Weighted summation and contrast normalization account for short-latency disparity vergence responses to white noise stimuli in humans

Boris M. Sheliga
Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD, USA

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

Edmond J. FitzGibbon
Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD, USA

Bruce G. Cumming
Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD, USA

Natural images are typically broadband, whereas detectors in early visual processing are selective for narrow ranges of spatial frequency. White noise patterns are widely used in laboratory settings to investigate how responses are derived from Fourier components in the image. Here, we report disparity vergence responses (DVRs) to white noise stimuli in human subjects and compare these with responses to white noise patterns filtered with bandpass filters and notch filters and to sinusoidal gratings. Although the contribution of these short-latency eye movements to the overall vergence response to a given stimulus is generally small, they have proven to be a valuable tool for the study of the early mechanisms that process disparity stimuli in human subjects. Removing lower spatial frequency (SF) components reduced DVR amplitude, whereas removing higher SF components led to an increase in DVR amplitude. For larger disparities, the transition occurred at lower SFs. All of these effects were quantitatively well described by a model that combined two factors: (a) an excitatory drive determined by a weighted sum of stimulus Fourier components, which was scaled by (b) a contrast normalization mechanism.

Introduction

Vergence eye movements serve to align the eyes on an object and are critical for binocular vision. There are several visual cues to viewing distance that influence vergence, although the binocular disparity—the difference in the positions of the images on the two retinas—is the most potent at short distances. Objects nearer than the plane of fixation (crossed disparities) elicit convergence, and those farther than the plane of fixation (uncrossed disparities) elicit divergence, as expected of a negative-feedback system working to achieve and maintain appropriate binocular alignment. Though horizontal vergence has received most of attention, largely because it is the means by which the binocular alignment is achieved between objects at different depth planes, binocular vision requires vertical as well as horizontal alignment of the two lines of sight, and vertical disparities have been shown to elicit appropriate vertical vergence (e.g., Howard, Allison, & Zacher, 1997; Howard, Fang, Allison, & Zacher, 2000; Stevenson, Lott, & Yang, 1997).

Applying small disparities to large textured patterns produces vergence eye movements at very short latencies (<80 ms in humans and <60 ms in monkeys) (Bussettini, FitzGibbon, & Miles, 2001; Bussettini, Miles, & Krauzlis, 1996; Masson, Bussettini, & Miles, 1997; Masson, Yang, & Miles, 2002; Rambold & Miles, 2008; Rambold, Sheliga, & Miles, 2010; Takemura, Inoue, & Kawano, 2002; Takemura, Inoue, Kawano, Quaia, & Miles, 2001; Takemura, Kawano, Quaia, & Miles, 2002;
Takemura, Murata, Kawano, & Miles, 2007; Yang, FitzGibbon, & Miles, 2003). Many characteristics of these short-latency disparity vergence responses (DVRs) were shown to be consistent with the behavior of disparity-selective neurons in the primate striate cortex (Cumming & DeAngelis, 2001; Cumming & Parker, 1997; Ohzawa, DeAngelis, & Freeman, 1990; Prince, Cumming, & Parker, 2002; Prince, Pointon, Cumming, & Parker, 2002). The medial superior temporal cortical area was also shown to play a critical role in DVR generation (Takemura, Inoue, et al., 2002; Takemura et al., 2001; Takemura, Kawano, et al., 2002). In sum, DVRs are driven by cortical visual neurons and, thus, can be utilized as a behavioral signature of cortical neuronal mechanisms (Sheliga, FitzGibbon, & Miles, 2006; Sheliga, FitzGibbon, & Miles, 2007).

Because natural images are typically broadband, one- and two-dimensional white noise patterns are widely used in laboratory settings. In this study, we set out to evaluate quantitatively the contributions of individual Fourier components of broadband stimuli in generating human DVRs. To accomplish this task, we compared responses to four stimuli: white noise, bandpass-filtered noise, notch-filtered noise (the complement of a bandpass filter), and sinusoidal luminance gratings. We propose a simple model that suggests that the results can be accounted for by the operation of two factors: (a) an excitatory drive, determined by a weighted sum of stimulus Fourier components, which is scaled by (b) a global contrast normalization mechanism. Some preliminary results of this study were presented in abstract form elsewhere (Sheliga, Quaia, FitzGibbon, Cumming, 2016a).

Materials and methods

Many of the techniques will be described only briefly, as they are similar to those used in this laboratory in the past (Rambold et al., 2010; Sheliga, FitzGibbon, et al., 2006). 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 of this study included two paid volunteers (AGB and WG) who were naïve as to the purpose of the experiments and one of the authors (BMS). All subjects had normal or corrected-to-normal vision. Viewing was binocular.

