Rapid encoding of relationships between spatially remote motion signals

Kazushi Maruya

NTT Communication Science Labs, Nippon Telegraph and Telephone Corporation, Japan

Alex O. Holcombe

School of Psychology, University of Sydney, Australia

Shin’ya Nishida

NTT Communication Science Labs, Nippon Telegraph and Telephone Corporation, Japan

For visual processing, the temporal correlation of remote local motion signals is a strong cue to detect meaningful large-scale structures in the retinal image, because related points are likely to move together regardless of their spatial separation. While the processing of multi-element motion patterns involved in biological motion and optic flow has been studied intensively, the encoding of simpler pairwise relationships between remote motion signals remains poorly understood. We investigated this process by measuring the temporal rate limit for perceiving the relationship of two motion directions presented at the same time at different spatial locations. Compared to luminance or orientation, motion comparison was more rapid. Performance remained very high even when interstimulus separation was increased up to 100°. Motion comparison also remained rapid regardless of whether the two motion directions were similar to or different from each other. The exception was a dramatic slowing when the elements formed an orthogonal “T,” in which two motions do not perceptually group together. Motion presented at task-irrelevant positions did not reduce performance, suggesting that the rapid motion comparison could not be ascribed to global optic flow processing. Our findings reveal the existence and unique nature of specialized processing that encodes long-range relationships between motion signals for quick appreciation of global dynamic scene structure.

Introduction

In early visual processing, local image features, such as edge orientation, patch color, and pattern motion, are encoded in parallel across the retina. The responsible neural hardware is a bank of neural sensors with small receptive fields (Hubel & Wiesel, 1968). Meaningful scene structures in the retinal image are often larger than the receptive field of these sensors and have more complex spatiotemporal structures than can be represented by a single local sensor. To analyze such scene structures, the subsequent visual processing must encode spatiotemporal relationships among local measurements, presumably in a hierarchical way (Connor, Brincat, & Pasupathy, 2007).

For motion processing in particular, early neural stages including primary visual area and medial temporal (MT) area compute local motion vectors from retinal images (see Burr & Thompson, 2011; Krekelberg, 2008; Nishida, 2011 for reviews). Subsequent processing analyzes the spatiotemporal patterns formed by those local motion signals. Certain multi-element motion pattern analyses have been studied extensively, including optic flow (Duffy, 2003; Gibson, 1950; Warren, 2008), structure from motion (Jain & Zaidi, 2011; structure from motion (Jain & Zaidi, 2011; Ullman, 1979), biological motion (Blake & Shiffrar, 2007; Johansson, 1973; Troje, 2008), and material perception (Doerschner et al., 2011). The basic computational components of motion pattern analysis remain poorly understood, however.

The temporal correlation of a pair of remote local motion signals, such as a common movement and a synchronous direction change, is a powerful cue for detecting global scene structure, as suggested by the Gestalt law of grouping by common fate (Koffka, 1935). The cue arises because points connected to each other in the world and belonging to the same global structure are likely to move together regardless of how far they are separated. Lappin and colleagues showed that the minimum motion speed necessary to detect relative motion within a pair of remote elements was


doi: 10.1167/13.2.4 Received September 3, 2012; published February 6, 2013 ISSN 1534-7362 © 2013 ARVO
very small even when the interelement separation was several degrees (Lappin, Donnelly, & Kojima, 2001; Mowafy, Blake, & Lappin, 1990). Their finding suggests that the visual motion system encodes the relationship between remote local motion signals very effectively. The process underlying this computation must be fundamental to motion pattern analysis in general, but relatively little is known about it.

Recent studies have established a useful psychophysical technique to test how effectively the brain encodes the relationship between two signals (Arnold, 2005; Bartels & Zeki, 2006; Fujisaki & Nishida, 2010; Holcombe, 2009; Holcombe & Cavanagh, 2001; Holcombe & Judson, 2007; Moutoussis & Zeki, 1997). The task is the temporal phase discrimination of two stimulus sequences, A and B, simultaneously presented. Each sequence is made up of the alternation of two feature values (e.g., leftward and rightward motions). Both sequences alternate in synchrony, but whether the first phase or second phase of A is at the same time as the first phase or second phase of B is randomly determined. The observers must discriminate between these two possibilities. Because temporal coincidence is a sign of two features belonging to the same event, this task is also known as the temporal binding task. When the rate of alternation (temporal frequency) is fast enough, observers can no longer make the discrimination. The rate corresponding to a particular level of performance (in this paper, 75%) will be referred to as the critical temporal frequency. Just as the flicker-fusion frequency provides a measure of the temporal resolution of early visual processing, a high critical frequency of phase discrimination implies rapid and efficient encoding of the pairwise relationship (although this does not necessarily mean short latency to reach the final perceptual decision).

The present study applied this technique to investigation of the process underlying pairwise comparison of remote local motion signals. Unlike a method used previously (Lappin et al., 2001), this method can evaluate the characteristics of the motion comparison process independent of those of the earlier motion detection processes. Rather than being limited to barely visible slow speeds, it is applicable to ordinary suprathreshold speeds. Indeed, it is also applicable to a variety of stimulus conditions, allowing comparison of the motion comparison process to that for other attributes.

