A method for generating a “purely first-order” dichoptic motion stimulus

Ryusuke Hayashi
NTT Communication Science Laboratories, Atsugi, Japan

Shin’ya Nishida
Max Planck Institute for Biological Cybernetics, Tuebingen, Germany

Andreas Tolias
Max Planck Institute for Biological Cybernetics, Tuebingen, Germany

Nikos K. Logothetis
Max Planck Institute for Biological Cybernetics, Tuebingen, Germany

In the present technical article, we describe a method for generating a new dichoptic motion stimulus, the monocular components of which are dynamic random noise without constant figural cues for feature-tracking-based motion. Our dichoptic motion stimulus adds a new line of evidence, which supports the original conclusion of M. Shadlen and T. Carney (1986) that motion detection can be solely derived from early binocular motion processing. Further, we describe novel motion displays in which monocular motion and binocular motion are in opposite directions with variable intensity ratios. Our dichoptic stimuli will serve as a useful tool to investigate the interaction between low-level binocular motion detectors and monocular motion detectors without requiring feature extraction before motion detection.

Keywords: motion perception, binocular vision, dichoptic, stereo, interocular phase


Introduction

Several lines of evidence have revealed that human visual motion perception is mediated by at least two separate motion systems: short-range versus long-range motion (Braddick, 1974) or first-order versus second-order motion (Cavanagh & Mather, 1989—note that a third-order mechanism has also been proposed; Lu & Sperling, 2001). The short-range or first-order motion systems are early motion processes, extracting motion from moving luminance modulation, which can be computationally described as Reichardt detectors (Reichardt, 1961) or the motion-energy model (Adelson & Bergen, 1985). The early motion-energy-based system had been considered as an exclusively monocular process because motion of random-dot patterns (Braddick, 1974) was not seen when successive images were presented alternately to the left and right eyes. On the other hand, the long-range or higher order motion system (third-order motion in the terminology of Lu & Sperling, 2001) detects movement of salient features or shapes and is considered to be a binocular process, which means that the motion systems can combine features across eyes for motion perception (Lu & Sperling, 1995; Pantle & Picciano, 1976; Shipley, Kennedy, & King, 1945).

Contrary to the classical studies, Shadlen and Carney (1986) found a dichoptic motion illusion and interpreted the perception of motion of their stimulus as the results of early level motion-energy-based mechanisms after binocular integration (Carney, 1997; Carney & Shadlen, 1992).

Shadlen and Carney developed dichoptic motion stimuli by decomposing a moving sinusoidal grating into the sum of two standing waves in spatiotemporal quadrature as described in Equation 1:

\[
L(x, t) = C + m \cos(\omega_x - \phi) \cos(\omega_t), \\
R(x, t) = C + m \sin(\omega_x) \sin(\omega_t), \\
L(x, t) + R(x, t) = 2C + m \cos(\omega_x - \phi),
\]

where \(L(x, t)\) and \(R(x, t)\) are the luminance profiles of the left and right eye’s inputs at position \(x\) and time \(t\).

doi: 10.1167/7.8.7
\( \omega_x, \ \omega_t, \ C, \) and \( m \) are spatial frequency, temporal frequency, mean luminance, and amplitude/contrast, respectively.

In this stimulus, a sinusoidal grating presented to one eye was modulated by counterphase flicker at a fixed temporal frequency (which is represented as a product of spatial and temporal sinusoidal functions), whereas a similar grating was presented to the other eye but shifted by 90° (one-quarter cycle) in spatial and temporal phase (which is represented by substituting cosine for sine). Although each monocular component alone appears as a flickering grating without any directional information, the superposition of the two patterns is perceived as a smoothly drifting grating.

