

# Human short-latency ocular vergence responses produced by interocular velocity differences

**B. M. Sheliga**

Laboratory of Sensorimotor Research,  
National Eye Institute, National Institutes of Health,  
Bethesda, MD, USA



**C. Quaia**

Laboratory of Sensorimotor Research,  
National Eye Institute, National Institutes of Health,  
Bethesda, MD, USA



**E. J. FitzGibbon**

Laboratory of Sensorimotor Research,  
National Eye Institute, National Institutes of Health,  
Bethesda, MD, USA



**B. G. Cumming**

Laboratory of Sensorimotor Research,  
National Eye Institute, National Institutes of Health,  
Bethesda, MD, USA



We studied human short-latency vergence eye movements to a novel stimulus that produces interocular velocity differences without a changing disparity signal. Sinusoidal luminance gratings moved in opposite directions (left vs. right; up vs. down) in the two eyes. The grating seen by each eye underwent  $\frac{1}{4}$ -wavelength shifts with each image update. This arrangement eliminated changing disparity cues, since the phase difference between the eyes alternated between  $0^\circ$  and  $180^\circ$ . We nevertheless observed robust short-latency vergence responses (VRs), whose sign was consistent with the interocular velocity differences (IOVDs), indicating that the IOVD cue in isolation can evoke short-latency VRs. The IOVD cue was effective only when the images seen by the two eyes overlapped in space. We observed equally robust VRs for opposite horizontal motions (left in one eye, right in the other) and opposite vertical motions (up in one eye, down in the other). Whereas the former are naturally generated by objects moving in depth, the latter are not part of our normal experience. To our knowledge, this is the first demonstration of a behavioral consequence of vertical IOVD. This may reflect the fact that some neurons in area MT are sensitive to these motion signals (Czuba, Huk, Cormack, & Kohn, 2014). VRs were the strongest for spatial frequencies in the range of  $0.35\text{--}1\text{ c}/^\circ$ , much higher than the optimal spatial frequencies for evoking ocular-following responses observed during

frontoparallel motion. This suggests that the two motion signals are detected by different neuronal populations. We also produced IOVD using moving uncorrelated one-dimensional white-noise stimuli. In this case the most effective stimuli have low speed, as predicted if the drive originates in neurons tuned to high spatial frequencies (Sheliga, Quaia, FitzGibbon, & Cumming, 2016).

## Introduction

Two binocular cues signal motion of an object toward or away from a subject (Rashbass & Westheimer, 1961). First, there are changes in the object's binocular disparity (CD cue). Second, images of the object on the two retinas move with different velocities: an interocular velocity difference (IOVD cue). Considerable progress has been made in elucidating the relative importance of each cue in perceptual judgments of motion in depth: It has been found to vary as a function of task requirements (e.g., detection vs. discrimination) and to be sensitive to the experimental design (e.g., the disparity pedestal or stimulus visibility; see, e.g., Brooks, 2002; Brooks & Stone, 2004; Cumming & Parker, 1994; Czuba, Rokers, Huk, & Cormack, 2010; Gray & Regan, 1996; Harris, Nefs, &

Citation: Sheliga, B. M., Quaia, C., FitzGibbon, E. J., & Cumming, B. G. (2016). Human short-latency ocular vergence responses produced by interocular velocity differences. *Journal of Vision*, 16(10):11, 1–13, doi:10.1167/16.10.11.



Grafton, 2008, for a review; Harris & Watamaniuk, 1995; Portfors-Yeomans & Regan, 1996; Tyler, 1971). Furthermore, large differences across subjects are commonly reported (Allen, Haun, Hanley, Green, & Rokers, 2015).

Visual stimuli moving in depth are also capable of inducing eye movements. Early work revealed that moving the whole visual field slowly in opposite directions in the two eyes evoked robust vergence eye movements but no sensation of motion in depth (Erkelens & Collewijn, 1985a, 1985b). This implied that eye movements were guided by changes in *absolute* retinal disparity, whereas a motion-in-depth percept depended crucially upon changes in *relative* disparities or motion of objects comprising the scene (i.e., disparity or motion between foreground and background; Erkelens & Collewijn, 1985a, 1985b; Regan, Erkelens, & Collewijn, 1986). Masson, Yang, and Miles (2002) studied the speed tuning of short-latency disparity vergence responses (DVRs) produced by a random-dot stereogram (RDS) moving in opposite directions in the two eyes. They compared a “symmetric” stimulus, where both eyes see motion, with an “asymmetric” one in which one eye sees a stationary pattern. For a given rate of disparity change, monocular velocities in the symmetric condition are half those in the asymmetric condition. The researchers found that tuning was invariant in the two conditions when expressed as a function of monocular speed but not when expressed in terms of disparity change. This suggested a dominant role for IOVD over CD, but the relative contributions of CD and IOVD cues were not systematically studied.

To study IOVD in isolation, in Experiment 1 we presented drifting sinusoidal gratings dichoptically in such a way that they generated no CD cues. We found that the IOVD cue alone is sufficient for evoking short-latency VRs. Importantly, we show that the vergence system responds to both horizontal and vertical IOVDs. Since only horizontal IOVDs give rise to a clear percept (motion in depth), all previous behavioral studies neglected vertical IOVDs. However, a recent study of IOVD coding in primate area MT (Czuba, Huk, Cormack, & Kohn, 2014) indicates that many neurons are responsive to vertical IOVD. Our observation of vertical VRs is the first demonstration of any behavioral consequence of vertical IOVD, perhaps a behavioral signature of these neuronal signals. In Experiment 2 we used uncorrelated random-line stereograms, a stimulus similar to that used previously by others to isolate the IOVD cue, and investigated the speed tuning of the mechanism. Finally, in Experiment 3 we demonstrated that these responses do depend on local binocular comparisons—binocular overlap of monocular images is necessary.

## Experiment 1: Sine-wave gratings

### Materials and methods

Many of the techniques will be described only briefly, since they are similar to those used in this laboratory in the past (e.g., Rambold, Sheliga, & Miles, 2010). Experimental protocols were approved by the institutional review committee concerned with the use of human subjects. Our research was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), and informed consent was obtained for experimentation with human subjects.

### Subjects

Three subjects took part in this study: Two were authors (BMS and CQ) and the third was a paid volunteer (WG) who was unaware of the purpose of the experiments. All subjects had normal or corrected-to-normal vision.

### Eye-movement recording

The horizontal and vertical positions of both eyes were recorded with an electromagnetic induction technique (Robinson, 1963). A scleral search coil was embedded in a silastin ring (Collewijn, Van Der Mark, & Jansen, 1975), as described by Yang, FitzGibbon, and Miles (2003). At the beginning of each recording session, a quantitative calibration procedure was performed for the coil in each eye using monocularly viewed fixation targets.