Previous work suggests that whereas temporal phase discrimination judgments mediated by higher-level cognitive processing have slow temporal limits, relationships encoded by low-level sensory mechanisms have fast temporal limits (Amano, Johnston, & Nishida, 2007; Arnold, 2005; Bartels & Zeki, 2006; Fujisaki & Nishida, 2010; Holcombe, 2009; Holcombe & Cavanagh, 2001; Holcombe & Judson, 2007; Motoyoshi & Nishida, 2002). Specifically, the critical temporal frequency is fairly low (2–3 Hz) when the two sequences differ from each other in position and attributes (e.g., color vs. orientation), or in sensory modality (e.g., vision and audition). When, as when originating in distinct modalities, the features of the two sequences are very discrepant, it is believed that the temporal judgment is limited by the time it takes for a high-level process to identify and label the two simultaneously presented stimuli (Fujisaki & Nishida, 2010; Holcombe & Cavanagh, 2001). When the two stimulus sequences involve the same feature values instead of different features, the critical frequency can be much higher. This is true for example when both sequences alternate horizontally oriented with vertically oriented stimuli, or when both involve alternation of the same two colors or luminances (Fujisaki & Nishida, 2010; Victor & Conte, 2002). Previous investigations of the fast category of tasks have shown that increasing the spatial separation between the two sequences results in a decline in the critical temporal frequency (Aghdaee & Cavanagh, 2007; Motoyoshi & Nishida, 2002; Victor & Conte, 2002). This distance dependence suggests the involvement of fast specialized sensory mechanisms, such as edge, luminance motion, or angle detectors that mainly encode short-range relationships between neighboring local signals of the same attribute. It is suggested that at large spatial separations, this mechanism is not effective, and the slower, higher-level process mediates the judgments instead (Aghdaee & Cavanagh, 2007; Battelli et al., 2001).

To understand processing of the relationship between two motion signals, we measured the critical temporal frequency to elucidate processing efficiency, spatial range, and direction tuning. We discovered characteristics of motion that differ from all previously measured visual attributes. First, the critical frequency to encode pairwise remote relationships was considerably higher for motion than for luminance and orientation. Second, the critical frequency for motion fell off only slightly as the inter-sequence separation was increased, remaining relatively high (∼ 10 Hz) even when the sequences were separated by 100°. Third, the critical frequency was high for a wide range of configurations, including different directions for the two local motions, but it dropped when they formed a T shape. Lack of masking by motion presented at task-irrelevant positions suggested that the rapid motion comparison could not be ascribed to global optic flow processing. These results indicate the existence of specialized motion processing for pairwise comparison of remote motion signals and show its unique properties suited to rapid analysis of the global motion structures in the visual scene.
General methods

Unless otherwise described, the methods described here were used in all series of the experiments.

Apparatus

The stimuli were created with custom software using the OpenGL library and running on a DOS/V computer with a gamma correction. They were displayed on a 21" CRT monitor (GDM-F520, SONY, Tokyo). The display had 1024 × 768 pixels and a refresh rate of 120 Hz. Each pixel subtended 4 arcmin at a viewing distance of 31 cm. A chinrest stabilized the observer’s head.

Stimulus

Each element sequence consisted of a patch of texture pattern (Figure 1a), whose spatial profile was defined by a stationary Gaussian envelope ($SD = 1.07\degree$). A fixation point was supplied by a small red cross at the center of display. Two patches were presented on both sides of the fixation point (Figure 1b). The texture within a patch was made by attenuating the amplitude of the spatial frequency spectrum of a gray-scale white noise in proportion to the inverse of the spatial frequency ("1/f" noise). The 1/f texture was chosen to reduce or avoid a sensitivity change with eccentricity. Although the spatial scale of visual processors changes with eccentricity, since frequency bandwidths of visual cortex neurons are nearly constant in octaves, the response of those neurons to 1/f patterns is similar across spatial scales (Bex, Brady, Fredericksen, & Hess, 1995; Field, 1987) much more than the responses to other stimuli including white noise and narrow-band patterns. The root-mean-square (RMS) contrast was approximately 70%. The background was a uniform gray field (57.5 cd/m², 68.3° × 51.2°).

Experimental procedure and data analysis

The critical frequency of temporal phase discrimination was measured by a temporal two interval forced choice (2IFC) method and a staircase procedure (Figure 1c). In each trial, observers viewed two stimulus movies in succession. Each lasted for 1.5 s, with a 250 ms contrast ramp at onset and offset—the contrast increased linearly with time from 0 to the maximum level (~70%) during the initial 250 ms and decreased linearly to 0% during the final 250 ms. For all conditions, in one movie the texture patterns of the two sequences oscillated in phase. In the other movie, they oscillated in the reversed phase. The order of the movies (first interval vs. second interval) was randomly determined. Observers were asked to judge which movie contained an in-phase pair (e.g., in which the two motions always moved in the same direction). After the observer responded with a key press, a tone indicated whether the response was correct. The range of temporal frequency of oscillation was 14 steps between 0.68 and 60 Hz. A session began with an oscillation of 2.7 Hz, and the oscillation frequency stepped up after every correct response until the observers made the initial error. Then the frequency stepped one down, and a modified 3-up, 1-down staircase began (Dixon & Mood, 1948). The frequency stepped up when observers made three consecutive correct responses and stepped down when observers made an error. The staircase was terminated after four staircase reversals. Data from the first ascending series of trials were not used for data analysis. This procedure resulted in data at many temporal frequencies, but concentrated around the discrimination threshold. At the end of data collection, a psychometric function was computed from the responses of all sessions, from which the 75% correct threshold was estimated by the maximum likelihood fitting of a logistic function.

Experiment 1

Experiment 1a: Attribute type and inter-sequence separation

In the first experiment, we measured the highest temporal frequency at which observers could reliably perform the remote motion comparison and investigated how this critical frequency is affected by the separation between the sequences. To compare performance with attributes other than motion, we also measured the critical frequency for luminance, luminance contrast, and orientation alternations under similar conditions.

Methods

Six observers including two authors participated in the experiment. All observers had normal or corrected-to-normal visual acuity. Informed consent was obtained before the experiment started, and the experiment was approved by the NTT Communication Science Laboratories Research Ethics Committee. Each observer ran at least six staircases for each condition. The two patches were presented on both sides of the fixation point, at the eccentricity of 5° or 20°, with a center-to-center sequence separation of 10° or 40°. Viewing distance was fixed at 31 cm with a chinrest.
For the task conditions involving motion, the texture in a stationary patch was an isotropic 1/f noise and made a horizontal oscillatory movement at the designated temporal frequency of that trial. The motion speed was always $32^\circ/s$ (16 min/frame). This implies that the position of the texture in a patch displaced with a triangular temporal waveform whose amplitude was reduced in inverse proportion to the oscillation frequency.