Moreover, Shadlen and Carney generalized the quadrature relationship exploited in Equation 1 to generate random texture motion displays in which an input pattern to one eye was replaced by an arbitrary spatial pattern expressed as a Fourier series or summation of many spatial frequencies with random amplitude and phase. A pattern in quadrature to the other eye was created by shifting the phase of “all spatial frequency components” by 90° or applying the Hilbert transformation (Carney & Shadlen, 1993). The patterns presented to the two eyes were temporally modulated at the single temporal frequency, one in the sine phase and the other in the cosine phase. The sum of the left and right eyes’ inputs is a superposition of traveling waves of common temporal frequency and direction.

However, it has been speculated that these dichoptic motion stimuli may be detected by higher level feature-tracking mechanisms (Georgeson & Shackleton, 1989, 1992; Lu & Sperling, 1995).

Here, we extend Shadlen and Carney’s stimulus configuration and propose a method for generating new dichoptic motion stimuli. The first point of our elaboration was the use of the Hilbert transformation to shift all frequency components by 90° not only in the space domain but also in the time domain. This enables us to employ a variety of spatiotemporal pattern as monocular inputs, including dynamic random noise that is expected to make feature tracking more difficult than in the previous stimulus (Carney & Shadlen, 1993). Together with the pedestal immunity of a dichoptic motion (Carney, 1997), perception of dichoptic motion with our new stimuli adds a new line of convincing evidence for the hypothesis that motion perception can be derived from binocular early motion mechanisms (Shadlen & Carney, 1986). In addition, we will show how to generate motion stimuli where the directions of monocular motion and binocular motion are opposite or different from each other. This stimulus could be served as a useful tool to analyze interactions of monocular and binocular motion mechanisms. Observations with our new motion stimuli indicate clear dissociations of binocular motion-energy-based mechanisms from monocular motion-energy-based mechanisms or from higher level feature-tracking mechanisms.

**Method**

We propose to apply the Hilbert transformation to shift all frequency components by 90° not only in the space domain but also in the time domain to generate a new dichoptic motion that has no cue for feature-tracking motion as follows.

\[
L(x, t) = I(x, t)
= \sum_{i=0}^{N-1} \sum_{j=0}^{T-1} \sum_{m=0}^{1} \sum_{n=0}^{1} m_{ijmn} \cos(\omega_i x + \phi_m) \cos(\omega_j t + \phi_n),
\]

\[
R(x, t) = H_x(H_t(I(x, t)))
= \sum_{i=0}^{N-1} \sum_{j=0}^{T-1} \sum_{m=0}^{1} \sum_{n=0}^{1} m_{ijmn} \sin(\omega_i x + \phi_m) \sin(\omega_j t + \phi_n),
\]

\[
L(x, t) + R(x, t)
= \sum_{i=0}^{N-1} \sum_{j=0}^{T-1} \sum_{m=0}^{1} \sum_{n=0}^{1} m_{ijmn} \cos(\omega_i x - \omega_j t + \phi_m - \phi_n),
\]

where \( I(x, t) \) is the luminance profile of an arbitrary spatiotemporal random-dot pattern. \( H_{x,t} \) denotes Hilbert transformation in the dimension of \( x \) or \( t \). The right side of the equation is the Fourier series expansion of the left side, where \( \omega_{ij} = 2\pi j/N \) is the spatial frequency, \( N \) is the number of pixels, \( \omega_{ij} = 2\pi j/T \) is the temporal frequency, \( T \) is the number of frames, and \( m_{ijmn} \) is the amplitude of the \((i, j, m, n)\)th component, respectively. The phase, \( \phi_{0/1} \) is 0 or \( \pi/2 \). As can be seen in Equation 2, the Hilbert transformation of dynamic random noise both in space and time domains remains dynamic random noise, whereas the sum of the left and right dynamic noise has motion energy in one direction (at different velocities \( \omega_j/\omega_i \)). We can extend 1D space dynamic random noise to 2D space without breaking the quadrature relationship in one orientation and time. Movie 1 is an example of the dichoptic motion stimuli generated by the method described. The Hilbert transformation was applied to the vertical direction to produce upward motion when fusing a stereo–movie pair. (The observations made on this movie
and on the following movies were confirmed by seven observes who had normal binocular vision.) **Movie 2** demonstrates dichoptic clockwise rotation motion made by applying the Hilbert transformation to the angular direction in the polar coordinate. Here, movements of the pattern of dots in a unique direction under dichoptical presentation are perceived, although the image presented to each eye is dynamic random noise with neither directional information nor trackable features.

