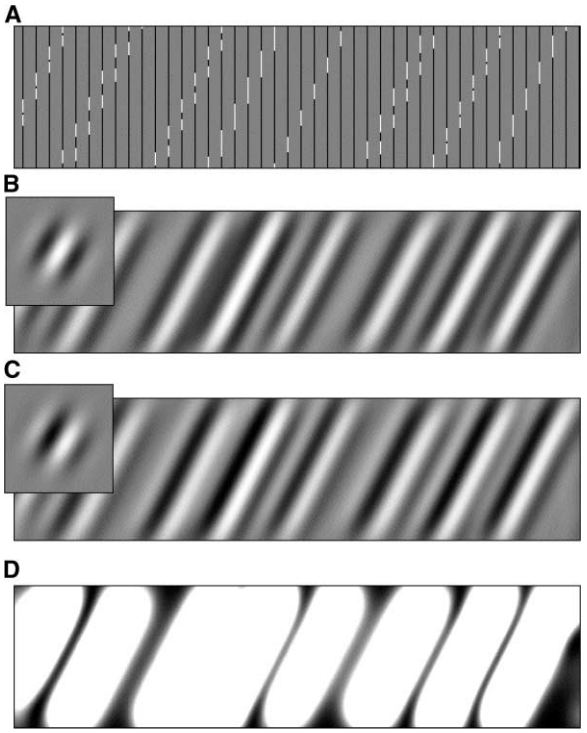

Figure 7. Possible Roles of the Two Stages of Early Motion Processing in Spatiotemporal Pattern Interpolation

One is direction-selective linear filtering and the other is motion energy computation, whose neural correlates may be direction-selective simple cells and complex cells in V1, respectively. The first stage response to a multislit stimulus (A) can be simulated by convolving the space-time plot of the input image by spatiotemporally slanted filters (shown in insets of [B] and [C]). The two filters differ in phase by 90° (quadrature pair). As shown by (B) and (C), the filtering is able to interpolate continuously moving edges. The second-stage response (D) is the squared sum of (B) and (C), in which the fine edge structure is blurred out. This suggests that the visual system may use V1 simple cells, but not complex cells, for spatiotemporal pattern interpolation.

The hypothesis is consistent with the observed effects of noise masking, since the direction-selective filtering mixes up signal with noise more when they move in the same direction than otherwise. It is also compatible with the effect of motion adaptation. The finding that the masking and adaptation effect was not very strongly selective to motion direction may suggest that the direction selectivity of the involved neurons is relatively weak or that both the direction-selective and nondirection-selective neurons are involved.

It should be noted, however, that the 2D spatial-frequency map obtained in the last experiment (Figure 6D) suggests that direction-selective filtering may not suffice as an explanation of motion-based pattern interpolation. Judging from the distribution of the green area (indicating positive impact factors), a wide range of orientation components did contribute to letter identification of the slit condition. This is not surprising since the target stimuli (letters) are 2D patterns. If various orientation components had not been visible, letter identification would have been difficult. The spatiotemporal interpolation of 2D patterns is theoretically important,

however, since the motion aperture problem makes it difficult to match locally measured velocities across different orientations [46, 58]. For correct integration of different orientation components, motion-based pattern interpolation should be based on the true 2D velocity, which is supposed to be recovered only in higher motion processing stages (e.g., MT pattern cells) [52, 56, 57]. The involvement of higher motion stages is also indicated by the recovery of over-Nyquist horizontal-frequency components. Because of aliasing by the slit sampling, it is not easy for the motion mechanisms to recover the true direction and speed of these components. Indeed, it was observed that the perceived motion direction always reversed when the moving pattern only contained the over-Nyquist (and below sampling) frequencies. To perceive these components to move in coherence with the other frequency components, the visual system may have to use a certain elaborate mechanism to attain phase congruency across different spatial frequencies, such as an across-frequency cooperative network. Possible neural correlates of such a mechanism have been reported only in the higher motion area, MT [59, 60]. In addition, although it is debatable whether multislit viewing and single-slit viewing are mediated by common neural mechanisms, it is worth mentioning that a computational model of single-slit viewing proposed that global motion must be computed prior to integration of pattern signals over time [61], and an fMRI study found activities in area MT+ during the presentation of single-slit displays [62].