In the luminance condition, we decreased and increased the mean luminance of an isotropic stationary 1/f noise carrier to make a dark patch and a bright patch, respectively. The Michelson contrast of the temporal mean-luminance modulation produced by alternation of the dark and light patches was 75%.

In the luminance contrast condition, an isotropic stationary 1/f noise pattern with a Gaussian envelope was alternated with the gray background at the desired frequency. This sequence produced an oscillation of the luminance contrast between $\sim70\%$ and 0%.

In the orientation condition, each patch consisted of a nonisotropic 1/f texture whose orientation content was restricted to $\pm4^\circ$ of one of two orientations ($0^\circ$ or $90^\circ$). At a given rate, the texture orientation alternated
between two orthogonal orientations (i.e., from 0° to 90° and back to 0°).

**Results**

Figure 2 shows the results. For motion, the critical frequency of temporal phase discrimination was fairly high at the separation of 10° (~15 Hz). At 40°, it was slightly decreased, t(5) = 4.55, p < 0.01, but still high (~12 Hz). The critical frequency was lower for luminance contrast, luminance, and lowest for orientation (ANOVA indicated a significant main effect of attribute, F(3, 15) = 124.92, p < 0.01). Increasing the separation between the sequences had a more deleterious effect on the luminance and orientation cases than it did on motion (separation-attribute interaction, F(3, 15) = 13.82, p < 0.01).

**Experiment 1b: Intersequence separation of 100°**

Experiment 1a found that the critical frequency for motion was high even with the widest separation of 40°, so we tested a larger separation—100°. For this experiment, rather than varying the physical separation of the stimulus elements, we kept the physical stimulus constant as that used for the 40° separation condition of the first experiment and varied the viewing distance. Varying the viewing distance from 12 to 31 to 120 cm yielded center-to-center intersequence separations of 10°, 40°, and 100°, respectively. Because the retinal element size varied in proportion to the retinal eccentricity, this procedure was expected to reduce or eliminate the potential effect of eccentricity on spatial visual processing (Murakami & Shimojo, 1996).

**Methods**

Four observers including two of the three authors participated. All observers also participated in Experiment 1a. Observers viewed the same pairs of motion or luminance contrast sequences used in the 40° separation condition of the last experiment, but the viewing distance varied from 12 to 31 to 120 cm, yielding center-to-center intersequence separations of 10°, 40°, and 100°, respectively. In each of at least six sessions, observers ran one staircase. Intersequence separation was fixed within each session.

**Results**

The results (Figure 3) show that the critical frequency for motion was again fairly high (~12 Hz) and nearly constant despite the change in separation from 10° to 100°. The critical frequency for luminance contrast was lower than that for motion and also declined more with separation (separation-attribute interaction, F(2, 6) = 6.81, p < 0.05).

**Experiment 1c: An additional experiment with a large number of observers**

Experiment 1b used only four observers that included two of the authors. To test the generality of the main finding, we ran an additional experiment with a larger number of observers. Ten observers discrimi-
inated the relative phase of two motion elements that were horizontally oscillating at 10 Hz and separated by 108, and we calculated the proportion correct of phase discrimination.

**Methods**

The stimulus was the same as for Experiment 1b except for the following points. The oscillation frequency was fixed at 10 Hz. The ten observers did not include any authors. None of the observers knew the purpose of the experiment, except for the two who also participated in Experiment 1b. The experiment was run in sessions. Each session consisted of 30 trials. In each trial, a pair of sequences oscillated horizontally with one of two opposite phase-relationships. Observers judged whether the oscillation was in-phase or reversed-phase. Each observer ran at least two sessions. We pooled the data across observers and calculated the proportion correct for each stimulus type.

**Results**

The average proportion correct was 82.8%. The 95% confidence interval, calculated from the standard error, was ± 7.4%. These results indicate that the critical frequency is higher than 10 Hz, confirming the result found for fewer observers in Experiment 1b.

---

**Experiment 1d: Local change detection?**

Conceivably the better temporal thresholds for the motion condition relative to the luminance contrast conditions might reflect a difference only in local signal detection sensitivity, rather than of local signal comparison. That is, the advantage for motion might be in detecting the individual signals (directions) compared to detecting the luminance contrast signals. We therefore compared the temporal limits for detecting local changes, using the stimuli used to test phase discrimination at 10° and 100° separation (5° and 50° eccentricity, respectively).

**Methods**

Three observers including an author participated, and all observers also participated in Experiment 1b. The temporal limit of local change detection was measured for the same stimulus as used in the first experiment. The separation was either 10° or 100°, accomplished by changing the viewing distance from 120 to 12 cm. The procedure was a 2IFC response combined with a staircase procedure as in the first experiment. In each trial, observers viewed two stimulus movies in succession. For motion sessions, the textures of the two sequences oscillated horizontally in one movie, and vertically in the other. For luminance contrast sessions, the textures alternated with a gray (the luminance corresponded to the mean luminance of the texture) uniform background in one movie and were unchanging in the other movie. The RMS contrast of the stationary pattern was half that of the alternated pattern (~35%) so that the two movies contained the same net luminance energy. In the motion sessions, observers indicated which interval contained the vertical motion. In the luminance contrast sessions, observers indicated which interval contained flicker. Observers performed at least three sessions for each of motion and luminance contrast conditions.

**Results**

We pooled the data for each temporal frequency condition and calculated the proportion correct. For all observers and conditions, observers performed the tasks perfectly with oscillations below 20 Hz. At 30 Hz (Figure 4), the performance with 100° separation was lower than that for 10° for both the luminance-contrast and the motion oscillations (probably because of the greater eccentricity of the stimuli at the larger separation). The performance for luminance contrast was similar to or slightly better than that for motion.

