Spatial pooling of one-dimensional second-order motion signals

Kazushi Maruya

Shin’ya Nishida

We can detect visual movements not only from luminance motion signals (first-order motion) but also from non-luminance motion signals (second-order motion). It has been established for first-order motions that the visual system pools local one-dimensional motion signals across space and orientation to solve the aperture problem and to estimate two-dimensional object motion. In this study, we investigated (i) whether local one-dimensional second-order motion signals are also pooled across space and orientation into a global 2D motion, and if so, (ii) whether the second-order motion signals are pooled independently of, or in cooperation with, first-order motion signals. We measured the direction-discrimination performance and the rating of a global circular translation of four oscillating bars, each defined either by luminance or by a non-luminance attribute, such as flicker and binocular depth. The results showed evidence of motion pooling both when the stimulus consisted only of second-order bars and when it consisted of first-order and second-order bars. We observed global motion pooling across first-order motion and second-order motions even when the first-order motion was not accompanied by trackable position changes. These results suggest the presence of a universal pooling system for first- and second-order one-dimensional motion signals.

Keywords: aperture problem, second-order motion, global motion


Introduction

Movements of various image attributes, including luminance, spatiotemporal luminance contrast, and binocular depth, yield motion sensations. While several motion classifications based on different theoretical viewpoints have been proposed (e.g., Cavanagh & Mather, 1989; Chubb & Sperling, 1988; Lu & Sperling, 1995), this study simply refers to luminance-based motion as first order and all non-luminance-based motions as second order. A large body of experimental evidence supports separate detection of first-order and second-order motions (Badcock & Derrington, 1985; Chubb & Sperling, 1989; Ledgeway & Hutchinson, 2005; Ledgeway & Smith, 1994; Lu & Sperling, 1995; Nishida, Ledgeway, & Edwards, 1997; Pantle & Turano, 1992; Smith & Ledgeway, 2001; Vaina, Cowey, & Kennedy, 1999), although the possibility of a partial overlap cannot be excluded (Benton, Johnston, McOwan, & Victor, 2001; Johnston, Bent, & McOwan, 1999; Johnston & Clifford, 1995a, 1995b).

When motion information is extracted from a one-dimensional (1D) edge or when an image motion is detected with a 1D local motion sensor with an oriented receptive field, the resulting motion signal is ambiguous in that it allows multiple 2D motion interpretations (aperture problem). To estimate true 2D object motion from 1D local motions, the visual system pools 1D motion signals across different orientations, as well as across different spatial locations. This process has been extensively studied for first-order motion but is not yet fully understood for second-order motion.

Using plaid motion displays (Adelson & Movshon, 1982), a standard stimulus for investigating the visual process underlying the aperture problem, several studies examined motion pooling of 1D second-order motions. Even on initial introduction of the notion of second-order motion, it was recognized that plaid motion stimuli consisting of a couple of second-order component motions could produce integrated coherent motion perception (Cavanagh & Mather, 1989). Coherent motion is perceived even when the plaid motion is type II (Cropper, Badcock, & Hayes, 1994; Kim & Wilson, 1993; Wilson & Kim, 1994). With regard to the interaction between first- and second-order signals, Wilson, Ferrera, and Yo (1992) proposed their famous model of luminance-based plaid motion perception, in which second-order motion is integrated with first-order motion. Stoner and Albright (1992) showed that coherent motion is perceived with a plaid motion stimulus consisting of luminance-defined and
The present study examined (i) whether 1D second-order motion signals are integrated across space and orientations, and if so, (ii) whether 1D second-order motion signals are integrated with motion signals defined by different attributes, including first-order motion signals. Instead of using plaid displays, we used a multiple aperture stimulus (Lorenceau & Zago, 1999; Takeuchi, 1998). The stimulus consisted of four sinusoidally oscillating bars, simulating a situation where four edges of a diamond are translating along a circular path, and were viewed through four Gaussian apertures, each located at the center of each edge. When observers could appropriately integrate coordinated oscillations of four bars across space and orientation, they perceived global translation of the virtual diamond along a circular path. To measure the objective performance of motion integration, we asked observers to judge the direction of global circular translation, i.e., clockwise or counterclockwise. We expected that this sensitive task could effectively detect the sign of coherent motion even when the overall impression of motion coherence was not high. In addition to the direction judgment, we asked observers to rate the comprehensive quality of global motion to obtain data directly comparable to coherent motion judgments for plaid stimuli.

The first experiment examined motion pooling using luminance-based (first-order) and flicker-defined (second-order) motions. The second experiment examined motion pooling using luminance-based (first-order) and stereo-disparity-defined (second-order) motions. The third experiment reexamined integration between first and second-order signals, using luminance-based motion that was not accompanied by position shift. The results show that the second-order 1D motion signals are pooled and that they contribute to the solution of the aperture problem in cooperation with first-order motion signals.

Experiment 1: Flicker-defined second-order motion

Methods

Observers

Eleven observers including the two authors participated in the main experiments. All observers had normal or corrected-to-normal visual acuity. In all the experiments described in this paper, informed consent was obtained before the experiment started, and the experiment was approved by NTT Communication Science Laboratories Research Ethics Committee.