Finally, when animating objects are presented by slit displays, the observers can perceive the pattern movement as well as the 3D structure defined by the object motion [63]. This phenomenon appears to indicate still more complicated pattern-motion interactions such that the spatial pattern recovered by the help of motion mechanisms is fed again to the standard motion analysis.

Conclusion

The present results suggest that the pattern information of moving image is spatiotemporally interpolated through early direction-selective filtering by V1 simple cells, but the integration of the local information represented by each simple cell into a global figure should be guided by the outputs of a higher-order motion mechanism, such as MT pattern cells. Spatiotemporal interpolation can be a powerful mechanism to improve the visibility of patterns, which never stand still in the retinal image due to movements of the visual scene and observer. The present results therefore suggest that the whole motion system has a significant contribution to dynamic pattern perception. Future studies should reconsider the functional roles of neural mechanisms that have been believed to contribute only to visual motion processing, especially V1 simple cells, MT pattern cells, and the feedback from MT to V1 [64].

Experimental Procedures

Observers

The author (S.N.) and three naive volunteers participated. They had normal or corrected-to-normal vision. Informed written consent was obtained before experiments.

Apparatus

The stimulus pattern was generated through a graphics card (VSG2/5, Cambridge Research Ltd) controlled by a host computer (Dell, OptiPlex GX200). It was presented on a CRT monitor (Clinton monoray monitor) with a fast greenish phosphor (DP104), a refresh rate of 200 Hz, and a spatial resolution of 512 $\times$ 512 pixels. Each pixel subtended 2′ at the viewing distance of 80.6 cm. The observer viewed the monitor while sitting in a dark room with his/her head fixed on a chin rest.

Standard Stimulus

The stimulus display (Figure 1) consisted of two arrays of vertical slits, presented on a uniform field (17.1° $\times$ 17.1°, 58 cd/m^2). The size of a slit array was 128 pixels (4.3°) in height and 512 pixels (17.1°) in width. The width of each slit was 1 pixel. The interslit interval was even across the stimulus. The gap between the upper and lower slit arrays was 20 pixels (40′) in height. A fixation bullseye was presented at the center of the gap.

Two sets of letters, moving in opposite directions, were shown through the upper and lower slit arrays, respectively. In each slit array, 16 uppercase letters of a given alphabet character were presented in Arial font (29–44 pixels in width, 46 pixels in height) as a bright image (116 cd/m^2) on a dark background ($<$0.1 cd/m^2). The letters were shown at irregular positions (without overlap) and orientations (varied within $\pm$30° from the vertical). This was to present differently sampled letters at a time. The alphabet character was chosen from 20 letters (IJQMWZ were excluded). Different characters were shown through the two slit arrays.

Standard Procedures

A trial started with a blank slit pattern. In response to observer's key press, letters were presented for 320 ms (64 refresh frames) at a drifting speed of 6.7°/s (1 pixel/frame), unless otherwise stated. Motion direction was randomly determined for each trial. During letter presentation, the observer was required to keep fixation at the center point. After a presentation of letters, the observer responded to the two presented alphabets one by one via keyboard. The chance level to get the correct response for both letters by random responses was only 0.25%.

To evaluate the performance of the two-letter identification task, we measured the percent of correct trials in some experiments (e.g., Figure 2A). In a block consisting of 10 trials, 20 different alphabets were shown in a random order. Stimulus conditions, including slit interval and letter movement, were varied between blocks. Each observer ran 4–8 blocks for each stimulus condition. The standard errors in the figures are based on a variation across blocks.