These results suggest that the critical frequencies shown in Figure 2 (and also in Figure 1) did not reflect the limits on local signal detection, but those of...
subsequent signal comparison. Certainly the advantage for motion in the phase discrimination cannot be attributed to an advantage in local signal detection, as performance for luminance contrast signal detection was better.

Discussion of Experiment 1

For luminance, luminance contrast, and orientation, the critical frequency of phase discrimination decreased with intersequence separation, in agreement with previous reports (Aghdaee & Cavanagh, 2007; Fujisaki & Nishida, 2010; Lappin, Tadin, & Whittier, 2002; Motoyoshi, 2004; Victor & Conte, 2002). In contrast, the critical frequency for motion was less affected by the increase in intersequence separation. The temporal limit for motion comparison was also much higher than the other features, and remained so (~12 Hz) even when the separation was 100°.

Experiment 2

Experiment 2a: Spatial configuration

The first series of experiments suggest the presence of a specialized mechanism in the brain that rapidly encodes long-range relationships between local motion signals. These experiments were restricted to pairings of horizontally oscillating elements with collinear oscillation axes. We next tested whether specific configurational relationships between local motions, such as similarity and collinearity, are necessary for rapid encoding of spatially remote local motions.

For comparison of remote orientations, the phase discrimination limit was previously found to be higher when the alternating orientations in the two sequences are the same than when they are different (Fujisaki & Nishida, 2010). We replicated this finding using the current setup, as shown in Figure 5a (right). This implies that the visual system is particularly sensitive in detecting simultaneous presentation of the same orientation at different locations. An advantage for identical sequences has also been suggested for color comparisons (Bartels & Zeki, 2006; Fujisaki & Nishida, 2010).

To examine the effect of similarity and configural relationship of local motions on remote motion comparison, we measured the critical temporal frequency for the five combinations of oscillation axes shown in the insets of Figure 5a. The two elements were placed symmetrically on opposite sides of fixation with a fixed intersequence distance of 40°. In the first three conditions, we rotated the oscillation axes while keeping them parallel. In the remaining two conditions, the oscillation axes were orthogonal to each other. “T-type” was a combination of horizontal and vertical axes, while “L-type” was a combination of two diagonal axes (45° and 135°).

Methods

Four observers including two of the authors participated. All of them also participated in Experiment 1a. The oscillation axes of motion sequences were varied to produce four combinations: 90° × 90° (vertical), 45° × 45° or 135° × 135° (diagonal), 45° × 135° (L-type), and 0° × 90° (T-type). In the data analysis, data from 45° × 45° or 135° × 135° conditions were merged for the fitting of psychometric functions after confirming that the data from these conditions did not show systematic differences. Observers were asked to judge which movie contained a pair with one of two possible relationships of motion direction. Each observer ran at least six staircases for each condition.

For the orientation sequences, the center orientation of the 1/f texture was one of four orientations (0°, 90°, 45°, or 135°). The texture orientation alternated between two orthogonal orientations (i.e., from 0° to 90° and back to 0°, or from 45° to 135° and back to 45°). One of the two texture patterns of motion sequence oscillated between 0° and 90°, and the texture pattern in the other sequence oscillated between 45° and 135°.

A follow-up experiment tested whether the main findings with motion stimuli were obtained with a larger number of observers. Ten observers did not include any authors. The intersequence separation was 100°. Fixing the oscillation frequency at 10 Hz, we...
measured the proportion correct for the following three configurations: horizontal parallel, L-type, and T-type. (Note that the result of the first control condition was already reported as Experiment 1c). The experiment was run in sessions, one for each of the three types of motion sequence. Each observer ran at least two sessions for each type of stimulus, for four sessions in total.

**Results**

The critical frequency was over 10 Hz in the conditions where the oscillation axes were parallel, regardless of whether the oscillation axes were horizontal (0°), diagonal (45° or 135°) or vertical (90°). Although the oscillation axes were collinear in the horizontal condition and not collinear in the diagonal and vertical cases, the critical frequencies for these three conditions were similar. In the remaining two conditions, the oscillation axes were orthogonal to each other. T-type was a combination of horizontal and vertical axes, while L-type was a combination of two diagonal axes (45° and 135°). In the L-type case, the critical frequency was again over 10 Hz. In the T-type case, however, the critical frequency fell down to ~4 Hz.

This difference between L-type and T-type configurations was confirmed by the results of a follow-up experiment where we tested a larger number of observers, using a larger intersequence separation. The performance of the L-type was as good as that of the parallel configuration (~80%, Figure 5b, left, middle), while the performance of the T-type was at chance level (Figure 5b, right). An ANOVA indicated a significant effect of configuration, $F(2, 18) = 33.38, p < 0.01$.

In agreement with the current results, Bartels and Zeki (2006) reported a fairly low temporal limit of pairwise motion comparison using a T-type configuration, but they used equiluminant chromatic motion known to be sluggish and did not compare the performance with other configurations.
Experiment 2b: Spatial tuning of the T-type configuration

The results shown in Figure 5 suggest that motion comparison is generally rapid except for the T-type configuration. However, since all four conditions other than the T-type were, in some sense, “good” configurations (with symmetry or repetition), the results do not exclude the possibility that motion comparison is generally slow except for certain good configurations. To distinguish between this and a specific impairment for the T-type configuration, we gradually changed the stimulus configuration from a T-type to one of the “good” configurations in two ways, examining whether slow comparison was observed only in the vicinity of the T-type configuration or if fast comparison was observed only in the vicinities of “good” configurations.

In the “position tuning” set of conditions (Figure 6), we gradually changed T-type configurations into L-type ones by vertically displacing one element from the center level line and then examined how the critical frequency varied as a function of the offset. In the “angle tuning” set of conditions (Figure 7), we gradually changed T-type configurations into parallel ones by rotating the axis of oscillation of one sequence from vertical to horizontal while keeping the oscillation axis of the other sequence horizontal.