Apparatus

Stimuli were generated using the ViSaGe system (Cambridge Research Systems) driven by a host DOS/V computer with a gamma correction. They were displayed on a 21-inch CRT monitor (SONY GDM-F520). The display had a refresh rate of 100 Hz with a spatial resolution of 1024 × 768 pixels. Each pixel subtended 3 arcmin at a viewing distance of 39 cm. A chin rest was used to stabilize the observer’s head.

Stimuli

The stimulus consisted of four elements that centered at vertices of a virtual rectangle (Figure 1a) and was presented on the uniform gray background (30 cd/m²). The distances between the centers of each element were 12.8 arc deg. Each element was a diagonal bar moving on a noise dot field (where bright and dark pixels appeared independently with equal probability).

Three types of stimulus were used (Figure 1b): all moving bars were defined by luminance (L-type, Movie 1); all bars were defined by flicker (F-type, Movie 2); a combination of luminance- and flicker-defined bars (L × F, Movie 3). For L-type stimulus, a luminance-defined moving bar was generated by modulating the mean luminance of the uniform dynamic noise field (85% contrast, 100-Hz
Figure 1. Across-space integration of flicker-defined motion. (a) A diagram of the arrangement of four local elements. (b) A diagram of three stimulus types for Experiment 1. Note that dotted lines in the flicker-defined component were not displayed in the actual stimulus display. (c) Results of Experiment 1. Red open rectangles denote mean correct ratio with a direction judgment task. Data are plotted on the left vertical axis. Gray bars denote mean scores with a quality rating task. Data are plotted on the right vertical axis. Dotted lines show a chance level (50%) and 75% and 100% performance levels in direction judgment. Error bars show the 95% bootstrap confidence intervals.
update)—a 15% step luminance increment within the bar and a 15% decrement outside of the bar. A flicker-defined bar was generated by modulating the update rate of dynamic noise (100% contrast)—100 Hz within the bar and 0 Hz (i.e., static) outside of the bar. In either case, the field luminance contrast was modulated by a static Gaussian envelope (standard deviation = 2.14 arc deg).

The width of a moving bar was 48 arcmin. The bars were oriented orthogonally with respect to the radial direction and oscillated sinusoidally at 1 Hz. The bars in the opposing positions (e.g., upper left and lower right) had the same oscillation phase, while bars in adjacent positions (e.g., upper left and upper right) had a 90-deg phase difference. This motion pattern was designed to generate a translational global motion along a circular path. The diameter of the path was 2.56 deg and the circulation speed was 1.0 rev/s, which yielded the tangential speed of approximately 8.0 deg visual angle/s. Stimulus duration was 500 ms.

Procedure

Observers performed two tasks in separate blocks. In the first block, observers judged the direction of global circular translation (clockwise/counterclockwise). The block consisted of sessions where one of three stimulus types (L/F/L × C2) was repeatedly displayed 20 times. The

Movie 1. A demonstration of L-type stimulus. All four element-motion bars are defined by luminance increments. The stimulus used in the actual experiments was slightly faster, since this QuickTime movie plays an original 100-fps movie at 60 fps.

Movie 2. A demonstration of F-type stimulus. All four element-motion bars are defined by flickers of random noise. The stimulus used in the actual experiments was slightly faster, since this QuickTime movie plays an original 100-fps movie at 60 fps.

Movie 3. A demonstration of L × F type stimulus. Two element-motion bars (upper right and lower left) are defined by luminance increments, and the other two bars are defined by flickers. The stimulus used in the actual experiments was slightly faster, since this QuickTime movie plays an original 100-fps movie at 60 fps.
order of the three stimulus conditions was randomized within a block. Observers performed two sessions for each stimulus condition in a randomized order. In total, they performed six sessions (120 trials). In the second block, observers evaluated the quality of global motion with using a 5-point rating procedure. Observers were explained the notion of quality of global motion in advance and asked to rate the quality of global motion based on several aspects, such as robustness of global form, smoothness of global motion, and difficulty in direction judgment. The basic procedure was the same as for the direction judgment blocks except for the number of trials. In this block, observers repeated ten trials for each session and performed two sessions for each stimulus type.

Results and discussion

For each observer, the correct ratio of the direction judgment was computed from 40 trials and the mean rating value from 20 trials. The correct ratio and the rating value averaged across observers are shown in Figure 1c. As for the L condition, observers generally showed high performance. The mean correct ratio was almost 100%, and the mean rating value was close to the maximum rating score (i.e., 5). As for the F condition, the correct ratio was fairly high (>95% for 10 of 11 observers), while the rating value was not so high (~3). For the L × F condition, the correct ratio was slightly lower than the L or F condition, but the performance was still high (~90%). The mean rating value (~3) was similar to the F-only condition. The results of a one-way ANOVA with repeated measures show that the differences of mean correct ratios (with an arcsine transformation) between conditions were significant \( F(2,20) = 6.78, p < 0.01 \). The results of a post-hoc multiple comparison by LSD revealed that there was a significant difference between the direction judgment performance in the L and L × F conditions \( p < 0.05 \). The results of a one-way ANOVA comparing the rating values showed that there was a significant difference between conditions, and post-hoc LSD revealed that differences between the L and F conditions and those between the L and L × F conditions were significant \( F(2,20) = 25.24, p < 0.01 \); LSD post-hoc: \( p < 0.05 \).