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 et al. (2003). The signals from each eye were sampled at 1 KHz. At the beginning of each recording session, a 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 21-inch CRT computer monitor (GDM-C520, Sony Corporation, Tokyo, Japan, or p1230, HP Inc., Palo Alto, CA) through a 45° mirror, creating a binocular image 521 mm straight ahead 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 personal computer (Precision 380; Dell USA, Round Rock, TX), but the outputs of the video card of each computer (Quadro FX 5600; NVIDIA, Santa Clara, CA) were frame-locked via NVIDIA Quadro G-Sync cards, so the two were synchronously refreshed at a rate of 150 Hz. This arrangement allowed the presentation of independent images simultaneously to each eye. The monitor screens were each 41.8° wide and 32.0° high and had 1024 × 768-pixel resolution (i.e., 23.4 pixels/° directly ahead of each eye). Each monitor was driven via an attenuator (Pelli, 1997) and a video signal splitter (AC085A-R2; Black Box Corporation, Mumbai, India), allowing presentation of black-and-white images with 11-bit grayscale resolution (mean luminance of 20.8 cd/m²). Visual stimuli were seen through a 32° × 32° (768 × 768 pixels) rectangular aperture centered directly ahead of the eyes.

Experiment 1: Filtered noise stimuli

Three types of stimuli

White noise: Vertical/horizontal one-dimensional (1D) binary white noise stimuli (1Dv/1Dh) were constructed by randomly assigning a “black” or “white” value to each successive column/row of pixels. The actual luminance of the “black” and “white” pixels was set to 0.32 and 0.68 of maximal luminance, respectively,
resulting in 32% root mean square (RMS) contrast. RMS contrast was calculated using the following formula:

\[
RMS = \sqrt{\frac{\sum_{i=1}^{N} (Lum_i - Lum_{mean})^2}{N}} / Lum_{mean}
\]

using actual pixel luminance values (Lum). An example of 1Dv is shown in the left panel of Figure 1A, whereas the right panel of Figure 1A shows the results of 1D fast Fourier transform (FFT) of the stimulus along the horizontal axis to illustrate the fact that, on average, all spatial frequencies (SFs) are equally represented. Bandpass-filtered noise: The 32% RMS contrast 1D white noise images were filtered using a bandpass filter whose attenuation function was Gaussian on a log scale. The central SF of the filter varied from 0.07 to 4.56 cpd in half-octave increments, whereas the full width at half maximum (FWHM) was always set to 2 octaves. We used a fixed bandwidth on a log scale because this approximately describes the SF tuning of striate cortex neurons (e.g., De Valois, Albrecht, & Thorell, 1982). Two panels of Figure 1B show an example of a bandpass-filtered 1Dv stimulus with a central SF of 1 cpd and its FFT. Because the bandwidth was fixed on a log scale, bandpass-filtered stimuli with a higher central SF had higher RMS contrasts (Figure 1D). Notch-filtered noise: A range of SFs was removed from 32% RMS contrast 1D white noise using the same Gaussian function of SF as used for bandpass filtering. The central SF of the filter varied from 0.07 to 4.56 cpd in half-octave increments, whereas the FWHM was 2 octaves. These filters were the complement of the bandpass filters: The sum of the bandpass-filtered image and the notch-filtered image reconstructs the 1D binary noise pattern. Two Figure 1C panels show an example of a 1Dv stimulus with a notch at 1 cpd and its FFT. Because the bandwidth was fixed on a log scale, notch-filtered stimuli with a higher central SF had lower RMS contrasts (Figure 1E).

**Disparity in Experiment 1**

In each trial of Experiment 1, identical static 1D noise images were presented to both eyes, although their relative positions on the left and right retinas were manipulated. The stimulus binocular disparity was computed by subtracting the horizontal (vertical) position of the right image from the horizontal (vertical) position of the left image. That is, the horizontal disparity was positive when the stimulus seen by the left eye was to the right from the stimulus seen by the right eye (right-hyper disparity), and it was negative when the disparity was uncrossed. Vertical disparity was positive when the stimulus seen by the left eye was shifted upward with respect to the stimulus seen by the right eye (left-hyper disparity), and it was negative when the stimulus seen by the left eye was shifted downward with respect to the stimulus seen by the right eye (right-hyper disparity). Horizontal and vertical disparity stimuli were run in separate experimental sessions. In a single session, subjects were shown either bandpass-filtered or notch-filtered 1D noise stimuli, with the binocular disparity set to ±0.51° or ±1.03° (±0.68° or ±1.37° in subject WG for horizontal disparity stimuli).

Every session also contained “full” (i.e., unfiltered) 32% RMS 1D noise stimuli of ±0.51° and ±1.03° disparity (±0.68° and ±1.37° for horizontal disparity stimuli in subject WG), and all of the DVRs, recorded in separately run bandpass-filtered or notch-filtered sessions, were normalized with respect to the DVRs recorded in one of these common conditions. A single block of trials, therefore, contained 56 randomly interleaved conditions: filtered stimuli having one of 13 central SFs plus an unfiltered white noise stimulus, two disparity values, and two disparity signs.

**Experiment 2: Varying disparity in unfiltered noise**

Subjects were shown 32% RMS 1D horizontal/vertical white noise stimuli (an example of 1Dv is shown in the left panel of Figure 1A), with vertical/horizontal disparity set to ±0.09°, ±0.17°, ±0.34°, ±0.68°, ±1.37°, or ±2.73°. Horizontal and vertical disparity stimuli were presented in a single block, which resulted in 24 randomly interleaved conditions: two disparity orientations, six disparity values, and two disparity signs.