Our stimuli have some advantages over the original random texture binocular motion display used by Carney and Shadlen (1993). Firstly, we can synthesize our dichoptic motion stimuli using arbitrary spatiotemporal patterns without limiting the frequency property to a single fixed frequency. Thus, we can stimulate the two eyes continuously instead of alternating presentation (or sinusoidally modulated in time) that might cause unexpected interaction between the two eyes, such as interocular inhibition. Secondly, our stimulus further camouflages monocular motion cues for dichoptic motion perception. In Carney and Shadlen’s stimuli, because each monocular input was modulated by merely single
temporal frequency, its figural features and textual patterns do not change except for contrast modulation. If some figural features become salient and their shape does not change that much after Hilbert transformation by chance, then such features could be interocularly tracked for motion perception. In our stimulus, however, all spatial frequency components are modulated by multiple temporal frequencies, which dynamically shuffle figural features. Therefore, motion perception induced by our stimuli is less likely to derive from feature-tracking mechanisms.

In our dichoptic motion stimulus, unidirectional motion is visible even when the refresh rate of the random pattern is more than 60 Hz (movies demonstrated here are 30 fps throughout). This, however, does not contradict the previous report by Carney and Shadlen (1993) that their dichoptic motion is visible up to 11.3 Hz using random-dot texture (and up to 32 Hz using sinusoidal grating patterns). Whereas Carney and Shadlen’s stimuli are modulated only by a single temporal frequency, our stimulus has a broad range of temporal frequencies. Therefore, our stimulus could activate interocular motion systems even if they can only detect low temporal frequencies. In a different perspective, our stimulus display has advantage to stimulate all early motion systems regardless of their temporal frequency property, which might contribute to enhance motion perception.

Although we can consider that motion perception induced by our dichoptic stimulus is detected by early motion-energy-based detectors, dichoptic motion perception is different from monocular motion perception in several respects, as previous studies using displays with interocular alternation have suggested. First, dichoptic motion is weak compared with motion perception when the quadrature pair is superimposed to one eye (Carney & Shadlen, 1993; Lu & Sperling, 2001). Second, figure/ground segregation is not clearly perceived when the figure is distinguished from the surround only by the opposing direction of dichoptic motion (Carney & Shadlen, 1993). Although motion per se is distinguishable between figure and surround in such stimuli, no striking contour emerges at the boundaries of figure and surround in the monocular motion (see Movie 3 for demonstration). It is worth noting that figure/ground segregation does not occur in so-called second-order motion stimuli such as contrast-modulation motion (Cavanagh & Mather, 1989; Dosher, Landy, & Sperling, 1989) and in second-order stereo images with disparity introduced to the features defined by contrast modulation (Landy, Dosher, Sperling, & Perkins, 1991; Ziegler & Hess, 1999). Therefore, early binocular motion detection seems to result from later processing than monocular motion and to share some properties with second-order (contrast-modulation-sensitive) systems.

As noted above, our method can synthesize dichoptic stimulus from arbitrary spatiotemporal patterns. The monocular pattern can even be a coherently moving stimulus whose direction differs from that of binocular motion. By rendering such monocular motion invisible after binocular fusion, we will show that we can make the direction of binocular motion opposite to that of monocular motion in a single display as follows:

\[
L(x, t) = a f_1(x, t) - H_x f_1(x, t) - (a + 1) H_x f_1(x, t)
\]

\[
X = \frac{1}{a + 1} \left[ \cos(a x + \phi_m) \cos(\alpha x + \phi_m) - \sin(a x + \phi_m) \sin(\alpha x + \phi_m) \right]
\]

where

\[
\cos \alpha = \frac{a}{a^2 + 1}, \quad \sin \alpha = \frac{1}{a^2 + 1},
\]