In summary, observers could judge the direction of global circular translation for all the conditions with high direction-discrimination performance, although the rating of motion quality was higher for the L condition than for the other two conditions that contained F-type elements. The high performance in the direction judgment task indicates that both the motion integration for flicker-defined motion and the integration between luminance- and flicker-defined motions are possible across space and orientation. The integration with flicker-defined motions, however, is not as effective as in the case where only luminance-defined motion signals are integrated.

Readers, upon watching the demonstrations of the F-type stimulus (Movie 2) and the L × F type stimulus (Movie 3), would see for themselves that the direction of global motion is discernable, but the four bars often appear to move discretely, and that the global motion of a large diamond consisting of four bars is not clearly perceived. This is in marked contrast to the appearance of the L-type stimulus (Movie 1), in which a coherently moving diamond is robustly perceived.

One might have a concern that the observed high direction-discrimination performance does not necessarily be the evidence of spatial pooling of 1D second-order signals, since observers might be able to cognitively infer the direction of global motion from the phase relationship of element motions. We have two arguments against this criticism. First, even when we halved the duration of stimulus (from 500 to 250 ms) to prevent cognitive strategies, the performance was still far above the chance level for all three observers that we tested (~95% for the F condition and ~90% for the L × F condition). Second, it is known that observers cannot perform this direction-discrimination task even when they can clearly see the phase relationship of element motions, such as when the spatial frequency of moving carrier is significantly different between elements (Maruya, Amano, & Nishida, 2010). We therefore consider that the observed high discrimination performance cannot be ascribed to the observers’ guessing from the phase relationship between element motions.

Supplemental experiment

In the main experiment, we used flicker-defined second-order motion to show motion integration within second-order motion and between first- and second-order motions. Another, and presumably more popular, type of second-order motion is the motion defined by contrast modulations. To investigate the generality of our finding, we also tested, with a few observers, whether similar results could be obtained with contrast-defined second-order motion. It is possible that flicker-defined and contrast-defined second-order motions are detected by the same neural mechanism. This is because the difference in the update rate between the moving bar and background in flicker-defined motion display can cause contrast modulations under the temporal averaging in some stages of our visual system. In addition, it is theoretically possible for a common preprocessing mechanism (spatiotemporal filters followed by a rectification) to make the second-order motion signals visible to the standard motion analyzers. Nevertheless, the relationship between different types of second-order motion has not yet been fully understood. We think it still a matter of empirical investigation whether flicker-defined and contrast-defined second-order motions behave similarly in various cases including motion integration. The effect of carrier temporal structure (static/dynamic) is another issue that we could see in this additional experiment.
Contrast-defined moving diagonal bars were 75% contrast patterns produced on the 25% contrast background (modulation contrast: 50%). The stimulus field was either a static or dynamic noise field (Cs or Cd motion). The background noise was updated at 100 Hz for Cd-type elements. Luminance artifacts potentially contained in these patterns were removed for each observer by using the minimum motion technique (Ledgeway & Smith, 1994). We tested all homogeneous and heterogeneous combinations of Cs-, Cd-, L-, and F-type elements. The total number of combinations was seven, excluding the three conditions of Experiment 1. Three observers, one naive and the two authors, participated in this experiment.

Figure 2 shows the correct ratio and mean rating value, separately for each observer. Data for the same observers in the main Experiment 1 are plotted together for the purpose of comparison. As for the Cs and Cd conditions, all of the three observers could judge the direction of global motion almost completely. The rating value was 3–4, which was slightly worse than that for the F-only condition obtained with the same observers. As for the L × Cs and L × Cd conditions, observers could judge the global motion direction with the accuracy of ~90%. The rating score was 3.5–4.5, except for the L × Cs condition of observer SN (~2). When different second-order elements were combined (Cs × Cd, Cs × F, Cd × F), although there were some individual differences, the direction-discrimination performance was always higher than 75%, and the rating score was 3–4 except for a few cases. These results indicate that motion integration is possible also for stimuli involving contrast-defined second-order motions, although the rating score was slightly lower than that for the stimuli containing flicker-defined elements.

Readers may wonder to what extent the visibility of local motion varied across L-type, F-type, and C-type stimuli and how the local motion visibility could affect the performance of motion integration for each stimulus. To address these questions, we carried out additional experiments with two observers (Appendices A and B). First, for each motion type, we measured the modulation threshold of local motion detection and evaluated the local motion visibility in terms of multiples of threshold (Appendix A, Table A1). The visibility was ~30 times the threshold for...
L-type, while it was significantly lower for F-type (×5), Cs-type (×6–7), and Cd-type (×4–5). Next, we measured the performance of global motion integration as a function of the local motion visibility (the magnitude of stimulus modulation expressed in terms of threshold multiples; Appendix B). The results showed that motion integration was improved as the local motion visibility was increased. The rate of improvement, however, varied among different motion types, with the slope being steeper for the L condition than for the other motion conditions (Figure B1). This implies that motion integration is easier for L-type than for the others even when they are compared at equal visibility. These findings indicate that the differences between stimulus types that we observed in the main experiment can be partially but not exclusively ascribed to the visibility difference of local motion signals.