In other experiments (e.g., Figure 2B), the slit interval that gave 50% correct performance was estimated by the double-random staircase procedure in which two staircases starting with small and large slit intervals were randomly interleaved. Within each staircase, the slit interval was increased after a correct trial or decreased after an incorrect trial. The step size was $\times$2 until the first reversal, $\times 2^{0.5}$ until the second reversal, and $\times 2^{0.25}$ until the staircase terminated at the sixth reversal. The arithmetic mean of the last four reversals was taken as an estimate of the threshold slit interval. Each observer ran 4–8 staircases for each stimulus condition. The standard errors in the figures are based on a variation across staircases.

Noise Masking

A binary noise field in which dark and bright pixels randomly appeared with equal probabilities filled interslit areas. The luminance contrast of the noise was varied, while the mean luminance was kept constant (58 cd/m^2). When moved, the noise field was horizontally shifted without considering occlusion by the slits. In the first experiment (Figure 4B), the noise speed was 6.7°/s (1 pixel/frame). In the second experiment (Figure 4C), it varied from 1.7°/s (1 pixel/4 frames) to 13.3°/s (2 pixels/frame). The slit interval used for the first experiment was determined in a preliminary experiment for each observer as the interval that gave $\sim$50% correct performance by a 25% contrast noise moving in the letter direction.

Motion Adaptation

The adaptation stimulus was a pair of random line patterns in which each vertical line was randomly painted in black or white with a

probability of 50%. It subtended the same areas as the slit patterns, moving in the opposite directions between the upper and lower areas. The luminance contrast was 100%, and the moving speed was 6.7°/s. The observer adapted to this motion pattern for 60 s at the start of experiment and readapted for 10 s during every intertrial interval.

Spatial Frequency Map Estimation

In each trial, a new letter image was filtered by a new random-window mask pattern. To avoid luminance overflow by filtering, the contrast of source letter image was reduced to 70%, and the DC level of the filtered image was adjusted so that all the luminance values were in the displayable range. The letter image was transformed to the spatial-frequency domain by FFT, filtered by a random-frequency mask, and then transformed back to the real-space domain by inverse FFT. The mask consisted of a grid that divided the spatial frequency space (one quadrant of the mask shown in Figure 6A) into 32 $\times$ 32 subband windows. The state of each mask window, open or close, was independently and randomly determined with a given window open rate. The random window pattern was symmetric about the horizontal and vertical spatial-frequency axes, which implies that no distinction was made between orientation components tilted clockwise and counterclockwise from the vertical. A pair of filtered moving letter patterns was shown without slits (no slit condition) or through slits (slit condition). In the latter case, the slit interval was 16 pixels (32′). The window-open rate was adjusted to make the single letter-identification performance $\sim$50%. It was 20%–30% (on average 27.5%) for the no slit condition, and 50%–70% (67.3%) for the slit condition. Each of three observers ran 2000 trials for each condition. The responses for the upper and lower letters were independently counted in computation of the correlogram for the purpose of increasing the number of observations.

Acknowledgments

The author thanks T. Tokimoto and M. Ohishi (AVIX, Inc.) for helpful discussions.

Received: March 1, 2004
Revised: March 17, 2004
Accepted: March 17, 2004
Published: May 25, 2004