### Visual display and stimuli

Dichoptic stimuli were presented using a Wheatstone mirror stereoscope. In a darkened room, each eye saw a computer monitor (Sony GDM-series 21-in. CRT) through a 45° mirror, creating a binocular image straight ahead at a distance of 521 mm from the eyes' corneal vertices, which was also the optical distance to the images on the two monitor screens. Each monitor was driven by an independent PC (Dell Precision 380), but the outputs of each computer's video card (PC NVIDIA Quadro FX 5600) were frame-locked via NVIDIA Quadro G-Sync cards. This arrangement allowed the presentation of independent images simultaneously to each eye. The monitor screens were each 41.8° wide and 32.0° high, had 1024 × 768 pixel resolution (i.e., 23.4 pixels/° directly ahead of each eye), and were synchronously refreshed at a rate of 150 Hz. Each monitor was driven via an attenuator (Pelli, 1997) and a video signal splitter (Black Box Corp., AC085A-

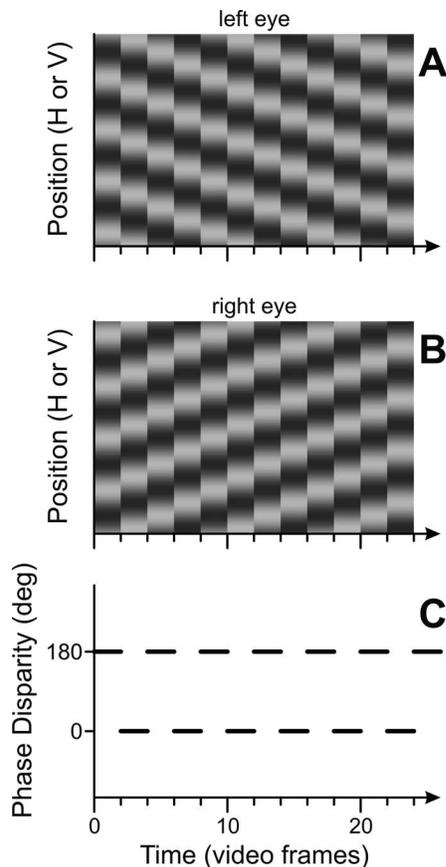


Figure 1. Visual stimulus. The sine wave seen by each eye—left (A) and right (B)—underwent a  $\frac{1}{4}$ -wavelength shift every other video frame. As a result, the phase difference between sine waves seen by each eye (C; phase disparity) changed by  $180^\circ$  every second frame. The interocular phase difference was always either  $0^\circ$  or  $180^\circ$ .

R2), allowing presentation of black-and-white images with 11-bit grayscale resolution (mean luminance of  $20.8 \text{ cd/m}^2$ ).

One-dimensional (1-D) horizontal or vertical sine-wave gratings moved in opposite directions (up vs. down; left vs. right) in the two eyes. They were seen through a central  $32^\circ \times 32^\circ$  rectangular aperture. The sine wave seen by each eye underwent a  $\frac{1}{4}$ -wavelength shift every second video frame (i.e., at a 75-Hz stimulus refresh rate; Figure 1A, B). This arrangement provided a near-optimal motion temporal frequency (18.75 Hz; Gellman, Carl, & Miles, 1990; Sheliga, Quaia, FitzGibbon, & Cumming, 2016) and eliminated CD cues, since the phase difference between sine waves seen by each eye changed by  $180^\circ$  every second frame. The interocular phase difference was always either  $0^\circ$  or  $180^\circ$  (Figure 1C). Since a  $180^\circ$  interocular phase difference is by definition ambiguous, there was no CD cue. All sine-wave gratings had 32% Michelson contrast, and their spatial frequencies (SFs) varied from 0.0625 to  $1 \text{ c/}^\circ$  in octave increments. A single block of

trials had 20 randomly interleaved stimuli: 5 SFs  $\times$  2 axes of motion (horizontal vs. vertical)  $\times$  2 directions (for a given eye: leftward/downward vs. rightward/upward).

### Procedures

The experimental paradigms were controlled by three PCs, which communicated via Ethernet using a TCP/IP protocol. One of the PCs was running the Real-time EXperimentation software (REX; Hays, Richmond, & Optican, 1982) and provided overall control of the experiment as well as acquiring, displaying, and storing the eye-movement data. Two other PCs utilized the Psychophysics Toolbox extensions of MATLAB (Brainard, 1997; Pelli, 1997) and generated the binocular visual stimuli upon receiving a start signal from the REX machine.

At the start of each trial, a stationary sine-wave stimulus (the same SF in each eye, randomly selected from a lookup table) and a central fixation cross (width =  $10^\circ$ , height =  $2^\circ$ , thickness =  $0.2^\circ$ ) appeared on both monitors. The phase difference between sine waves seen by each eye was  $0^\circ$  (zero disparity)—i.e., the binocular image was located at the plane of fixation. After the subject's eyes had been positioned within  $2^\circ$  of the centers of the crosses for 500–1000 ms and no saccades had been detected (using an eye velocity threshold of  $18^\circ/\text{s}$ ), the contrast of the sine wave seen by one of the eyes (randomly chosen) was flipped (analogous to  $180^\circ$  phase shift). Therefore, the binocular image was not located at the plane of fixation anymore and instead had an ambiguous disparity (horizontal or vertical). After another 100 ms, the fixation crosses disappeared and motion commenced (opposite directions in the two eyes), after which the screen changed to uniform gray, marking the end of the trial. We adopted this procedure because pilot experiments revealed that the change in phase alone was sufficient to produce transient (but idiosyncratic) vergence responses. We therefore introduced this 100-ms delay before motion onset to ensure that this did not contaminate our measures of vergence produced by the IOVD itself. The magnitude of the effect was minor in two of our subjects but not in the third. In all subjects, any transient ocular response caused by the contrast flip faded within 100 ms—i.e., it was completed before the moment of motion onset. After an intertrial interval of 500 ms the fixation crosses and new stationary sine-wave stimuli reappeared, signaling the start of another trial. The subjects were asked to refrain from blinking or shifting fixation except during the intertrial intervals but were given no instructions relating to the motion stimuli. If no saccades were detected for the duration of the trial, then the data were stored; otherwise, the trial was aborted and repeated within the same block. Data were

collected over several sessions until each condition had been repeated an adequate number of times to permit good resolution of the responses (through averaging).

### Data analysis

The horizontal and vertical eye-position data obtained for the right and left eyes during the calibration procedure were each fitted with second-order polynomials whose parameters were used to linearize the corresponding eye-position data collected during the experiment proper. The eye-position data were then smoothed with an acausal sixth-order Butterworth filter (3 dB at 30 Hz), and mean temporal profiles were computed for each stimulus condition. Trials with small saccadic intrusions that failed to reach the eye-velocity criterion of 18°/s used during the experiment were deleted. The horizontal (vertical) vergence angle was computed by subtracting the horizontal (vertical) position of the right eye from that of the left eye. We used the convention that rightward and upward motion of stimuli (and the eyes) was positive. Convergence and left-sursumvergence, therefore, had positive signs. To improve the signal-to-noise ratio, the mean VR profile for each condition in which the left eye saw leftward (downward) motion and the right eye saw rightward (upward) motion was subtracted from the mean VR profile for the corresponding (same sine-wave SF) condition in which the left eye saw rightward (upward) motion and the right eye saw leftward (downward) motion. This produced mean pooled-difference vergence measures,

$$VR = VR_{Lr/u+Rl/d} - VR_{Ll/d+Rr/u}, \quad (1)$$

where the first (uppercase) subscript letter denotes the eye which sees the image (left or right) and the lowercase subscript letters denote the direction of motion seen by each eye (right or left, up or down). As convergence and left-sursumvergence were positive in our sign convention, the horizontal (vertical) pooled-difference measures were positive when in the direction dictated by the motion stimuli. Velocity responses (mean vergence velocity) were estimated from differences between samples 10 ms apart (central difference method) and evaluated every 1 ms. The horizontal (vertical) version response was computed by adding the horizontal (vertical) position of the right eye to that of the left eye and halving the result of the sum. Mean pooled-difference version measures were calculated as the difference in mean version-response profiles between the same pairs of conditions that were used to calculate pooled-difference vergence measures. Response latencies were estimated by determining the time after motion onset when the mean vergence velocity first exceeded 0.1°/s. Responses of this magnitude are almost always significant (the standard error of the