Methods

Three observers including one of the three authors participated. All observers also participated in Experiment 1a.
For measurement of the position tuning, the position of the texture pattern in either the left or right sequences was displaced vertically, either upward or downward, from the horizontal center line, and the horizontal distance between the two sequences was varied in three steps from 10° to 40° (Figure 6a). The vertical offset was varied in six steps (0°, 1.25°, 2.5°, 5°, 10°, and 20°) for 20° horizontal distance conditions, and in five steps for other two conditions (0°, 1.25°, 2.5°, 5°, and 10° for 10° and 0°, 2.5°, 5°, 10°, and 20° for 40°). The oscillation axis in the left patch was vertical, and horizontal in the right patch. For each spatial configuration, observers participated in at least two staircase runs. In the data analysis, data from runs of the four conditions (at least eight runs in total) with the same vertical offset were merged for the computation of psychometric functions regardless of the direction of offset.

For measurement of orientation tuning, the oscillation axis of the right sequence was fixed to horizontal. The oscillation axis in left sequence took one of six angles from 0° to 90° (0°, 5°, 11°, 22°, 45°, and 90°). For each angle observers participated in at least six staircase runs.

The procedure was basically the same as for Experiment 1a. The spatial configuration of stimulus was fixed for each session. Observers viewed a pair of movies and judged which movie contained a designated temporal phase relationship. The oscillation temporal frequency was changed according to a modified 3-up, 1-down staircase rule (Dixon & Mood, 1948).

**Results**

In the position tuning experiment, the critical frequency, which was lowest at zero offset (T-type configuration), varied as a function of vertical displacement and horizontal distance.

Figure 7. The angle tuning of “T-type” configuration. (a) Schematic diagrams of stimulus configuration. (b) Results from three observers. Solid lines show the results of exponential regression. (c) Critical angle rotation from “T-type” configuration. Error bars show 95% confidence intervals (1.96 × SE).
configuration), showed a rapid improvement as the offset increased (Figure 6b). To estimate the threshold offset at which the critical frequency was the midpoint of the lowest (at zero offset) and the highest (asymptote) values, we fitted an exponential function ($y = a - be^x$, where $x$ is offset, $y$ is the critical frequency, and $a$, $b$, and $c$ are free parameters). The estimated threshold offset (average of three observers) was 1.1°, 1.5°, and 3.3° for 10°, 20°, and 40° separations, respectively. This narrow tuning (only about 10% of the intersequence separation, Figure 6c) implies that the critical frequency drops only in the vicinity of T-type configuration.

The angle tuning conditions showed that the critical frequency, which was the lowest at zero rotation (T-type configuration), increased rapidly as the rotation angle increased (Figure 7b). To estimate the threshold angle, we fitted the same exponential function as for the position tuning conditions. The estimated threshold was only ~15° for both 10° and 40° separations (Figure 7c). This shows that the critical frequency drops only in the vicinity of T-type configuration, just as it drops only in the vicinity of the L-type when angle is varied.

In sum, our data indicate that the phase-discrimination limit for motion is high and nearly invariant for a variety of oscillation axis pairings, including the cases where oscillation axes are different. We found a disadvantage of a specific configuration, the T-type. This specificity shows that the difference is not due to a simple advantage of pairing identical signals (such an advantage has been documented for orientation and color). Instead, these findings reveal configurational specificity of the process that encodes the pairwise relationship between remote local motion signals.

Experiment 3

Experiment 3a: Dichoptic presentation

Dichoptic stimulus presentation assesses whether the level of signal interaction resides before or after convergence of the signals from the two eyes. We measured the critical frequency of phase discrimination while presenting two motion elements to different eyes.

Four observers including two authors participated. All of them participated in the first series of experiments. A black matte partition was set between each observer’s eyes that divided the visual field vertically so that each eye could see only one of the two motion sequences. A red fixation cross was supplied for each eye. Observers were asked to fuse these crosses and maintain the fusion during each trial. The intersequence distance was 10°. Oscillation axes were vertical in both sequences. The phase discrimination limit was estimated in the same way as for the main experiment by merging data from six sessions for each observer.

The obtained temporal limit was almost as high as that obtained under binocular presentation (Figure 8a; $F(1, 3) = 2.34$, n.s.), suggesting that the comparison mechanism resides after convergence of monocular signals.

Dichoptic presentation reduces the magnitude of induced motion (Day & Wade, 1988). This suggests that the short-range center-surround interaction, which is supposed to cause induced motion, is not entirely binocular, and this view was further supported by a later study using binocular rivalry (Paffen, van der Smagt, te Pas, & Verstraten, 2005). If the short-range center-surround interaction is indeed partially monocular, it is distinct from the mechanism responsible for long-range motion comparison examined here. In agreement with the current finding, Mowafy et al. (1990) found that dichoptic presentation, along with spatial frequency difference, had little effect on discrimination of correlated and uncorrelated relative movements of two elements separated by a few degrees.

Experiment 3b: Inter- vs. intra-hemifield

In most of the conditions, we presented two motion elements symmetrically to the left and right of the fixation point. The two elements were projected at the same eccentricity of different hemifields. This might be a special condition for long-range signal comparison, because callosal connections connecting higher visual cortices preferentially connect points of the same eccentricity and distance from the vertical midline (Innocenti, 2009). If this type of cortical connection underlies the relative encoding, the critical frequency should be significantly worse when measured at different locations.

On the other hand, for some judgments independent mechanisms in the two hemispheres are involved. In multiple object tracking (MOT), for example, approximately twice as many targets can be successfully tracked when they are divided between the left and right hemifields as when they are all presented within the same hemifield (Alvarez & Cavanagh, 2005; Holcombe & Chen, 2012). The MOT task is thought to tap high-level attentive motion processing mediated by parietal and frontal cortices (Culham et al., 1998; Culham, Cavanagh, & Kanwisher, 2001; Jovicich et al., 2001). If the long-range motion comparison reflects the same high-level attentional resource as object tracking, performance should be worse when the two elements are presented within the same hemifield.