Experiment 2: Disparity-defined second-order motion

There is a variety of second-order motion, depending on the moving feature attributes. The second-order motion used in the first experiments could be processed, at least theoretically, by relatively simple non-linear preprocessing followed by the standard motion analysis (Chubb & Sperling, 1988). In contrast, it has been suggested that the movements carried by more complex attributes, such as binocular disparity and motion direction, might be processed by “high-level” motion mechanisms distinct from “low-level” second-order mechanisms (Lu & Sperling, 1995, 2001; Zanker, 1993; but also see Patterson, 1999, 2002). In this experiment, we examined motion integration with disparity-defined motion and asked whether such complex motion also contributes to the solution of the aperture problem.

Methods

Ten observers including the two authors participated in this experiment. All of them participated in the first experiment.

A stereo stimulus was presented using liquid crystal shutter goggles (NuVision 60GX) controlled by the ViSaGe system. A movie refreshed at 120 Hz was alternately displayed on each eye through the pair of liquid crystal shutters. Thus, observers viewed a dichoptic stimulus pattern refreshed at 60 Hz on each eye. All other details of apparatus were the same as for the previous experiment.

The diameter of the path for global circular translation was 3.84 deg and the speed was 0.6 rev/s, corresponding to 7.2 deg visual angle/s. The stimulus duration was 750 ms. We used three types of stimulus combinations for the disparity-defined motion condition. In the disparity (D) condition, all four elements were disparity-defined moving bars generated using a dynamic random dot stereogram. The random dot pattern was made of numerous binary 12 × 12 arcmin dots and updated to a new pattern at 60 Hz. A moving bar was rendered on the zero-disparity plane and the surround noise field was rendered on a far plane presented at an uncrossed disparity of 24 arcmin. A preliminary experiment that measured the performance of direction judgment for a variety of depth arrangements of the bar and surround indicated that the depth arrangement affected the performance, with the best performance being obtained with the depth arrangement used in this experiment (see Appendix C). The diagonal bars shifted every four frames, i.e., every 67 ms. In the other two conditions, named L × D and F × D conditions, luminance- or flicker-defined motion elements were displayed in combination with disparity-defined motion elements. Parameters for luminance- and flicker-defined elements were the same as in the first experiment except for the following points caused by a change of the refresh rate of the display. The background noise of luminance elements was updated at 60 Hz. A flicker-defined bar was generated by modulating the update rate of dynamic noise, 60 Hz within the bar and 0 Hz (i.e., static) outside of the bar. The diagonal bars in these elements shifted every 67 ms and their displacements were calculated on the basis of the same global circular translation as for the disparity-defined elements.

Results and discussion

Figure 3 shows the correct ratio of the direction judgment and the rating value averaged across observers. Under the D condition, observers showed good performance in judging the direction of circular translation (around 90%). However, they rated the quality of global motion fairly low. The mean rating score was around 2. Under the L × D and F × D conditions, the mean correct ratio was 80%. This performance level was slightly lower than that observed for flicker-defined motion conditions (Experiment 1) but much higher than the chance level, which indicates that global motion integration is possible. The rating values of these two conditions were around 3, which was similar to the rating value of the L × F condition in Experiment 1. In summary, the results indicate that motion integration across space and orientation is possible either when the stimulus consists of disparity-defined motion or when the stimulus is a combination of disparity-defined motion and motion of another type.

In comparison with the D condition, the rating score significantly increased for the L × D and F × D conditions where disparity-defined elements were combined with other types of elements (F(2,18) = 5.57, p < 0.05; LSD post-hoc: p < 0.05). In contrast, the direction judgment for these heterogeneous conditions was slightly worse than for the D condition yet the difference was not
statistically significant \( F(2,18) = 1.52, \text{ n.s.} \). One may wonder why the rating score obtained with D-type stimuli was low yet the direction discrimination was accurate. Since the rating score includes evaluations of smoothness and quality of global motion, which are affected by the visibility of local motion, the low rating score could be due to the difficulty in seeing local disparity-defined motion. Several naive observers reported difficulty in seeing D-type local motion in the post-hoc oral assessment. An additional experiment with a couple of observers also showed that the strength of local motion was only 1.3–2.5 times the direction-discrimination threshold (Table A1).

Experiment 3: Four-stroke first-order motion

So far, the results indicate that global 2D motion can be perceived for any combinations of 1D motion elements. These results support the presence of a universal motion pooling system that pools any type of 1D local motion signals to recover the true 2D motion.

However, there remains an alternative explanation that does not need to assume universal motion pooling. The movement of luminance-defined bars used in the previous experiments can be detected not only by first-order motion sensors but also by higher level mechanisms, such as that tracking the position shift of salient features (Cavanagh, 1992; Lu & Sperling, 1995; Seiffert & Cavanagh, 1998, 1999). Therefore, it is possible that global motion perception with our heterogeneous stimuli might be mediated by integration of homogeneous local motion signals detected by a high-level second-order motion system.

For the purpose of clarifying whether second-order motion is integrated with luminance-based motion via first-order motion signals or second-order signals of the luminance-based motion, the third experiment used a special luminance-based motion that was not effectively detected by second-order mechanisms. We generated a motion stimulus that did not contain any consistent...
position shift of a particular feature, by using an illusory motion called “four-stroke apparent motion,” which is a variant of reversed-phi motion display (Anstis, 1970; Anstis & Rogers, 1986).