mean [*SEM*] of velocities in the period 0–50 ms ranged from 0.001°/s to 0.004°/s)—applying an absolute criterion avoids introducing any changes in estimated latency with statistical power. The amplitude of a VR to a given stimulus was calculated by measuring the changes in the mean pooled-difference vergence measures over the initial open-loop period—i.e., over the period up to twice the minimum response latency. However, to permit within-subject comparisons across different paradigms and experiments, the duration of this measurement window for a given subject was always the same throughout the entire study (71 ms for subject BMS; 69 ms for CQ; 79 ms for WG); and for all of the data obtained from a given subject with a given stimulus set, this window always commenced at the same time after the stimulus onset (stimulus-locked measures), the actual time being determined by the shortest response latency in the particular stimulus set.<sup>1</sup> All error bars in the figures are  $\pm 1$  *SEM*; actually they were smaller than a symbol size in many cases, and therefore not visible on the graphs).

## Results

Sine-wave gratings moving in opposite directions in the two eyes elicited robust VRs in both horizontal and vertical directions. In all subjects the horizontal (vertical) stimulus motion caused almost exclusively horizontal (vertical) VRs. VRs in the direction orthogonal to the axis of stimulus motion were negligible (median was 3.0% of the response in the stimulus direction; see also Figure 2B through D) and will not be discussed further. Figure 2A shows mean horizontal vergence velocity profiles over time obtained from subject WG in response to sine-wave gratings moving horizontally: Sine-wave SF is noted by grayscale coding of velocity traces. With low SFs, the responses were the weakest; they peaked at intermediate SFs, before slightly falling again for the highest SF tested. Response latencies were short (80–90 ms) and showed only a modest dependence on the SF of the stimulus. Figure 2B through D quantifies observations from the three subjects. We estimated the peak by fitting a Gaussian function of log SF (median optimal SF = 0.525 c/°; range = 0.35–1.0 c/°; median  $r^2 = 0.969$ ; range = 0.861–0.995; see Table A1 for the full list of fit parameters), though the horizontal VRs of subject BMS and vertical VRs of subject CQ peaked for the highest SF tested (1 c/°), and their dependencies on SF were quite linear on a semilog scale. Overall, the horizontal VRs were stronger than the vertical ones in subjects CQ and WG, but the opposite was true for subject BMS (intersubject variability of this kind can also be seen in Mulligan, Stevenson, & Cormack, 2013).

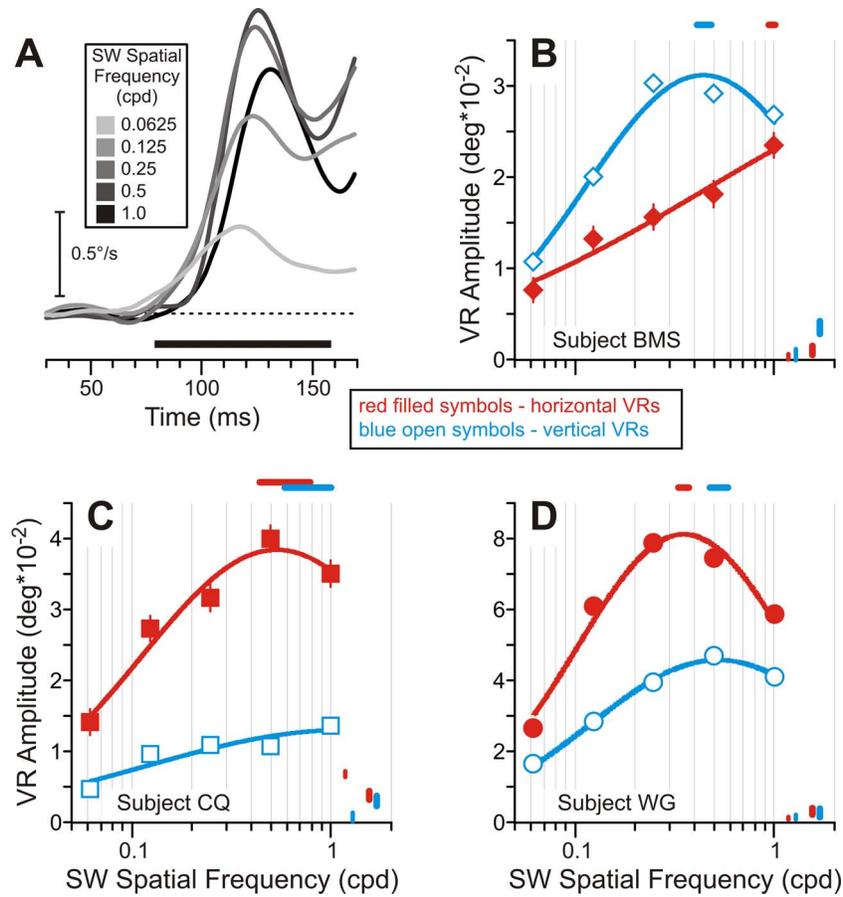


Figure 2. Experiment 1. (A) Mean horizontal vergence velocity profiles over time to sinusoidal luminance gratings moving horizontally in opposite directions in the two eyes. Spatial frequency is denoted by grayscale coding of velocity traces; see the inset. Subject WG: Each trace is the mean response to 142–161 stimulus repetitions. Abscissa shows the time from the stimulus onset; horizontal dotted line marks zero velocity; horizontal thick black line beneath the traces indicates the response measurement window. (B–D) Semilog plots of VR amplitude against grating SF. Red filled symbols: horizontal VRs; blue open symbols: vertical VRs. (B) Subject BMS: diamonds, 274–292 trials per condition; (C) subject CQ: squares, 128–255 trials per condition; (D) subject WG: circles, 142–161 trials per condition. Color-coded horizontal lines above each graph show the 95% confidence intervals of the fits' peak locations. Color-coded vertical lines in the right lower corner of each graph show the mean amplitude ( $\pm$  SEM) of VRs orthogonal to the axis of stimulus motion (thinner lines) and version responses along the axis of stimulus motion (thicker lines). Since the calculation of these measures was done in exactly the same way as described in Materials and methods for the VRs along the axis of motion, the sign of version (and orthogonal vergence) responses is arbitrary. We therefore plot absolute values.