We therefore compared the performance of phase discrimination when the two elements were displaced...
horizontally across the two hemifields and when they were displaced vertically within the same hemifield.

Four observers including the two authors participated. All of them participated in the first series of experiments. The oscillation frequency in the two sequences was fixed at 12 Hz. The oscillation axes were either vertical or horizontal and were the same for the two sequences. In the bilateral condition, the two patches were equally and horizontally displaced in opposite directions from the vertical midline, either in the upper or the lower visual field. In the unilateral condition, the sequences were vertically displaced from the horizontal midline in either the left or right visual field (Figure 8b). The intersequence separation and eccentricity of sequences were fixed at 10 and 7.1 arc deg, respectively. In each session of the experiment, oscillation orientation was either vertical or horizontal. Observers performed a phase discrimination task for 80 trials for each of the four spatial configurations with a randomized order. We pooled the data for each observer and calculated the proportion correct for the bilateral and unilateral configurations. We merged data of the vertical and horizontal oscillation conditions and the two spatial configurations for each laterality condition because we did not find any systematic difference between them.

The results showed no significant difference in performance between the two hemifield conditions (Figure 8b). It seems that neither callosal connections connecting higher visual cortices nor hemisphere-specific resources of attention mediate the performance of long-range motion comparison.

**Experiment 4: Noise masking**

One might suggest that rapid motion comparison is based not on encoding of pairwise relationships of local motions, but on detection of global optic flows induced by the relative motion between an observer and the scene, such as large-field translation, rotation, and expansion, which has already been extensively studied (Duffy & Wurtz, 1991; Gibson, 1950; Graziano, Andersen, & Snowden, 1994; Tanaka & Saito, 1989). This might explain the deficit for the T-type configuration, because parallel and L-type configurations are parts of optic flow patterns, while the T-type configuration is not.

This interpretation (which we do not favor) requires that optic flow detectors be effectively activated even by only a pair of local motions. This has not been confirmed, but we cannot exclude it. Another assumption is that optic flow processing is sensitive to temporal asynchronies among component motions. This remains unexplored in both psychophysics and

Figure 8. The effects of dichoptic presentation and stimulus laterality. (a) Critical temporal frequencies with dichoptic presentation. Data are shown as means of estimated critical frequencies from four observers. Error bars show 95% confidence intervals (1.96 × SE). (b) The performance with bilateral and unilateral configurations. The mean correct ratios with a pair of motion elements with 12 Hz oscillation are shown. Data from two spatial configurations (upper/lower for bilateral and left/right for unilateral) and two oscillation orientations (horizontal and vertical) were merged. Error bars show 95% confidence intervals.
neurophysiology. Although it has been shown that optic flow processing has a long temporal integration window (Burr & Santoro, 2001; Domini, Vuong, & Caudek, 2002; Eby, 1992; Treue, Husain, & Andersen, 1991), this could reflect a property of processing subsequent to detection of the flow.

A critical difference between the two hypotheses is the range of predicted spatial pooling. While we propose a mechanism that compares a small number of local motion signals, an optic flow detector is expected to pool all local motion signals presented anywhere within its large receptive field (Morrone, Burr, & Vaina, 1995; Ohtani, Tanigawa, & Ejima, 1998; Snowden & Milne, 1997). This difference suggested the following noise masking experiment to test the optic flow hypothesis.

We measured the phase discrimination limit for a pair of vertically moving elements (targets) with concurrent presentation of four or two masker elements that were irrelevant to the task. Targets and maskers were positioned along a common circular path. Each masker element oscillated in directions tangential to the circular path, at a constant frequency of 10 Hz and in a random temporal phase. These maskers were designed to be inconsistent with the global rotation detector that might otherwise contribute to the target phase discrimination. Therefore, if the global rotation detector is indeed responsible for rapid phase discrimination between the target motions, the critical frequency should be impaired regardless of the distance between targets and maskers (Figure 9a).

**Methods**

Four observers including the two authors participated in this experiment. All of them also participated in Experiment 1a. The signal and distractor sequences were located on a 40° diameter circle centered at fixation (the “circular path”). The signal sequences occupied the same locations as they did in the configuration with 40° intersequence separation in the first experiment. The distractors were separated from the nearest signals by four distances (6.3°, 12.6°, 25.2°, and 28.2°; see Figure 9b). The temporal frequency of oscillation for distractors was fixed at 10 Hz. The oscillations were along the direction tangential to the circular path. The frequency of the signals’ oscillation was either 5, 8.6, 10, 12, or 15 Hz. The phase discrimination limits were measured using the method of constant stimuli. Within a session, the spatial configuration of signal and distractors was fixed and each temporal frequency condition was repeated 20 times, in randomized order. For each trial, observers...
saw the movie for 1500 ms and judged whether the pairs were moved in the same (in-phase) or opposite (reversed-phase) direction. Each observer performed three sessions for each spatial configuration. We also measured the upper limit for the signal pairs when presented without any distractors.

**Results**

The data did not support the prediction of the optic flow hypothesis (Figure 9b). The size of the masking effect was critically dependent on the target-masker distance. $F(4, 12) = 28.67, p < 0.01$. At the longest distance ($28.2^\circ$), we observed no significant drop-off from the control performance. Yet the maskers were potent in that as the target-masker distance decreased, the critical frequency gradually dropped, reaching half of the control performance at the shortest distance ($6.3^\circ$). Thus, the masking effect was mostly local, not global.

The local masking effect may reflect the well-documented crowding phenomenon (Toet & Levi, 1992). To compare the two target elements presented in the periphery, we believe observers had to select the targets with attention (rather than using a global optic flow integration process), and it is likely that the selection process was impaired by mask elements presented nearby.

The optic flow hypothesis is also not consistent with the spatial tunings of the T-type configuration (Figures 6 and 7). According to the optic flow hypothesis, the critical frequency should drop when the configuration deviates from those that match optic-flow templates. This is at odds with our finding that the critical frequency returned to the peak level when the configuration deviated only slightly from the T-type configuration. We therefore conclude that our findings do not reflect optic flow processing.