A diagram of a simple four-stroke display is shown in Figure 4a. A dark bar is displayed in the first frame, and the bar shifts to the right in the second frame. In the third frame, the contrast of the stimulus image is reversed and the position of the bar is reset to the initial position. In the fourth frame, the bar again shifts to the right. When these four operations are repetitively applied, the observer perceives unidirectional rightward motion. Here, the unidirectional motion perception is supposed to be mediated by early motion energy detection (Adelson & Bergen, 1985; Sato, 1989). The spatiotemporal pattern of the stimulus described above matches the receptive field of rightward motion detectors elongated diagonally right and down but not that of leftward detectors. Nevertheless, the mere tracking of the moving bar produces no unidirectional motion.

Applying this rule, we generated apparently oscillating motion elements while keeping the apparent position of the bar almost the same and thus making position tracking ineffective in perceiving the oscillation (Lf-type elements; Figure 4b). We displaced bars in every one of two motion frames and reversed the contrast of the whole element pattern in the other frame. The displacement was set to produce the same speed as a normally oscillating bar at that motion frame. That is, the generated motion was a motion sampled from the original motion in every two frames. Since the frame rate was fairly high, the displacement of bars was too small for positional tracking, while the motion was perceived approximately with the correct speed and phase of oscillation.

Figure 4. Across-space integration of the four-stroke first-order motion and second-order motion. (a) Diagram of original four-stroke apparent motion display. (b) Diagram of Lf-type element. Mean luminance within the local area is plotted in the space–time plot. (c) Results of Experiment 3. Data are shown in the same format as in Figure 1.
Methods

Eleven observers including the two authors participated in the main experiments. All of them participated in the first experiment.

When the stimulus contained the disparity-defined motion, we used the same experimental setup as in the second experiment. For the other stimuli, we used the same experimental setup as in the first experiment.

In this experiment, all the stimuli contained special luminance-based elements making four-stroke apparent motion (Lf-type element, see Movie 4). Each frame of an Lf-type element was a diagonal bar defined by a 15% increment/decrement of the mean luminance of dynamic noise pattern (85% contrast, 100-Hz update). In the way described above, the motion of diagonal bars produced first-order motion without coherent, unidirectional displacements. The duration of each motion frame was 20 ms (or 33 ms when combined with disparity-defined elements). The diagonal bars shifted at every odd frame, with the displacement corresponding to the instantaneous speed of the sinusoidal oscillation. At every even frame, the bars jumped back to their home positions with the polarity of luminance contrast reversed. Stimulus duration was 500 ms (or 750 ms when combined with disparity-defined motion).

In one condition, the stimulus consisted of four Lf elements (Lf condition, Movie 4). In the other three conditions, two Lf elements were combined with two luminance-, flicker-, or disparity-defined elements (L × Lf, Lf × F, Lf × D conditions). The L-, F-, and D-type elements were the same as those used in the first two experiments, except for the luminance setting of L-type bars—a 50% luminance increment within the bar and a 50% decrement outside of the bar, against a 50% contrast dynamic noise field.

The observers’ task was the same as in the previous experiments, i.e., to judge the direction of global rotation (clockwise/anti-clockwise) or evaluate the quality of global motion perception in five steps. The two tasks were performed in different sessions. Observers ran 20 trials for direction judgments, and 10 trials for rating in each session, and repeated two sessions for each stimulus-type condition. The session order was randomized for each observer and stimulus-type condition. All observers ran all four stimulus conditions except for one observer who ran only three (Lf, L × Lf, Lf × D) conditions.

Results and discussion

Figure 4c shows the results. The direction judgment was generally accurate for the Lf, L × Lf, and Lf × F conditions. The mean rating value was high (~4) for the L × Lf and Lf × F conditions. It was reduced but still larger than 2 for the Lf condition. In contrast, for the Lf × D condition, the mean correct rate was close to the chance level, and the mean rating was less than 2. These findings support the existence of a universal 2D motion process that can integrate first-order motion signals with second-order motion signals, at least those produced by flicker-defined stimuli.

The results shown in Figure 4c do not provide clear evidence for the integration of first-order motion signals with disparity-defined motion signals. This result can be interpreted to indicate separate pooling of first-order motion and high-level position-based second-order motion. However, our failure to find motion pooling for the Lf × D condition might be simply because the visibility of local motions was too low to perform the task. The Lf-type motion was very noisy. The disparity-defined motion was also noisy as had been suggested in Experiment 2. This might be why the integration of this combination was very ineffective. In agreement with this idea, two of the ten observers showed 70–75% direction discrimination for the Lf × D condition. Furthermore, in an additional test with three observers [the two authors (KM, SN) and a naive observer, MT], we reduced the stimulus speed (a quarter of the original speed) and increased the stimulus duration (1 s) and found a significant improvement in task performance for the Lf × D condition. The correct direction-discrimination ratio was
95, 100, and 87.5%, and the global motion ratings were 3.95, 3.6, and 3.0 for MT, KM, and SN, respectively (while they were 58, 75, 50% and 2.8, 2.35, and 1.38 in the original condition). Therefore, it is premature to conclude that integration of first-order motion and disparity-defined motion is impossible.