Although this experiment used sine-wave gratings moving with the same speed but in opposite directions in the two eyes, it is possible that an ocular imbalance in the effect on version eye movements could lead to version responses. In fact, version responses were small (median was 10.7% of the VRs; see also Figure 2B through D).<sup>2</sup>

## Discussion

The results of Experiment 1 clearly show that the IOVD cue alone is sufficient for evoking short-latency VRs. Here the sine wave seen by each eye during successive video frames underwent  $\frac{1}{4}$ -wavelength shifts. This eliminates the CD cue, because the phase difference

between the eyes alternated between 0° and 180°. This is a novel visual stimulus that, to our knowledge, has not been used before in the study of motion in depth. Several previous studies have used uncorrelated RDSs moving in opposite directions to deliver IOVD in isolation (e.g., Brooks, 2002; Sanada & DeAngelis, 2014; Shioiri, Saisho, & Yaguchi, 2000). One problem with such stimuli is that while they have no mean correlation, it is quite possible that local patches of any one image contain “false matches” that approximate a stimulus with a certain disparity. The stimulus motion then ensures that these local matches will produce a local CD cue in the same direction as the IOVD. Our grating stimulus ensures that there is no CD cue available, even from local false matches. That these new stimuli elicit VRs further

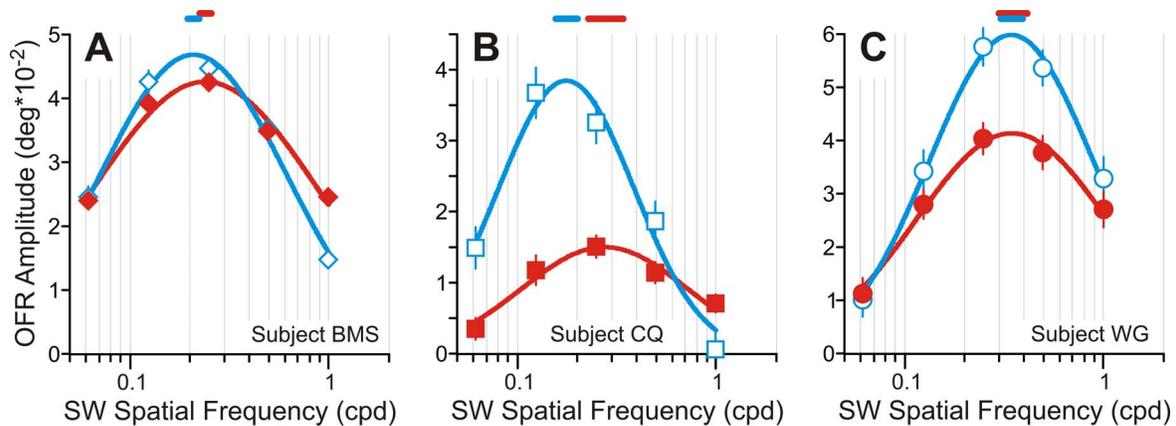


Figure 3. Semilog plots of OFR amplitude as a function of grating SF. Symbol and color conventions are as in Figure 2B–D. (A) Subject BMS: 145–150 trials per condition; (B) subject CQ: 50–71 trials per condition); (C) subject WG: 54–59 trials per condition. Color-coded horizontal lines above each graph show the 95% confidence intervals of the fits' peak locations.

strengthens the case that IOVD cues are represented separately from static disparity cues.<sup>3</sup>

In Experiment 1 the SF tuning curves appear to peak at higher SFs (median = 0.525 c/°; range = 0.35–1 c/°) than those usually observed with the short-latency ocular-following responses (OFRs) recorded during binocular frontoparallel sine-wave motion (Quaia, Sheliga, Fitzgibbon, & Optican, 2012; Sheliga, Chen, FitzGibbon, & Miles, 2005). In order to confirm this difference in SF selectivity<sup>4</sup> we ran a control experiment whose stimuli and procedures were identical to those of Experiment 1 except that the direction of motion in the two eyes was the same—i.e., frontoparallel motion. The results of this control experiment are shown in Figure 3: The median location of the SF tuning-curve peaks was 0.26 c/° and ranged from 0.18 to 0.35 c/°. These values are clearly lower than those found in Experiment 1 ( $p < 0.05$ , bootstrapping procedure; the horizontal data of subject WG are the only exception). Our previous work has shown for the OFR that speed selectivity is largely determined by the SF channels that a given stimulus activates (Sheliga et al., 2016). Thus, this shift in optimal SFs towards higher values for VRs driven by IOVD might reflect the larger behavioral relevance of slow speeds—projection geometry implies that (except at very short viewing distances) a fixed object velocity produces lower IOVD magnitudes than the mean monocular speed.

## Experiment 2: 1-D white-noise stimuli

Because of the required  $\frac{1}{4}$ -wavelength shifts between updates, our stimuli cannot be easily used to determine the temporal-frequency tuning of the IOVD-sensitive mechanism. Uncorrelated noise patterns can, however,

be used to measure its speed tuning. Furthermore, the relatively high SF preference observed in Experiment 1 predicts that, relative to the OFR system, this mechanism should be tuned to lower speeds (under the assumption that the selectivity for temporal frequency is similar for both mechanisms). In Experiment 2 we therefore measured VRs induced by uncorrelated white-noise line stimuli. This stimulus more closely matches those used in a number of previous studies (e.g., Brooks, 2002; Sanada & DeAngelis, 2014; Shioiri et al., 2000).

## Methods

Only methods that were different from those used in Experiment 1 will be described.

### Visual stimuli

Horizontal or vertical 1-D white-noise stimuli moved in opposite directions (up vs. down; left vs. right) in the two eyes. They were seen through a  $32^\circ \times 32^\circ$  rectangular aperture centered directly ahead of the eyes. All stimuli had 32% Michelson contrast, and their monocular speed varied from 6.4°/s to 102.5°/s (1–16 pixels/frame) in octave increments. A single block of trials had 20 randomly interleaved stimuli: 5 speeds  $\times$  2 axes of motion (horizontal vs. vertical)  $\times$  2 directions (for a given eye: leftward/downward vs. rightward/upward).

### Procedures

At the start of each trial, a stationary 1-D noise stimulus and a central fixation cross (width =  $10^\circ$ , height =  $2^\circ$ , thickness =  $0.2^\circ$ ) appeared on both monitors. 1-D noise stimuli were 100% correlated in the two eyes and had zero disparity—i.e., the binocular image was located at the plane of fixation. After the

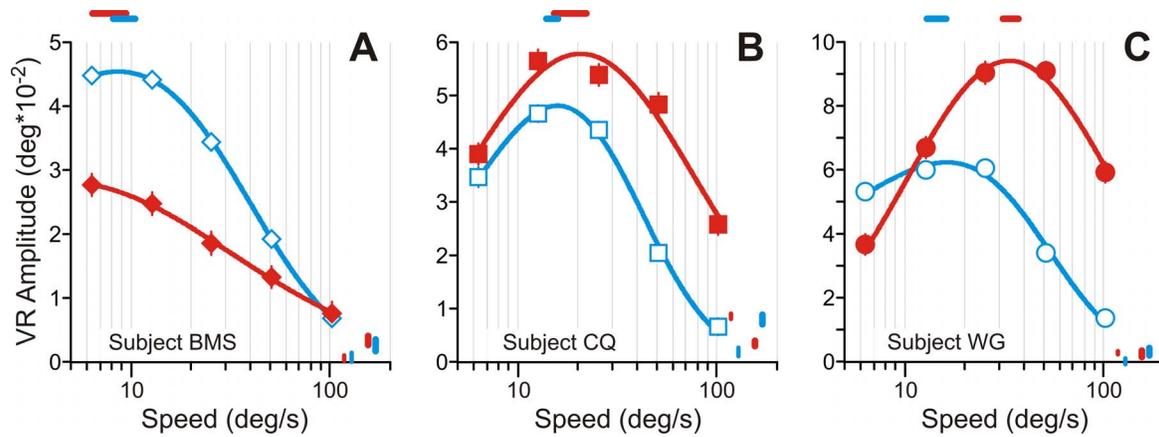


Figure 4. Experiment 2. (A–C) Semilog plots of VR amplitude against 1-D noise speed. Symbol and color conventions are as in Figure 2B–D. (A) Subject BMS: 134–140 trials per condition; (B) subject CQ: 127–148 trials per condition; (C) subject WG: 87–94 trials per condition. Color-coded horizontal lines above each graph show the 95% confidence intervals of the fits’ peak locations. Color-coded vertical lines in the right lower corner of each graph show the mean amplitude ( $\pm$ SEM) of VRs orthogonal to the axis of stimulus motion (thinner lines) and version responses along the axis of stimulus motion (thicker lines). Since the calculation of these measures was done in exactly the same way as described in Methods for the VRs along the axis of motion, the sign of version (and orthogonal vergence) responses is arbitrary. We therefore plot absolute values.