References

1. Livingstone, M.S., and Hubel, D.H. (1987). Segregation of form, color, movement and depth: Anatomy, physiology, and perception. Science *240*, 740–749.
2. Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In Analysis of Visual Behavior, D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield, eds. (Cambridge, MA: MIT Press), pp. 549–586.
3. Zeki, S. (1993). A Vision of the Brain (Oxford: Blackwel).
4. Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. Zeitschrift für Psychologie *61*, 161–265.
5. Kulikowski, J.J., and Tolhurst, D.J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. J. Physiol. *232*, 149–162.
6. Lu, Z.L., Lesmes, L.A., and Sperling, G. (1999). Perceptual motion standstill in rapidly moving chromatic displays. Proc. Natl. Acad. Sci. USA *96*, 15374–15379.
7. Wallach, H., and O'Connell, D.N. (1953). The kinetic depth effect. J. Exp. Psychol. *45*, 205–217.
8. Braddick, O. (1974). A short-range process in apparent motion. Vision Res. *14*, 519–527.
9. Stoner, G.R., Albright, T.D., and Ramachandran, V.S. (1990). Transparency and coherence in human motion perception. Nature *344*, 153–155.
10. De Valois, R.L., and De Valois, K.K. (1991). Vernier acuity with stationary moving Gabors. Vision Res. *31*, 1619–1626.
11. Nishida, S., and Johnston, A. (1999). Influence of motion signals on the perceived position of spatial pattern. Nature *397*, 610–612.
12. Whitney, D., and Cavanagh, P. (2000). Motion distorts visual space: shifting the perceived position of remote stationary objects. Nat. Neurosci. *3*, 954–959.
13. Geisler, W.S. (1999). Motion streaks provide a spatial code for motion direction. Nature *400*, 65–69.
14. Ross, J., Badcock, D.R., and Hayes, A. (2000). Coherent global motion in the absence of coherent velocity signals. Curr. Biol. *10*, 679–682.
15. Geisler, W.S., Albrecht, D.G., Crane, A.M., and Stern, L. (2001). Motion direction signals in the primary visual cortex of cat and monkey. Vis. Neurosci. *18*, 501–516.
16. Burr, D.C., and Ross, J. (2002). Direct evidence that "speedlines" influence motion mechanisms. J. Neurosci. *22*, 8661–8664.
17. Krekelberg, B., Dannenberg, S., Hoffmann, K.P., Bremmer, F., and Ross, J. (2003). Neural correlates of implied motion. Nature *424*, 674–677.
18. Ross, J. (1977). A new type of display relying on vision's sensitivity to motion. J. Physiol. *271*, 2P–3P.
19. Burr, D.C. (1979). Acuity for apparent vernier offset. Vision Res. *19*, 835–837.
20. Zöllner, F. (1882). Über eine neue Art anorthoskopischer Zerrbilder. Annalen der Physic *117*, 477–484.
21. Parks, T.E. (1965). Post-retinal visual storage. Am. J. Psychol. *78*, 145–147.
22. Bruno, N., and Bertamini, M. (1990). Identifying contours from occlusion events. Percept. Psychophys. *48*, 331–342.
23. Shipley, T.F., and Kellman, P.J. (1993). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? Spat. Vis. *7*, 323–339.
24. Mateeff, S., Popov, D., and Hohnsbein, J. (1993). Multi-aperture viewing: perception of figures through very small apertures. Vision Res. *33*, 2563–2567.
25. Fahle, M., and Poggio, T. (1981). Visual hyperacuity: spatiotemporal interpolation in human vision. Proc. R. Soc. Lond. B. Biol. Sci. *213*, 451–477.
26. Burr, D.C., and Ross, J. (1986). Visual processing of motion. Trends Neurosci. *9*, 304–307.
27. Burr, D.C., Ross, J., and Morrone, M.C. (1986). Seeing objects in motion. Proc. R. Soc. Lond. B. Biol. Sci. *227*, 249–265.
28. Burr, D.C., Ross, J., and Morrone, M.C. (1986). Smooth and sampled motion. Vision Res. *26*, 643–652.
29. Fahle, M., Biester, A., and Morrone, C. (2001). Spatiotemporal interpolation and quality of apparent motion. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *18*, 2668–2678.
30. Burr, D.C. (1980). Motion smear. Nature *284*, 164–165.
31. Hogben, J.H., and di Lollo, V. (1985). Suppression of visible persistence in apparent motion. Percept. Psychophys. *38*, 450–460.
32. Nishida, S., Motoyoshi, I., and Takeuchi, T. (1999). Is the size aftereffect direction selective? Vision Res. *39*, 3592–3601.
33. Anstis, S.M., and Atkinson, J. (1967). Distortions in moving figures viewed through a stationary slit. Am. J. Psychol. *80*, 572–585.
34. Morgan, M.J., Findlay, J.M., and Watt, R.J. (1982). Aperture viewing: a review and a synthesis. Q. J. Exp. Psychol. *34A*, 211–233.
35. Morgan, M.J., and Watt, R.J. (1983). On the failure of spatiotemporal interpolation: a filtering model. Vision Res. *23*, 997–1004.
36. Barlow, H.B., and Hill, R.M. (1963). Evidence for a physiological explanation of the Waterfall phenomenon and figural after-effects. Nature *200*, 1345–1347.
37. Sekuler, R.W., and Ganz, L. (1963). Aftereffect of seen motion with a stabilized retinal image. Science *139*, 419–420.
38. De Valois, R.L., and De Valois, K.K. (1988). Spatial Vision (New York: Oxford University Press).
39. Ahumada, A.J., Jr., and Lovell, J. (1971). Stimulus features in signal detection. J. Acoust. Soc. Am. *49*, 1751–1756.
40. Chubb, C., and Nam, J.H. (2000). Variance of high contrast textures is sensed using negative half-wave rectification. Vision Res. *40*, 1677–1694.