**General discussion**

Although encoding the relationship between separated local motion signals may be fundamental to the analysis of spatial motion patterns, it is poorly understood. We found that the limit to discriminate the temporal phase of two local motion oscillations was high (>10 Hz) over an extraordinarily wide range of intersequence separations (up to the maximum tested of $100^\circ$). Rapid comparison of local motions was found for a variety of configurations, including a case where local oscillations were orthogonal to each other, but the temporal limit dropped considerably when local oscillations made a T-type configuration. In this T-type configuration, uniquely, the oscillations did not share a common motion component.

**Rapid comparison of remote motion signals**

Discrimination of relative temporal phase between two sequences is increasingly used to evaluate the performance of feature binding based on temporal coincidence. Previous studies suggest that the temporal limit measured in terms of critical frequency is low when the binding is mediated by late attentive mechanisms, while high when it is mediated by early sensory mechanisms, and that early sensory comparators are most effective at short separations or for identical stimulus pairings (Aghdaee & Cavanagh, 2007; Amano et al., 2007; Arnold, 2005; Bartels & Zeki, 2006; Battelli et al., 2001; Fujisaki & Nishida, 2010; Holcombe, 2009; Holcombe & Cavanagh, 2001; Holcombe & Judson, 2007; Motoyoshi & Nishida, 2002; Victor & Conte, 2002). According to this theoretical framework, the present study indicates the existence of low-level yet long-range comparators of local motion signals. These comparators rapidly encode the relationship between local motions irrespective of how far they are separated in the visual field, unless they form T-type configurations.

The pairwise comparison processing investigated here may contribute to efficient detection of large-scale relative motion (Golomb, Andersen, Nakayama, MacLeod, & Wong, 1985; Watson & Eckert, 1994). It may also contribute to a variety of aspects of perception of motion patterns, including perceptual grouping and segmentation based on the Gestalt law of common fate, structure from motion, material perception, optic flow analysis, and biological motion perception.

The present findings indicate that the human visual system is more sensitive to remote relationships of local motion signals than those of luminance or orientation signals. This difference across attributes might reflect differences in natural image statistics. Since the luminance distributions of natural scenes typically have $1/f$ amplitude spectra (Burton & Moorehead, 1987; Field, 1987; Geisler, 2008; Ruderman & Bialek, 1994), there are some positive correlations of luminance across short separations, but they fall as the separation increases. Similarly, correlations of orientation exist for short separation due to the presence of smooth contours in natural scenes, but they rapidly fall with separation (Geisler 2008; Geisler & Perry, 2009; Geisler, Perry, Super, & Gallogly, 2001). In comparison with these attributes, long-range correlation is likely higher for motion, since global retinal motions could be often produced by movements of the observer’s eye, head, and body, as well as by movements of objects sitting on a common ground and those of crowds in...
harmony. In addition, considering the characteristics of our task, relevant natural image statistics are not only pairwise correlations of visual attributes, but also pairwise correlations of temporal changes in these attributes. One may suggest that changes in illumination could introduce a correlated temporal variation in luminance across a whole visual field, but this change is typically slow. It seems that rapid and long-range correlated changes in luminance and orientation are uncommon in natural situations, except for those accompanied by global movements. In order to verify our speculations, however, it is necessary to systematically investigate pairwise statistics of motion vectors in the retinal image under natural viewing conditions and to compare them with luminance and orientation statistics.

Slow comparison of T-type configurations

We found that the critical frequency dropped only in the vicinity of the T-type configuration. We also documented the change of the intersequence separation with position and angle tuning (Figures 6 and 7). We now consider the possibilities for what the critical configurational variable is that determines the critical frequency.

One hypothesis suggested by the current findings is that motion comparison is rapid if the two motions have components that displace in a common direction or opposite directions, while it is slow otherwise. Since a motion vector is two-dimensional (2-D), it can be decomposed into two orthogonal components. Our hypothesis states that fast comparison occurs thanks to same/different judgment of at least one of the two components. To define the two components, the component axes must be specified. The current results can be accounted for if we assume that the axes are defined in relation to the locations of two elements—one axis is along the line connecting the two locations, and the other axis is orthogonal to it. For these coordinates, there is no common component for T-type configurations.

Three considerations support this hypothesis of vector component decomposition defined by an axis connecting the two stimuli. First, it is an extension of the notion of fast comparison of identical pairing that is true for other attributes. Second, that the visual system decomposes motion vectors into common and uncommon components is already well known (vector analysis; Johansson, 1973). Third, this hypothesis is consistent with the effects of intersequence separation on the position tuning (Figure 6c) and angle tuning (Figure 7c) for the T-type configuration. According to the hypothesis, the critical condition for rapid motion comparison is the presence of a common motion component. Then the threshold is expected to be dependent on the angle of the axis of each sequence relative to the line connecting the two sequences. In the case of position shift (Figure 10a), given the threshold angle is constant, the threshold shift should increase with the separation. In the case of angle rotation (Figure 10b), on the other hand, the threshold angle should remain constant regardless of separation. These predictions are consistent with the obtained results (Figures 6 and 7).

An alternative hypothesis is that motion comparison is fast if the visual system can find an imaginary path that smoothly links the two local motions, and slow otherwise (Figure 11). Unlike other configurations, in
T-type configurations any continuous connection path would have high curvature and an inflection point. Although this explanation is intuitively appealing, we do not think it as promising as the vector analysis hypothesis. First, this hypothesis cannot explain rapid motion comparison for similar motion directions that are near parallel rather than collinear. Second, the definition of smoothness can be arbitrary. Third, if a smooth connection is the critical factor, the shorter the path, the harder it would be to connect different directions of motion (Figure 11b). Then the threshold shift and angle should decrease with the intersequence separation, but the results did not support this prediction (Figures 6 and 7).