subject’s eyes had been positioned within  $2^\circ$  of the centers of the crosses for 500–1000 ms and no saccades had been detected, the stimulus seen by one of the eyes (randomly chosen) was substituted by a new one (randomly selected from a lookup table): The images were now uncorrelated in the two eyes. After another 100 ms, the crosses disappeared and motion commenced (opposite directions in the two eyes), after which the screen changed to uniform gray, marking the end of the trial. After an intertrial interval of 500 ms, the fixation crosses and new stationary 1-D noise stimuli reappeared, signaling the start of the next trial.

## Results and discussion

Uncorrelated 1-D noise stimuli elicited robust VRs in all subjects. Figure 4A through C plots VR amplitude against log stimulus speed. We fitted skewed Gaussians to log stimulus speed to estimate the optimal speed from these data (median optimal speed =  $16.1^\circ/\text{s}$ ; range =  $6.4^\circ/\text{s}$ – $33.6^\circ/\text{s}$ ; median  $r^2$  for fits = 0.995; range = 0.960–1.000; the full list of fit parameters can be found in Table A2), though the horizontal VRs of subject BMS peaked for the lowest speed tested ( $6.4^\circ/\text{s}$ ), so that the response decreased monotonically with speed in this subject. Like in Experiment 1, the horizontal VRs were stronger than vertical ones in subjects CQ and WG, but not in BMS. VRs in the direction orthogonal to the axis of stimulus motion were negligible (see also Figure 4).

Thus, Experiment 2 confirmed the observation that IOVD cues alone are sufficient to produce short-latency VRs. The intersubject idiosyncrasy in the magnitude of horizontal versus vertical VRs found in Experiment 1

was replicated. Optimal stimulus speeds ranged from  $6.4^\circ/\text{s}$  to  $33.6^\circ/\text{s}$ —i.e., they appear to be lower than those found for the OFRs during frontoparallel motion ( $\sim 40^\circ/\text{s}$ ; Gellman et al., 1990;  $35^\circ/\text{s}$ – $47^\circ/\text{s}$ ; Sheliga et al., 2016).<sup>5</sup> Also note that within each subject, if the peak SF was higher (lower) for horizontal than vertical sine gratings, the peak speed was higher (lower) for vertical than horizontal random-line stereograms (i.e., there is an inverse relationship between SF and speed, as expected if temporal-frequency tuning is constant across conditions). These observations complement the finding of Experiment 1 that the SF tuning curves of VRs peaked at SFs higher than those usually seen with the OFRs, and suggest that a common mechanism is responsible for all these responses.

## Experiment 3: The importance of the binocular overlap of images seen by each eye

Since the behavioral measure we use is a reflexive eye movement, the response represents a signal that has been integrated across the visual field. IOVD may then be computed in two different ways. It may be that the integration of motion signals across the field is done separately in each eye, followed by binocular comparison. Or it may be that the binocular comparison is made locally (presumably within binocular receptive fields), and then the binocular IOVD signal is integrated across the visual field. We therefore explored the effect of binocular overlap on responses to IOVD.