**Supplemental experiment**

As in the first experiment, we also examined whether typical contrast-defined motion also contributes to the solution of aperture problem in cooperation with first-order motion signals, which were now produced by Lf-type elements. The apparatus and stimuli were basically the same except that contrast-defined elements were used. The contrast-defined elements were the Cs- and Cd-type elements used in Experiment 1 (supplement) with 100% contrast. Three observers, one naive and the two authors, participated in this experiment.

Figure 5 shows the results, together with the results for the same observers in the main Experiment 3. The direction discrimination was fairly accurate for the Lf × Cs and Lf × Cd conditions. This performance was comparable to that for the Lf × F condition, although the rating was slightly lower. The results imply successful global motion integration for combinations of contrast-defined second-order motion and pure first-order motion generated by four-stroke apparent motions, supporting the notion of a universal 2D motion process that can integrate first-order motion signals with second-order motion signals.

![Figure 5](https://jov.arvojournals.org/)

Figure 5. Results for the supplemental experiment of Experiment 3. Data for each observer (MT, KM, and SN) are shown in the same format as in Figure 1.
This supplemental experiment also included another version of the Lf condition whose parameters were matched with the Lf × D condition (60-Hz update; 750-ms duration). Under this condition, all observers could judge the direction of global motion almost perfectly.

General discussion

The results show that observers could judge the direction of global 2D circular translations by integrating local 1D motion signals carried by second-order motions. Even in the conditions where the global motion stimulus contained both first- and second-order local motions, observers could judge the global motion direction. The performance in these conditions was at the same level as in the conditions with homogeneous second-order motion component (F, D conditions). Furthermore, observers could judge the motion direction even when the stimulus consisted of a combination of second-order motion and pure first-order motion generated by a four-stroke motion illusion. These results from the direction judgment task consistently indicate that second-order motion signals are pooled across space and orientation. The pooling includes first-order and different types of second-order signals. The present results are consistent with Stoner and Albright (1992), who reported motion integration with a plaid stimulus whose two components were defined by luminance and flicker contrast, and support the notion of non-selective pooling as assumed in the model proposed by Wilson et al. (1992).

The rating scores obtained with stimuli containing second-order motions were generally lower than those obtained with a stimulus consisting of only luminance elements. The low rating scores suggest that the pooling of second-order motion is not very efficient, which is in agreement with observations by Victor and Conte (1992). The direction-discrimination task could elucidate the relatively inefficient integration of second-order signals presumably because observers could judge the rotation direction even when the integration was partial and the global pattern was not rigid. In contrast, the rating task would be more complicated since observers scored the extent of the integration synthetically on the basis of several aspects. Naturally, results could be vulnerable to the rigidity of global figure, the quality of local motions, and other various factors. Thus, the discrepancy between results from the two tasks can be explained by the difference in the task sensitivity and by other contributing factors.

The present findings suggest that there is a universal global motion process that pools first- and second-order motion signals, while the efficiency is higher for pooling within first-order motion signals than for pooling including second-order motion signals. We can suggest two possible hypotheses to account for this apparent discrepancy.

One hypothesis is that, in addition to the universal pooling system, there is a pooling system specific to first-order signals. Assuming that the first-order pooling system is more efficient than the universal pooling system, this hypothesis fits well with the present findings. In somewhat agreement with this idea, independent global motion pooling of 2D local motions for first- and second-order motions has been suggested by noise masking studies (Cassanello, Edwards, Badcock, & Nishida, 2009; Edwards & Badcock, 1995).

The second hypothesis seeks the difference in the efficiency of motion pooling between first-order and second-order motions not in the processing mechanism but in the stimulus structure. Pure [drift-balanced (Chubb & Sperling, 1988)] second-order motions cannot be perfectly rigid and coherent. In carriers of second-order motion, they contain first-order motion signals that are directionally balanced as a whole. Almost all of these first-order motion signals are inconsistent either with the local 1D second-order motion or the global 2D motion. Due to these “incoherent” components, the quality of coherent global motion may be degraded for a stimulus display containing second-order motions. In contrast, first-order motion does not have to include such incoherent components. Although luminance-defined motion stimuli can be accompanied by incoherent random noise, as were our stimuli, there is a marked difference in phenomenology of the motion perception. For first-order motions, the motion and noise components are properly decomposed as belonging to separate layers in additive transparency. On the other hand, for second-order motions, second-order motion components and first-order noise components are perceptually inseparable, presumably because their relationship is not additive but multiplicative. Here we assume the importance of perceptual segmentation of relevant motion and irrelevant noise components in global motion pooling. This assumption is based on previous findings that the rules of perceptual grouping for static images can significantly modulate the magnitude of global motion integration (Lorenceau & Alais, 2001; Lorenceau & Zago, 1999; McDermott, Weiss, & Adelson, 2001; Stoner & Albright, 1992). See also Appendix C for an argument relevant to the importance of figure–ground segregation in 1D motion integration. Note that our global motion stimuli had a diamond configuration that had been shown to be best configuration for motion integration (Lorenceau & Alais, 2001) and this could be a critical factor to obtain pooling across first-order and second-order signals (cf., Cassanello et al., 2009).