Even if it is true that the lack of common motion vector components causes a slow motion comparison for T-type configurations, why this would be so is unknown. In our subjective experience, it is hard to see a global structure in common across two motions in a T-type configuration. Our result, therefore, is consistent with the idea that the relationship of local motions should be rapidly encoded only when the two local motions are connectable or relatable (Kellman & Shipley, 1991), and likely to belong to a common rigid/nonrigid entity. The presence of common motion components could be a critical stimulus property for the visual system to find connectability, and this might reflect natural image statistics in the retinal image of the natural scene. It is plausible that correlated changes in motion are very uncommon without the presence of common vector components. It is also plausible that this is because correlated changes in motion are associated with physical connection of the objects in the scene. Future investigation of natural image statistics of retinal motion would clarify the relationship with the present findings. We also suspect that the critical role of common motion vectors may reflect a general processing structure of visual motion analysis, including vector decomposition of complex motion flows.

Related studies

The work of Lappin and colleagues (Lappin et al., 2001; Lappin et al., 2002; Mowafy et al., 1990; Mowafy, Lappin, Anderson, & Mauk, 1992) is also consistent with specialized and efficient long-range processing of the relationship of two motions, although they tested only very slow speeds and relatively small intersequence separations. Rainville and Makous (2001, 2002) also investigated sensitivity to the synchrony of two random oscillations, although so far these were published only as conference abstracts. They used one-dimensional (1-D) local motions, while we used 2-D ones. Since the rule of spatial interaction is different for 1-D and 2-D motion signals (Amano, Edwards, Badcock, & Nishida, 2009), it remains open how much the current findings are relevant to those made with 1-D stimuli (Lorenceau & Zago, 1999; Rainville & Makous, 2001, 2002).

The fast temporal limit suggests that local motion comparison is mediated by a low-level pre-attentive process since it is much higher than that for the temporal limit for the comparison by high-level processes (Fujisaki & Nishida, 2010; Holcombe & Cavanagh, 2001). The resulting motion grouping can then modulate how attention propagates over space (Festman & Braun, 2010, 2012). However, when the display contains task-irrelevant motions in addition to a pair of targets, as in our noise masking experiment, the observer must select the two target locations, probably by using selective attention. Our finding, therefore, is not inconsistent with a recent suggestion that visual search for a common-fate target pair is based on serial attentional selection (Levinthal &

Figure 11. Explanation by “smoothness of connecting path” hypothesis. (a) Schematic diagrams of position tuning. (b) Schematic diagrams of angle tuning. The dotted red lines denote the path connecting between two motion vectors. Red circles show the curvature of connecting path. When the intersequence distance increases, curvature of the connecting path becomes shallow, i.e., connecting with a smooth path becomes easier.
Franconeri, 2011). That is, we conjecture that attention is necessary for selection of a pair of local motion positions under cluttered situations as suggested by Levinthal and Franconeri (2011), while attention is unnecessary for synchrony detection of the selected motion pair as shown in the present experiments.

The specialized processing for pairwise motion comparison that we revealed here is distinct from that responsible for segregation of a region of local motion elements that make a synchronous direction change (Lee & Blake, 1999, 2001). In that segregation task, observers can sometimes discriminate synchrony from asynchrony without necessarily being able to do what was required in our task, where the stimuli were always synchronous and observers identified the relationship of the feature values.

Neural correlate

The results obtained suggest that the mechanism of remote local motion comparison is motion-specific, rapid, binocular, and covers a wide field. These characteristics suggest that early visual areas such as V1 or V2 are unlikely to be the responsible cortical area. If local 2-D motion is computed by the V1-MT circuit, it is likely that comparison processing occurs in subsequent stages. These considerations suggest the medial superior temporal (MST) area as a candidate for the neural correlate of rapid local motion comparison. The receptive fields of MST neurons often subtend over 50° and extend across the hemifield border (Raiguel et al., 1997). MST responds quickly to motion input (Lagae, Maes, Raiguel, Xiao, & Orban, 1994; Thiele, Henning, Kubischik, & Hoffmann, 2002). In MST, many neurons are sensitive to optic flow patterns, such as rotation and expansion (Andersen, Snowden, Treue, & Graziano, 1990; Duffy & Wurts, 1997; Graziano et al., 1994; Tanaka et al., 1986). As noted above, our data indicates that the mechanism to compare remote local motions is distinct from the mechanism that pools local motion signals over a wide visual area to compute optic flow caused by movement of the observer. However, since optic flow processing requires spatial comparison of local motions, the remote motion comparison may be the precursor of optic flow processing, as well as that of other motion pattern analyses. Alternatively, an optic flow detector might dramatically change its pooling area via attentional control, being able to become a pairwise motion comparator when it is the task requirement (Burr, Baldassi, Morrone, & Verghese, 2009). For the time being, we suggest that encoding of the relationships between local motion signals with high temporal precision may be a visual function of MST.

Conclusions

The human visual system has a low-level and long-range mechanism that rapidly encodes the relationship of local motions presented anywhere in the visual field whenever they have a common component vector. This unique mechanism for visual motion processing likely contributes to quick appreciation of global dynamic scene structure.

Keywords: visual motion perception, temporal processing, long-range processing

Acknowledgments

SN was supported by KAKENHI (Grant-in-Aid for Scientific Research on Innovative Areas No. 22135004). AOH was supported by an ARC Future Fellowship. The demonstration program is available at Visiome platform (“http://visiome.neuroinf.jp/”).

Commercial relationships: none.
Corresponding author: Kazushi Maruya.
Email: kazushi.maruya@gmail.com.
Address: NTT Communication Science Labs, Nippon Telegraph and Telephone Corporation, Japan.

Footnotes

1 This result is also shown as the data of a control condition of Experiment 2a (Parallel condition in Figure 5b).

2 Although “optic flow” has a broader meaning in other contexts, here we use this term only to refer to the global flow patterns induced by self-movement.

References


Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images...