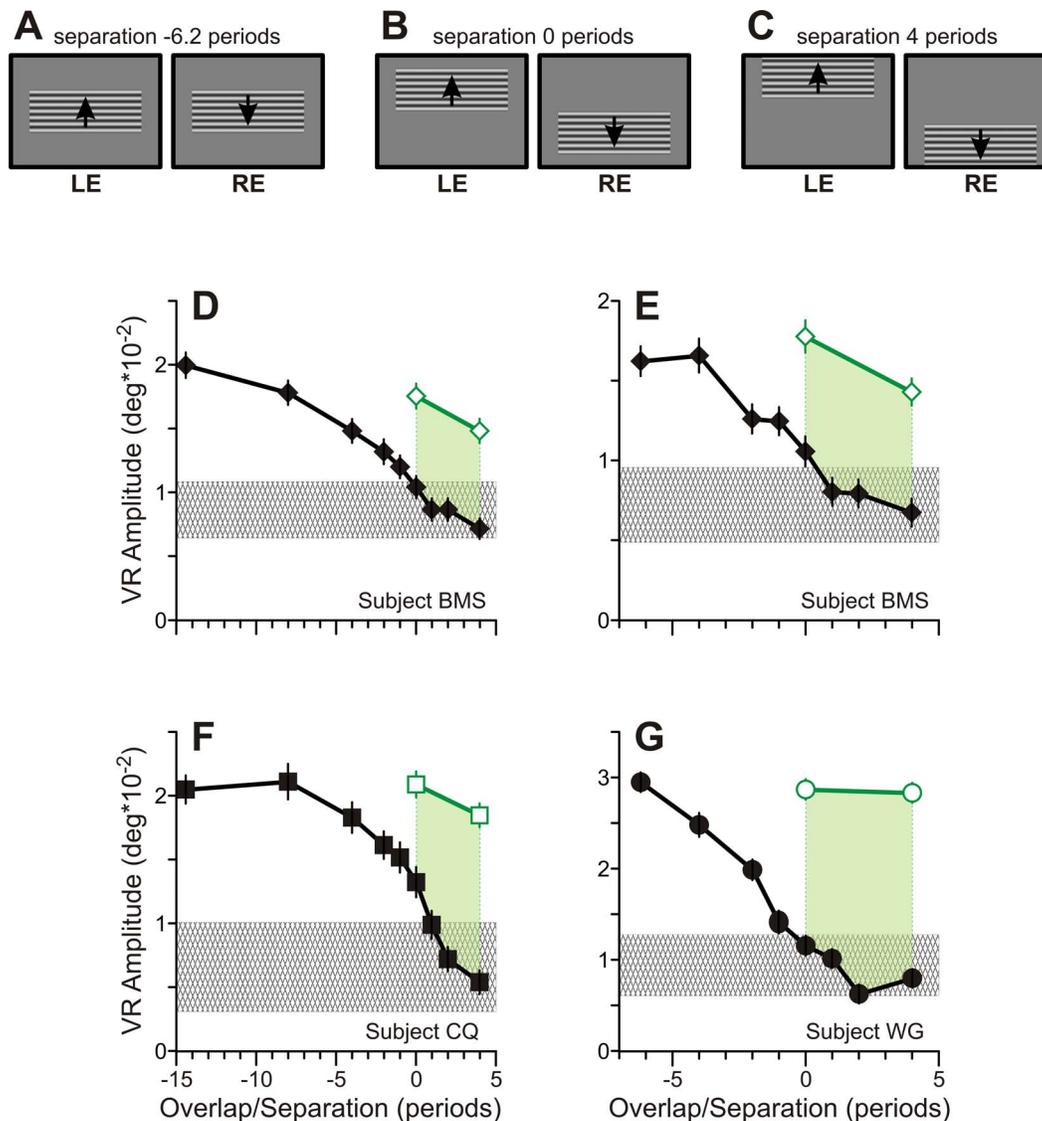


Figure 5. Experiment 3. (A–C) Three examples of stimulus spatial layout (used in E, G): (A) full binocular overlap (separation =  $-6.2$  periods); (B) abutting stimuli (separation =  $0$ ); (C) no binocular overlap, maximum vertical separation (separation =  $4$  periods). Gratings are scaled versions of  $0.5\text{-}c^\circ$  32% contrast stimuli. Vertical arrows indicate grating motion direction (up or down; opposite directions in the two eyes). LE = left-eye display; RE = right-eye display. (D–G) Black filled symbols show vertical VR amplitude as a function of binocular overlap (negative on abscissa) and separation (positive on abscissa). Green open symbols = same-eccentricity full-binocular-overlap controls. Gray hatched rectangles = 95% confidence interval for responses attributable to monocular stimulation ( $VR_{\text{mono}}$ ). (D–E) Subject BMS ( $1\text{-}c^\circ$  and  $0.5\text{-}c^\circ$  stimuli, respectively; 135–151 trials per condition); (F) subject CQ ( $1\text{-}c^\circ$  stimuli; 95–114 trials per condition); (G) subject WG ( $0.5\text{-}c^\circ$  stimuli; 113–120 trials per condition).

## Methods

Only methods that were different from those used in Experiment 1 will be described.

### Visual stimuli

1-D horizontal sine-wave gratings were utilized (i.e., vertical motion). All gratings had 32% Michelson contrast; their SF was either  $1\text{ }c^\circ$  (subjects BMS and CQ) or  $0.5\text{ }c^\circ$  (BMS and WG). The gratings' width was

always  $32^\circ$ , whereas the height was  $\sim 14.4^\circ$  for  $1\text{-}c^\circ$  stimuli and  $\sim 12.4^\circ$  for  $0.5\text{-}c^\circ$  stimuli. A pair of gratings (one grating per eye) was centered horizontally in each eye, while the vertical extent was manipulated separately in each eye, so that the pair could fully overlap, partially overlap, or be separated vertically with no binocular image overlap. When they were fully overlapping, the image in both eyes straddled the center of the screen (Figure 5A). At the maximum separation, the image in one eye was confined to the upper field while the image in the other eye was confined to the

lower field (Figure 5C). We use the term separation to describe the vertical distance between the nearest edges of the two images. Negative values of separation indicate overlapping stimuli, while zero separation is the case where the inner edges of the gratings seen by each eye abutted at the horizontal meridian (Figure 5B). Separation, expressed in multiples of the grating period, ranged from  $-14.4$  to  $4$  for  $1\text{-c}/^\circ$  stimuli and from  $-6.2$  to  $4$  for  $0.5\text{-c}/^\circ$  stimuli. The most negative values indicate complete overlap for each frequency. This arrangement means that the spatial separation covaries with vertical eccentricity, so we include two control conditions where overlapping stimuli were shown eccentrically. In these controls, the images seen by each eye fully overlapped but were placed at locations corresponding to one of the monocular images (upper or lower field at random) in the separated cases. We used configurations corresponding to separations of  $0$  or  $4$  cycles. We also presented monocular images alone (while the other eye was presented with a uniform gray screen,  $20.8\text{ cd}/\text{m}^2$ ), to control for the possibility that unequal drive from the two eyes might explain some of the response. These were located in positions corresponding to the largest binocular separation. A single block of trials had 58 (subject BMS; was shown sine waves of both SFs in a single block), 26 (CQ;  $1\text{ c}/^\circ$ ), or 28 (WG;  $0.5\text{ c}/^\circ$ ) randomly interleaved stimuli: 1 or 2 SFs  $\times$  8 or 9 overlap/separation conditions  $\times$  6 or 10 controls  $\times$  2 directions (for a given eye: downward vs. upward).

## Results and discussion

The filled black symbols in Figure 5D through G plot the dependence of the VR amplitude on spatial separation or overlap of the monocular gratings. Figure 5D and E shows data for subject BMS for  $1$  and  $0.5\text{ c}/^\circ$  stimuli, respectively; Figure 5F shows data for subject CQ ( $1\text{ c}/^\circ$ ) and Figure 5G shows data for subject WG ( $0.5\text{ c}/^\circ$ ). In all cases, reduced overlap led to weaker VRs. The control data with overlapping stimuli (open green symbols) show that the change in stimulus eccentricity had little impact.

VRs at the largest separations (rightmost black symbols in Figure 5) were small but still significantly different from zero. These might reflect a response to IOVD but could also be explained by asymmetries in the monocular responses. Although monocularly presented gratings produce a robust version response, there is also a small vergence component—the stimulated eye moves slightly faster than the unstimulated eye. We summed responses from appropriate pairs of monocular presentations to estimate the magnitude of the VR produced by this purely monocular component. Here the calculation involved combining VRs not from

two conditions (as in Equation 1) but from four:

$$VR_{\text{mono}} = (VR_{LU} + VR_{RD}) - (VR_{LD} + VR_{RU}) \quad (2),$$

where the first subscript letter denotes the eye which sees the image (left or right) and the second subscript letter denotes the direction of motion of this monocular image (up or down). A bootstrapping procedure was used to construct the 95% confidence intervals of these  $VR_{\text{mono}}$  measures, which are shown by gray hatched rectangles on each panel of Figure 5. This analysis revealed that the VRs evoked by nonoverlapping stimuli in the two eyes are no larger than those produced by monocular stimulation.

Small VRs to nonoverlapping stimuli were also reported by Masson et al. (2002), who used RDSs constructed from horizontal bands of dots that were offset between the eyes. These small responses may also have been caused by the asymmetries in monocular responses we demonstrate here. Because the overlapping stimulus in that study contained a static disparity signal, it left open the possibility that IOVD might be calculated after a broader monocular summation. Our results, combined with those of Masson et al. (2002), suggest that the requirement for binocular overlap is similar for both IOVD and static disparity processing.

## General discussion

We devised a new stimulus manipulation that isolates IOVD from changes in disparity. A grating stimulus is displaced by  $90^\circ$  with each image update. Because this displacement is in opposite directions in the two eyes, the interocular phase difference only ever takes two values— $0$  and  $180$ . Since neither of these can produce a signed response, the interocular phase difference alone cannot explain any responses. This is a subtle modification of the stimulus used by Czuba et al. (2014), which had no change in mean disparity but did have a defined disparity derivative (it used smaller phase increments). These stimuli represent an improvement on a more commonly used stimulus—uncorrelated RDSs moving in opposite directions in the two eyes—and so may also be useful for perceptual studies of motion in depth. While the uncorrelated RDS contains no systematic disparity signals, it is always possible that local false matches produce a disparity signal that changes systematically over time (locally). We found that the speed tuning of responses to IOVD in uncorrelated RDSs follows predictions based on the SF tuning of the responses we observed with gratings, suggesting that a common mechanism drives both responses. This in turn suggests that the theoretical possibility of a contribution from local matches in uncorrelated RDSs plays a minor role.