Another factor that could impair motion integration across first-order and second-order stimuli is the apparent speed of local motion. The apparent speed depends on stimulus attribute and intensity (e.g., Gegenfurtner & Hawken, 1996). While we used a common physical speed, the perceptual speed was not exactly the same across different motion types. The resulting disagreement in local motion speed could impair motion integration into a rigid global motion.
The present findings are consistent both with the notion of multiple pooling mechanisms, including the one specific to first-order motion signals (hypothesis 1), and with the notion of a single universal pooling mechanism under control of perceptual organization processing (hypothesis 2). It is worth noting that we drew a similar conclusion in our recent study on the effects of spatial frequency on 1D motion pooling (Maruya et al., 2010). To account for the spatial-frequency selectivity and non-selectively (Amano, Edwards, Badcock, & Nishida, 2009; Maruya et al., 2010), each consistently obtained under different conditions, we conjecture that there are both spatial-frequency-selective and non-selective pooling mechanisms, or that there is a single universal pooling mechanism to which spatial frequency contributes as an image segmentation cue. Although there remains an ambiguity in conclusions about the underlying process, the present study and the studies cited above agree in that they both indicate the presence of a universal motion pooling mechanism in the brain.

This study also highlights a novel aspect on the functional role of second-order motion. The present results indicate that second-order motions are pooled across space and could play an effective role in the solution of the aperture problem. In contrast, several studies have suggested that second-order signals are not as effective as first-order signals in a number of motion tasks requiring comparisons of the spatial distribution of motion signals. Surface segmentation does not occur with second-order signals (Cavanagh & Mather, 1989). Three-dimensional structures are not extracted from a second-order motion field (Dosher, Landy, & Sperling, 1989; Landy, Dosher, Sperling, & Perkins, 1991). In addition, simultaneous motion contrast is not generated by second-order signals (Nishida, Edwards, & Sato, 1997). What is the difference between the cases where second-order motion is effective and where it is not? An intriguing point is that all of these results reporting the ineffectiveness of second-order signals were obtained with tasks requiring differentiation of motion strength across space. The second-order signals are generally noisy and the spatial resolution of their detection is relatively low (Smith, Hess, & Baker, 1994). The results of spatial differentiation based on such blurred signals are not reliable. This would lead to the lower effectiveness of second-order signals in these tasks. In contrast, motion integration is a situation where estimation becomes more accurate by using more signals even if they are noisy. Second-order signals could be effective in such situations. This view is rather speculative at this point; however, it would work as a working hypothesis for further discussion of the functional role of second-order motion signals.

The present results do not suggest specific physiological mechanisms involved in universal motion integration. The universal pooling system indicated by this study would be consistent with a report that some types of MT cells have the same direction tuning for motion defined by luminance and temporal and spatial textures (Albright, 1992). Recently, however, Majaj, Carandini, and Movshon (2007) reported that the pattern of direction selectivity with multiple aperture stimuli is based on component motion direction, not pattern motion direction, for cells in the Macaque MT (V5) area. This suggests that this type of MT cell does not support the motion pooling across space. Thus, the situation is still controversial, and the question of whether there is a physiological mechanism involved in the universal motion pooling remains open.

In conclusion, the results of the present study indicate the presence of a universal motion integration system that pools 1D motion signals across space irrespective of the attributes defining the motion. Although specification of the detailed structure of the integration mechanism and its neural implementation awaits future investigation, the present results demonstrate an important functional role played by second-order motion in the estimation of global motion.

Appendix A

Measurement of local motion visibility for each stimulus type

In this experiment, we measured the thresholds of direction judgment of local motion for three observers. The stimulus used for the measurement of threshold was the same as the L-, F-, Cs-, Cd-, or D-type stimulus used in the main experiment, except that four vertical bars made a global horizontal translation (leftward or rightward) for 500 ms (or 750 ms for D-type) at a constant speed (8.0 deg/s for L-, F-, Cs-, or Cd-type, and 7.2 deg/s for D-type). To continuously change the visibility of

<table>
<thead>
<tr>
<th>Element type</th>
<th>Observer</th>
<th>Percent signal at threshold level</th>
<th>Visibility for 100% signal (×threshold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>SN</td>
<td>3.73</td>
<td>26.8</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>3.37</td>
<td>29.7</td>
</tr>
<tr>
<td>F</td>
<td>KM</td>
<td>19.5</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td>SN</td>
<td>19.5</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>23.1</td>
<td>4.33</td>
</tr>
<tr>
<td>Cs</td>
<td>KM</td>
<td>13.1</td>
<td>7.63</td>
</tr>
<tr>
<td></td>
<td>SN</td>
<td>16.8</td>
<td>5.95</td>
</tr>
<tr>
<td>Cd</td>
<td>KM</td>
<td>18.6</td>
<td>5.38</td>
</tr>
<tr>
<td></td>
<td>SN</td>
<td>27.2</td>
<td>3.68</td>
</tr>
<tr>
<td>D</td>
<td>KM</td>
<td>37.2</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td>SN</td>
<td>78</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>71.7</td>
<td>1.39</td>
</tr>
</tbody>
</table>