41. Helmholtz, H.v. (2000). Helmholtz's Treatise on Physiological Optics, Volume 3 (Bristol: Thoemmes Press).
42. Shannon, C.E. (1949). Communication in the presence of noise. Proc. Institute of Radio Engineers *37*, 10–21.
43. Barlow, H.B. (1979). Reconstructing the visual image in space and time. Nature *279*, 189–190.
44. Click, F.H.C., Marr, D.C., and Poggio, T. (1981). An information-processing approach to understanding the visual cortex. In The Organization of the Cerebral Cortex, F.O. Schmitt, ed. (Cambridge, MA: MIT press), pp. 505–503.
45. Solomon, J.A., and Pelli, D.G. (1994). The visual filter mediating letter identification. Nature *369*, 395–397.
46. Adelson, E.H., and Movshon, J.A. (1982). Phonomenal coherence of moving visual patterns. Nature *300*, 523–525.
47. Adelson, E.H., and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A *2*, 284–299.
48. Watson, A.B., and Ahumada, A.J., Jr. (1985). Model of human visual-motion sensing. J. Opt. Soc. Am. A *2*, 322–341.
49. Smith, A.T., and Snowden, R.J. (1994). Visual Detection of Motion (London: Academic Press).
50. Simoncelli, E.P., and Heeger, D.J. (1998). A model of neuronal responses in visual area MT. Vision Res. *38*, 743–761.
51. Nishida, S., and Ashida, H. (2000). A hierarchical structure of motion system revealed by interocular transfer of flicker motion aftereffects. Vision Res. *40*, 265–278.
52. Movshon, J.A., Adelson, E.H., Gizzi, M.S., and Newsome, W.T. (1985). The analysis of moving visual patterns. Exp. Brain Res. *11*, 117–151.
53. McLean, J., and Palmer, L.A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of the cat. Vision Res. *29*, 675–679.
54. Emerson, R.C., Bergen, J.R., and Adelson, E.H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. Vision Res. *32*, 203–218.
55. DeAngelis, G.C., Ohzawa, I., and Freeman, R.D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. J. Neurophysiol. *69*, 1091–1117.
56. Pack, C.C., and Born, R.T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature *409*, 1040–1042.
57. Huk, A.C., and Heeger, D.J. (2002). Pattern-motion responses in human visual cortex. Nat. Neurosci. *5*, 72–75.
58. Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information (New York: Freeman).
59. Perrone, J.A., and Thiele, A. (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. Nat. Neurosci. *4*, 526–532.
60. Priebe, N.J., Cassanello, C.R., and Lisberger, S.G. (2003). The neural representation of speed in macaque area MT/V5. J. Neurosci. *23*, 5650–5661.
61. Shimojo, S., and Richards, W. (1986). Seeing" shapes that are almost totally occluded: A new look at Parks's camel. Percept. Psychophys. *39*, 418–426.
62. Yin, C., Shimojo, S., Moore, C., and Engel, S.A. (2002). Dynamic shape integration in extrastriate cortex. Curr. Biol. *12*, 1379–1385.
63. Fujita, N. (1990). Three-dimensional anorthoscopic perception. Perception *19*, 767–771.
64. Maunsell, J.H., and van Essen, D.C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. J. Neurosci. *3*, 2563–2586.