We found that the vertical IOVDs were almost as effective in evoking VRs as horizontal IOVDs. This result is surprising, since a vertical VR is not necessary when presented with motion in depth, and stimuli moving in depth do not produce a useful vertical IOVD signal: Stimuli moving in depth will cause retinal motion signals along epipolar lines. Although these may not be horizontal, they usually have opposite rotations in upper versus lower hemifields (Read, Phillipson, & Glennerster, 2009). Furthermore, they are very close to horizontal near the fovea. Consequently, we cannot think of a naturally occurring situation for which these vertical VRs are adaptive. We suggest that this might arise as a consequence of the underlying neurophysiological substrate. It has recently been shown that a substantial fraction of neurons in primate area MT are activated by opposite-direction motion in the two eyes, in a way that renders them selective for motion in depth (Czuba et al., 2014; Sanada & DeAngelis, 2014; see also Zeki, 1974a, 1974b). They thus carry a signed signal that may be a neural substrate for the eye movements we report. Such responses are much rarer in striate cortex (Pettigrew, 1973; Poggio & Talbot, 1981). Czuba et al. (2014) found that while this property is more common in neurons with preferred directions that are near horizontal, it can also be found in neurons with near-vertical preferred directions. The activity of neurons with oblique preferred directions could contribute to both responses. Note, however, that for neurons with oblique preferred directions, a different pooling is required to produce responses to vertical IOVD from that required to produce responses to horizontal IOVD. This highlights the fact that the mere existence of these signals in area MT is not sufficient to explain the responses we observe to vertical IOVD—it also requires that these be appropriately connected to the motor circuitry driving vertical vergence. What advantage this produces in natural viewing is not clear. Nonetheless, these eye movements are the first behavioral signature that has been identified for the component of the responses in area MT that responds to vertical IOVD—it is not clear that these are perceptually distinguishable. In informal observations, our subjects are not able to differentiate opposite vertical IOVD signals.

We observed considerable intersubject variability in the relative magnitude of responses to vertical IOVD compared with horizontal IOVD. Two out of three subjects showed stronger responses to horizontal IOVD. These data seem commensurate with the preference for horizontal IOVD found by Czuba et al. (2014). One subject showed the opposite pattern, with stronger responses to vertical IOVD. One possibility is that there is a similar heterogeneity in the responses of

MT neurons between animal subjects, so that the overall bias found by Czuba et al. (2014) may depend on the subjects used.

If the responses we report do depend on the activity of MT neurons, then they should require binocular overlap of the stimuli. Experiment 3 shows that this is the case. Perhaps the strongest reason for believing that neurons in area MT (or similar neurons elsewhere) play a role is that no one has previously studied the effect of vertical IOVD—the stimuli we used here cannot be produced by any stimulus under natural viewing conditions. It was the observation that neurons in area MT are selective for IOVD along axes that are not horizontal that prompted us to conduct this study.

In the perceptual domain, Rokers, Czuba, Cormack, and Huk (2011) used drifting plaid stimuli and found that motion in depth was clearly perceived when the “pattern” direction of motion—not that of the two “component” gratings—was opposite in the two eyes, which strongly suggested that the extraction of the IOVD signals occurs later than striate cortex. These authors also showed that dichoptic “pseudoplaid,” in which the components were at different spatial locations in the two eyes, produced a sensation of motion in depth. Since these stimuli have no binocular overlap over the extent of each patch, it suggests that motion signals are pooled over a broad area prior to computing IOVD. This may reflect a difference between perception and eye-movement control, since we found that without stimulus overlap, no short-latency VRs were produced (Figure 5). However, these observations are not necessarily inconsistent, depending on the spatial integration area used for computing IOVD. If this is just a few degrees of visual angle, then our abutting condition would still be expected to generate very weak responses.

## Summary

We introduce a novel visual stimulus that has advantages over more traditional ones for isolating IOVD, and show that the IOVD cue alone is sufficient for evoking short-latency VRs. VRs to differences in interocular vertical velocity presumably reflect the neuronal responses to such stimuli recently shown in area MT (Czuba et al., 2014), for which no other behavioral consequence was previously known. The responses to vertical IOVD illustrate that VRs might provide a useful behavioral signature of the activity in such neuronal populations.

*Keywords:* motion in depth, interocular velocity difference, binocular disparity, vergence eye movements

## Acknowledgments

This research was supported by the Intramural Program of the National Eye Institute at the NIH.

Commercial relationships: none.

Corresponding author: Boris M. Sheliga.

Email: bms@nei.nih.gov.

Address: Laboratory of Sensorimotor Research, National Institutes of Health, Bethesda, MD, USA.

## Footnotes

<sup>1</sup> For example, in Experiment 1 all conditions of horizontal motion for gratings of various SFs were considered to be a stimulus set. Therefore, the window for each SF condition commenced at the same time after the motion onset. The quantification of VRs to vertical motion might have a different time window determined by the shortest response latency recorded with this stimulus set.

<sup>2</sup> Version was quantified over the same temporal window as vergence (see Methods).

<sup>3</sup> Our informal observations indicate that the subjects also perceived motion in depth in our motion displays, strongest at higher SFs, but only in the case of horizontal motion.

<sup>4</sup> The OFR SF tuning to frontoparallel motion depends on a number of experimental variables, with stimulus size being one of the most important (Sheliga et al., 2013; Sheliga, Quaia, Cumming, & Fitzgibbon, 2012).

<sup>5</sup> Two of our three subjects did not participate in experiments in which we studied the horizontal frontoparallel motion of 1-D noise (reported in Sheliga et al., 2016). For these subjects, therefore, we cannot perform a statistical comparison. However, subject BMS participated in both projects, and we found that the optimal speeds for the frontoparallel stimulus motion were significantly higher than those for motion in depth ( $p < 0.05$ ; bootstrapping).

## References

- Allen, B., Haun, A. M., Hanley, T., Green, C. S., & Rokers, B. (2015). Optimal combination of the binocular cues to 3D motion. *Investigative Ophthalmology & Visual Science*, *56*(12), 7589–7596. [PubMed] [Article]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
- Brooks, K. R. (2002). Interocular velocity difference contributes to stereomotion speed perception. *Journal of Vision*, *2*(3):2, 218–231, doi:10.1167/2.3.2. [PubMed] [Article]
- Brooks, K. R., & Stone, L. S. (2004). Stereomotion speed perception: Contributions from both changing disparity and interocular velocity difference over a range of relative disparities. *Journal of Vision*, *4*(12):6, 1061–1079, doi:10.1167/4.12.6. [PubMed] [Article]
- Collewijn, H., Van Der Mark, F., & Jansen, T. C. (1975). Precise recording of human eye movements. *Vision Research*, *15*, 447–450.
- Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision Research*, *34*(4), 483–495.
- Czuba, T. B., Huk, A. C., Cormack, L. K., & Kohn, A. (2014). Area MT encodes three-dimensional motion. *The Journal of Neuroscience*, *34*(47), 15522–15533.
- Czuba, T. B., Rokers, B., Huk, A. C., & Cormack, L. K. (2010). Speed and eccentricity tuning reveal a central role for the velocity-based cue to 3D visual motion. *Journal of Neurophysiology*, *104*(5), 2886–2899.
- Erkelens, C. J., & Collewijn, H. (1985a). Eye movements and stereopsis during dichoptic viewing of moving random-dot stereograms. *Vision Research*, *25*(11), 1689–1700.
- Erkelens, C. J., & Collewijn, H. (1985b). Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Research*, *25*(4), 583–588.
- Gellman, R. S., Carl, J. R., & Miles, F. A. (1990). Short latency ocular-following responses in man. *Visual Neuroscience*, *5*(2), 107–122.
- Gray, R., & Regan, D. (1996). Cyclopean motion perception produced by oscillations of size, disparity and location. *Vision Research*, *36*(5), 655–665.
- Harris, J. M., Nefs, H. T., & Grafton, C. E. (2008). Binocular vision and motion-in-depth. *Spatial Vision*, *21*(6), 531–547.
- Harris, J. M., & Watamaniuk, S. N. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Research*, *35*(7), 885–896.
- Hays, A. V., Richmond, B. J., & Optican, L. M. (1982). A UNIX-based multiple process system for real-time data acquisition and control. *WESCON Conference Proceedings*, *2*(1), 1–10.
- Masson, G. S., Yang, D. S., & Miles, F. A. (2002). Version and vergence eye movements in humans: Open-loop dynamics determined by monocular