and \(a(>1)\) is a parameter manipulating the ratio of the amplitudes of monocular motion and binocular motion. In Equation 3, each monocular input contains moving waves in the negative direction (the first term inside the large pair of brackets), whereas the summation of the two is a moving wave in the positive direction. Therefore, binocular motion is perceived in the direction opposite to that of monocular motion when the stimulus is dichoptically observed. Movies 4 and 5 are examples of motion stimuli.
generated by this method. Monocular motion and binocular motion cause motion rivalry where observers usually see only one of two directions at one time. This means binocular motion perception can exist independently of monocular motion perception and provides strong evidence that early binocular motion detectors represent mechanisms separable from monocular motion detectors (Anstis & Duncan, 1983). We can define the intensity of monocular motion $MI$ as the ratio of amplitudes of standing waves (noise) and traveling wave (motion signal), which is a function of parameter $a$.

$$MI = \frac{1}{1 + \sqrt{2(a^2 + 1)}}$$

(5)

In a similar way, we can define the intensity of binocular motion $BI$ as the amplitude of traveling wave after binocular summation normalized by the amplitudes of monocular motion and noise as follows.

$$BI = \frac{\sqrt{2}a}{1 + \sqrt{2(a^2 + 1)}}$$

(6)

Therefore, the intensity of monocular motion is a decreasing function of parameter $a$, whereas the intensity of binocular motion is an increasing function of parameter $a$ in Equation 3. Pure binocular motion stimulus defined by Equation 2 is the limit of Equation 3 when parameter $a$ is infinite.

Another method to make the direction of dichoptic motion different from that of monocular motion is as follows:

$$L(x, y, t) = I(x, y, t) + H_x(H_y(I(x, y, t)))$$

$$= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=0}^{T-1} \sum_{l=0}^{1} \sum_{m=0}^{1} \sum_{n=0}^{1} m_{ijklmn}$$

$$\times \cos(\omega_{x,i} x - \omega_{x,i} t + \varphi_l - \varphi_n) \cos(\omega_{y,j} y + \varphi_m),$$

$$R(x, y, t) = H_y(H_y(I(x, y, t))) - H_x(H_y(I(x, y, t)))$$

$$= - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=0}^{T-1} \sum_{l=0}^{1} \sum_{m=0}^{1} \sum_{n=0}^{1} m_{ijklmn}$$

$$\times \sin(\omega_{x,i} x - \omega_{x,i} t + \varphi_l - \varphi_n) \sin(\omega_{y,j} y + \varphi_m),$$

and

$$L(x, y, t) + R(x, y, t)$$

$$= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=0}^{T-1} \sum_{l=0}^{1} \sum_{m=0}^{1} \sum_{n=0}^{1} m_{ijklmn}$$

$$\times \cos(\omega_{x,i} x + \omega_{y,j} y - \omega_{x,i} t + \varphi_l + \varphi_m - \varphi_n),$$

where $I(x, y, t)$ is the luminance profile of an arbitrary dynamic random noise at 2D space $(x, y)$ and time $t$. As can be seen in Equation 7, each monocular input is moving in the $x$ direction, whereas the summation of the two eyes’ inputs has motion energy in the diagonal
direction, on average. In Movie 6, random-dot patterns for each eye move toward the left, whereas dichoptic motion is in a left–down direction. Again, motion rivalry occurs between monocular and binocular motion.