Table A1. Direction judgment thresholds and estimated visibilities for 100% signal stimulus.
moving bars, we varied the rate of signal dots in the moving bar. For example, when the signal ratio was 25%, a quarter of the dots in the target area had the same pixel statistics as the target in the original stimulus (e.g., higher mean luminance for L-type, 100-Hz refresh for F-type), while the remaining 75% dots had the same pixel statistics as the surrounding area (e.g., lower mean luminance for L-type, 0-Hz refresh for F-type). The Observers’ task was to judge the direction of horizontal local motion. The signal ratio was fixed at one of seven values within a block, which consisted of 20 trials. The range of steps was adjusted for each observer. Observers performed at least two sessions for each of the signal ratio (7 levels) and stimulus type conditions in a randomized order. The threshold was estimated by fitting a logistic function to the obtained psychometric function.

Table A1 shows the estimated thresholds for three observers. The threshold signal ratio was the lowest for L-type, moderate for F-, Cs-, and Cd-types, and the highest for D-type. When the visibility of the original 100% signal stimulus was evaluated in terms of threshold multiples, Table A1 indicates that L-type had the highest visibility, while D-type had the lowest visibility.

**Appendix B**

**Comparison of motion integration performance at equal visibilities**

In this experiment, we examined the extent of influence of these differences in visibility by measuring the performance of direction judgment and quality rating of global motion for the L, F, Cs, and Cd conditions at several visibility levels. The stimulus used in this experiment was almost the same as in the main experiments except that the visibility of moving bars was varied in the way described in Appendix A. The signal ratio of each element was fixed at one of 5 levels within a block, which consisted of 20 trials. Observers ran at least two sessions for direction judgment and one session for rating at each signal ratio in a randomized order.

The results (Figure B1) showed that the direction discrimination was better for the L condition than for the other conditions even when compared at equal visibilities. A similar conclusion can be drawn from the rating results, except for high ratings for the F condition by KM.

![Figure B1](https://jov.arvojournals.org/)
Appendix C

Influence of depth structure on global motion integration

In Experiment 2, we used a depth arrangement in which moving bars were located on the fixation plane and the surrounds were on a far plane. In this experiment, we examined the effects of the depth arrangement on the performance of direction judgment for the global circular translation.

Methods

The apparatus and stimulus were the same as used in Experiment 2 except for the following points. The stimulus had six types of depth arrangement including the same configuration as used in Experiment 2 (Figure C1a). In three of the six arrangements (“Bar” types), the disparity was fixed at zero for the surround noise and varied for the bars as follows: all four bars had a near disparity (Bar++); all four bars had a far disparity (Bar−−); or two had a near and two had a far disparity (Bar+-). In the remaining three arrangements, moving bars were on the zero-disparity plane. The relative depth order was varied in the same manner as in the first three conditions: all of the four surrounds had a near disparity (BG++); four surrounds had a far disparity (BG−−, the same condition as used in Experiment 2); or two had a near and two had a far disparity (BG+-). Note that the depth orders between the bar and the surround were the same between Bar++ and BG−−, between Bar−− and BG++, and between Bar+- and BG+- conditions. The disparity size and motion parameters were the same as used in Experiment 2.

Figure C1. Influence of depth structure on global motion integration. (a) Diagram of stimulus arrangement for the six stimulus types. (b) Results. The dotted lines show chance levels and 75% and 100% performance levels. Error bars show 95% confidence intervals.
In each session, one of six depth arrangements was repeatedly displayed for 20 times. Observers judged the direction of circular translation after each stimulus presentation. They performed two sessions for each stimulus condition. In total, they performed 12 sessions (240 trials). Eight observers including the two authors participated in this experiment. All had normal or corrected-to-normal vision.

Results and discussions

These results shown in Figure C1b indicate that the motion pooling is influenced by the depth structure of the stimulus. A one-way ANOVA with repeated measures indicated a significant main effect of stimulus type \((F(5,35) = 10.61, p < 0.01)\). A post-hoc LSD test revealed that performance was significantly higher for Bar++ and BG−− conditions than for the other conditions \((p < 0.05)\). No significant difference was found within the first two or within the latter four conditions. These results indicate that the performance was similar when the depth order between the bars and the surrounds was the same, and that good motion pooling was obtained only when the moving bars were seen in front of the background (Bar++, BG−− type).

The second point suggests the importance of figure–ground segregation in 1D motion integration. That is, motion integration might primarily occur for moving elements seen as figures but not for those seen as backgrounds. In agreement with this idea, we informally observed that even for F-type stimuli, the direction-discrimination performance dramatically dropped when in each element a moving bar with static texture was drawn on the dynamic noise background. Although a detailed examination of the effects of figure–ground segregation is beyond the scope of this study, it is worthwhile to note that motion integration with complex motion stimuli, including many second-order motions, could be affected by this factor, as well as by the attributes that define local motions.

Acknowledgments

We would like to thank Alan Johnston, Mark Edwards, Kaoru Amano, and David Badcock for their helpful comments.

Commercial relationships: none.
Corresponding author: Kazushi Maruya.
Email: kazushi.maruya@gmail.com.
Address: 3-1 Morinosato Wakamiya Atsugi, Kanagawa 243-0198, Japan.

References