- rather than binocular image speed. *Vision Research*, 42(26), 2853–2867.
- Mulligan, J. B., Stevenson, S. B., & Cormack, L. K. (2013). Reflexive and voluntary control of smooth eye movements. In B. E. Rogowitz, T. N. Pappas, & H. de Ridder (Eds.), *Human vision and electronic imaging XVIII, Proceedings of SPIE* (Vol. 8651). Bellingham, WA: SPIE.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Pettigrew, J. D. (1973). Binocular neurons which signal change of disparity in area 18 of cat visual cortex. *Nature: New Biology*, 241(108), 123–124.
- Poggio, G. F., & Talbot, W. H. (1981). Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *The Journal of Physiology*, 315, 469–492.
- Portfors-Yeomans, C. V., & Regan, D. (1996). Cyclopean discrimination thresholds for the direction and speed of motion in depth. *Vision Research*, 36(20), 3265–3279.
- Quaia, C., Sheliga, B. M., Fitzgibbon, E. J., & Optican, L. M. (2012). Ocular following in humans: Spatial properties. *Journal of Vision*, 12(4):13, 1–29, doi:10.1167/12.4.13. [PubMed] [Article]
- Rambold, H. A., Sheliga, B. M., & Miles, F. A. (2010). Evidence from vergence eye movements that disparities defined by luminance and contrast are sensed by independent mechanisms. *Journal of Vision*, 10(14):31, 1–34, doi:10.1167/10.14.31. [PubMed] [Article]
- Rashbass, C., & Westheimer, G. (1961). Disjunctive eye movements. *The Journal of Physiology*, 159, 339–360.
- Read, J. C., Phillipson, G. P., & Glennerster, A. (2009). Latitude and longitude vertical disparities. *Journal of Vision*, 9(13):11, 1–37, doi:10.1167/9.13.11. [PubMed] [Article]
- Regan, D., Erkelens, C. J., & Collewijn, H. (1986). Necessary conditions for the perception of motion in depth. *Investigative Ophthalmology & Visual Science*, 27(4), 584–597. [PubMed] [Article]
- Robinson, D. A. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. *Institute of Electronic and Electrical Engineers: Transactions in Biomedical Engineering, BME-10*, 137–145.
- Rokers, B., Czuba, T. B., Cormack, L. K., & Huk, A. C. (2011). Motion processing with two eyes in three dimensions. *Journal of Vision*, 11(2):10, 1–19, doi:10.1167/11.2.10. [PubMed] [Article]
- Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area MT. *The Journal of Neuroscience*, 34(47), 15508–15521.
- Sheliga, B. M., Chen, K. J., FitzGibbon, E. J., & Miles, F. A. (2005). Initial ocular following in humans: A response to first-order motion energy. *Vision Research*, 45, 3307–3321.
- Sheliga, B. M., Quaia, C., Cumming, B. G., & Fitzgibbon, E. J. (2012). Spatial summation properties of the human ocular following response (OFR): Dependence upon the spatial frequency of the stimulus. *Vision Research*, 68, 1–13.
- Sheliga, B. M., Quaia, C., FitzGibbon, E. J., & Cumming, B. G. (2013). Retinal visual processing constrains human ocular following response. *Vision Research*, 93, 29–42.
- Sheliga, B. M., Quaia, C., FitzGibbon, E. J., & Cumming, B. G. (2016). Ocular-following responses to white noise stimuli in humans reveal a novel nonlinearity that results from temporal sampling. *Journal of Vision*, 16(1):8, 1–22, doi:10.1167/16.1.8. [PubMed] [Article]
- Shioiri, S., Saisho, H., & Yaguchi, H. (2000). Motion in depth based on inter-ocular velocity differences. *Vision Research*, 40(19), 2565–2572.
- Tyler, C. W. (1971, Nov 26). Stereoscopic depth movement: Two eyes less sensitive than one. *Science*, 174(4012), 958–961.
- Yang, D. S., FitzGibbon, E. J., & Miles, F. A. (2003). Short-latency disparity-vergence eye movements in humans: Sensitivity to simulated orthogonal tropias. *Vision Research*, 43(4), 431–443.
- Zeki, S. M. (1974a). Cells responding to changing image size and disparity in the cortex of the rhesus monkey. *The Journal of Physiology*, 242(3), 827–841.
- Zeki, S. M. (1974b). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of Physiology*, 236(3), 549–573.

## Appendix

$$A * \exp\left(-\frac{(X - X_0)^2}{2 * \sigma^2}\right) * \left(1 + \operatorname{erf}\left(\alpha * \frac{X - X_0}{\sqrt{2} * \sigma}\right)\right)$$

Motion axis	Peak amplitude	Peak SF	Log $\sigma$ (base e)	Optimal SF	$r^2$
Subject BMS					
Horizontal	0.031	13.25	3.35	1	0.966
Vertical	0.031	0.44	1.36	0.44	0.973
Subject CQ					
Horizontal	0.038	0.53	1.56	0.53	0.960
Vertical	0.013	1.18	2.31	1	0.861
Subject WG					
Horizontal	0.081	0.35	1.23	0.35	0.973
Vertical	0.046	0.52	1.46	0.52	0.995

Table A1. Experiment 1: Gaussian best-fit parameters (on a log abscissa) for the dependences of vergence-response amplitudes ( $^\circ$ ) on spatial frequency (SF) of sine-wave gratings ( $c/^\circ$ ).

Motion axis	Scaling parameter (A)	Offset ( $X_0$ )	Log $\sigma$ (base e)	Skewness ( $\alpha$ )	Optimal speed	$r^2$
Subject BMS						
Horizontal	0.028	4.7	1.93	-0.01	6.4	0.999
Vertical	0.027	36.6	3.25	-3.48	8.6	1.000
Subject CQ						
Horizontal	0.048	45.3	1.59	-0.92	20.7	0.960
Vertical	0.033	36.4	1.55	-1.91	15.9	0.997
Subject WG						
Horizontal	0.094	33.6	1.18	0	33.6	0.994
Vertical	0.037	50.5	2.52	-3.40	16.2	0.991

Table A2. Experiment 2: Skewed Gaussian best-fit parameters (on a log abscissa) for the dependences of vergence-response amplitudes ( $^\circ$ ) on monocular 1-D noise speed ( $^\circ/s$ ).