**Experimental results**

To verify the effect of our new dichoptic motion stimuli, we have performed psychophysical experiments using motion stimuli (described in Equation 3 and demonstrated in Movie 4) in which monocular and binocular motions move in the opposite direction in the vertical orientation. We created six motion stimuli by choosing parameter $a = 1, 1.25, 1.5, 2, 3, \text{ and } 4$, respectively. The motion stimuli were then presented with a duration of 250 ms in pseudorandom order (6 stimuli $\times$ 2 direction $\times$ 30 trials). Left-eye images were output from PC through green signal (8 bits), whereas right-eye images were output through blue signal. Stimuli were rear projected from a CRT projector whose blue gun was replaced by a green gun with a horizontally polarizing filter while the original green gun was covered with a vertically polarizing filter. Images were then segregated by eyeglasses that covered the left eye with a vertically polarizing filter and vice versa. Observers viewed the
image screen (81° × 81°, 512 × 640 pixel resolution, 60 Hz refresh rate) from a distance of 50 cm. Seven healthy observers (of whom six were naive to the purpose of the experiment) whose visual acuity and stereoscopic depth detection performance are normal (or corrected to a normal level) participated in the experiments. All seven observers verbally reported that both monocular and binocular motions were visible but qualitatively different from each other. Binocular motion tends to be perceived stronger at the fixation depth around the central visual area, whereas monocular motion tends to be perceived stronger around the peripheral area at the beginning of stimulus presentation. To direct observer’s attention to binocular motion, we instructed the observers to report the direction of dominant motion (up or down by mouse button pressing, 2AFC) around the central fixation depth at the end of stimulus presentation. We then plotted the response rate, selecting the direction of binocular motion as dominant against motion intensity (binocular motion intensity − monocular motion intensity).

Figure 1 shows the results of the experiment. We can see that observers could report the direction of binocular motion quite accurately if the intensity of binocular motion was high. By reducing the intensity of binocular motion (and increasing the intensity of monocular motion), the binocular motion response decreased monotonically, which verifies that our motion stimuli can also control the intensity ratio of monocular and binocular motion.

**Discussion**

Our dichoptic motion stimuli further camouflage monocular motion cues by containing multiple temporal frequencies as compared with the previous dichoptic motion stimuli (Carney & Shadlen, 1993). The perception of dichoptic motion with our new stimuli supports the existence of interocular first-order energy-based motion mechanisms (Shadlen & Carney, 1986).

Another line of strong evidence showing the interocular first-order motion mechanisms was provided by Carney (1997). It is known that the early level energy-based motion detectors ignore stationary pattern and that they respond linearly when stimuli consist of several sine
waves modulated by different temporal frequencies (van Santen & Sperling, 1984). Thus, adding a stationary sine pattern (the pedestal pattern) to any moving pattern would not change the output of motion-energy detectors while such operation removes all possible feature tracking that would enable direction discrimination. Carney found that observers could perceive dichoptic motion stimuli in the presence of pedestal patterns.

Physiological studies have shown that the neural receptive field has a temporal as well as spatial structure and that both stereoscopic depth and motion are jointly encoded at the level of binocular simple and complex cells in V1 and V2 (Anzai, Ohzawa, & Freeman, 2001). Such a binocular spatiotemporal integration mechanism, which can be described as a motion–stereo hybrid energy model (Qian & Andersen, 1997), is considered to underlie a variety of psychophysical effects such as the Pulfrich-type effect and depth from motion parallax. This view is supported by the fact that binocular disparity (interocular spatial shift) and an interocular time delay are both theoretically (Qian & Andersen, 1997) and experimentally (Anzai et al., 2001; Carney, Paradiso, & Freeman, 1989) indistinguishable. It is worth noting that the stimulus configuration we used for generating dichoptic motion is similar to that of stereograms and Pulfrich-type effects in the sense that all of these binocular stimuli include spatial and/or temporal shifts between the two eyes; dichoptic motion requires a 90° phase shift in both space and time, whereas the Pulfrich-type effect requires an interocular time delay/temporal phase shift and stereoscopic depth requires binocular disparity (space shift; see Table 1). Given that our dichoptic motion stimulus derives from motion energy detection after simple summation of the two eyes have no correlation with each other at each frame, we can see a moving plane at the fixation depth. Therefore, activation of the binocular neurons selective for fixation depth, which are also sensitive to motion energy in the summation of the left- and right-eye images, is more essential for depth plane perception than the correlation of ongoing input images (Hayashi, Miyawaki, Maeda, & Tachi, 2003). This observation is additional evidence of spatiotemporal integration mechanism of stereo and motion perception. In this sense, our dichoptic motion stimuli are also useful to investigate the stereo-depth mechanism as well as motion detection.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Dynamic random-dot stereogram</th>
<th>Pulfrich effect</th>
<th>Morgan and Fahle’s (2000) display</th>
<th>Carney and Shadlen’s (1993) random texture motion display</th>
<th>Our dichoptic motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left:</td>
<td>(l(x,y,t))</td>
<td>(l(x,y,t))</td>
<td>(l(x,y)\cos(\omega t))</td>
<td>(l(x,y)\cos(\omega t + \varphi))</td>
<td>(l(x,y))</td>
</tr>
<tr>
<td>Right:</td>
<td>(l(x + \Delta x,y,t))</td>
<td>(l(x,y,t + \Delta t))</td>
<td>(l(x,y)\cos(\omega t))</td>
<td>(l(x,y)\cos(\omega t + \varphi))</td>
<td>(H_x(\mathcal{H}_y(l(x,y,t))))</td>
</tr>
</tbody>
</table>

Table 1. Summary of stimulus configuration for binocular illusions.

It is still not clear exactly why Braddick (1974) failed to observe motion in his dichoptic presentation. However, one possibility is that alternating presentation of two images may cause interocular suppression, which may disturb the binocular summation of two eyes’ input and reduce dichoptic motion perception.

It is known that depth impression is absent and luster perception is evoked when the image in one eye is an uncorrelated or contrast-reversed image of the other eye’s image (known as an “anticorrelated stereogram”; Cogan, Lomakin, & Rossi, 1993; Cumming & Parker, 1997). In our dichoptic motion stimuli, although the input images to the two eyes have no correlation with each other at each frame, we can see a moving plane at the fixation depth. Therefore, activation of the binocular neurons selective for fixation depth, which are also sensitive to motion energy in the summation of the left- and right-eye images, is more essential for depth plane perception than the correlation of ongoing input images (Hayashi, Miyawaki, Maeda, & Tachi, 2003). This observation is additional evidence of spatiotemporal integration mechanism of stereo and motion perception. In this sense, our dichoptic motion stimuli are also useful to investigate the stereo-depth mechanism as well as motion detection.

Conclusions

Our dichoptic motion stimulus has the following features. (a) Possibility of feature tracking as a source for motion perception is less likely compared with previous stimuli. (b) Its broad band spectral property enables us to stimulate interocular early motion mechanisms regardless of individual detectors’ spatiotemporal frequency limit, which may enhance binocular motion perception. (c) We can manipulate the direction (same or opposite) and the energy of both monocular and binocular motion components independently by extending our
stimulus configuration, although the motion perception in the dichoptic motion stimulus is weak compared with motion perception when identical inputs are superimposed to one eye.

Therefore, our stimulus provides evidence that dichoptic motion can be solely derived from early binocular motion processing and supports Shadlen and Carney’s original conclusion in 1986. Our findings are also very consistent with the more recent position of Lu and Sperling (2001) that there indeed is an interocular first-order motion computation, but it is less sensitive than monocular first-order motion. Our stimuli will be a useful psychophysical tool to investigate low-level binocular motion detectors and their interaction with monocular motion detectors.

Acknowledgments

This work was supported by the Max Planck Institute.

Commercial relationships: none.

Corresponding author: Ryusuke Hayashi.
Email: rhayashi@brain.med.kyoto-u.ac.jp.
Address: Department of Integrative Brain Science, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyoku, Kyoto, 606-8501, Japan.

Footnote

1 The Pulfrich-type effect is seen when dynamic visual noise is viewed with a dark filter over one eye (or interocular delay (Ross, 1974), or interocular temporal phase shift (Morgan & Fahle, 2000). The noise is usually perceived as dots moving in an ellipse around the vertical axis of the stimulus display (the dots in front of the fixation plane appear to move in the direction of the filtered eye, and dots behind the fixation plane appear to move in the opposite direction).

References