**Experiment 3: Response to interocular phase difference in sinusoidal gratings**

The disparity stimuli were 32% Michelson contrast (≈22.6% RMS contrast) vertical/horizontal sinusoidal gratings that differed in phase at the two eyes. The SF of gratings ranged from 0.0625 to 1 cpd in octave increments (from 0.044 to 0.35 cpd for vertical gratings in subject WG). Interocular difference in phase (phase disparity [PD]) ranged from 30° to 150° in 30° increments. A single block of trials contained 50 randomly interleaved conditions: five SFs, five PDs, and two PD signs. Vertical and horizontal sine wave grating were run in separate experimental sessions.

**Control experiment: Notch filters after lowpass filtering**

For two subjects (BMS and WG), vertical 1D 32% RMS contrast white noise stimuli (1 pixel wide) were randomly generated. They were lowpass filtered,
Figure 1. Noise stimuli of Experiment 1. (A) The left panel is an example of vertical 1D white noise stimulus (“vertical barcode”), a scaled version of a $32^\circ/32^\circ$ 1-pixel-wide pattern. The right panel shows the Fourier composition of a 1D 1-pixel-wide white noise stimulus. (B) The left panel is an example of bandpass-filtered (1-cpd central spatial frequency) noise, a scaled version of a $32^\circ/32^\circ$ pattern. The right panel shows the Fourier composition of 1-cpd bandpass-filtered noise. (C) The left panel is an example of notch-filtered noise, a scaled version of a $32^\circ/32^\circ$ pattern. “Notch”-filtered noise stimuli result from removing various spatial frequency bands (here, 1-cpd bandpass-filtered noise) from white noise stimuli. The right panel shows the Fourier composition of 1-cpd notch-filtered noise. (D) RMS contrast of bandpass-filtered stimuli having different central SFs. (E) RMS contrast of notch-filtered stimuli having different central SFs.
removing all components with SFs greater than 1 cpd, and were further passed through notch filters as in Experiment 1. The central SFs of the filter varied from 0.07 to 0.81 cpd in half-octave increments. In addition, the stimulus set included a low-pass-filtered stimulus without the notch filtering. Finally, an unfiltered 32% RMS 1D white noise stimulus was also shown, so that the DVRs of this control experiment and Experiment 1 could be normalized with respect to responses recorded in this common condition. The horizontal disparity applied to all stimuli was ±0.51° (subject BMS) or ±0.68° (subject WG). A single block of trials, therefore, contained 20 randomly interleaved conditions.

### Experiment procedures

Experimental paradigms were controlled by three PCs, which communicated via Ethernet (TCP/IP protocol). The first PC utilized Real-time EXPERimentation software (REX) (Hays, Richmond, & Optican, 1982), which provided the overall control of the experimental protocol, acquisition, display, and storage of the eye-movement data. Two other PCs utilized the Psychophysics Toolbox extensions of MATLAB (MathWorks, Natick, MA) (Brainard, 1997; Pelli, 1997) and generated the visual stimuli.

At the start of each trial, a central fixation cross (width 10°, height 2°, thickness 0.2°) appeared at the center of the screen on both monitors. After the subject’s eye had been positioned within 2° of the fixation target and no saccades had been detected (using an eye velocity threshold of 18°/s) for a randomized period of 800 to 1100 ms, the crosses were replaced with the disparity stimuli (randomly chosen from a lookup table) for 200 ms (30 frames), after which the screen changed to uniform gray, marking the end of the trial. After an inter-trial interval of 500 ms the fixation crosses reappeared, signaling a new trial. The subjects were asked to refrain from blinking or shifting fixation except during the inter-trial intervals but were given no instructions relating to the disparity 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 collection usually occurred 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 calibration procedure provided eye position data which were fitted with second-order polynomials and later used to linearize the horizontal and vertical eye position data recorded during the experiment. Eye-position signals 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 microsaccadic intrusions (that had failed to reach the eye-velocity cut-off of 18°/s used during the experiment) were deleted. Being consistent with our conventions for defining the polarity of the disparity stimuli, rightward and upward eye movements were defined as positive, and vergence position was computed by subtracting the horizontal (vertical) position of the right eye from the horizontal (vertical) position of the left eye. This meant that convergence and left-sursumvergence had positive signs. To improve the signal-to-noise ratio, the mean vergence response profile to each uncrossed (right-hyper) disparity stimulus was subtracted from the mean vergence response profile to the corresponding crossed (left-hyper) disparity stimulus: DVR amplitude (“pooled-difference measures”). As convergence and left-sursumvergence were positive in our sign convention, the horizontal (vertical) pooled-difference measures were positive when in the compensatory direction defined by our disparity stimuli. Response latency was estimated by determining the time after stimulus onset when the mean pooled-difference vergence velocity first exceeded 0.1°/s. The initial vergence responses to a given stimulus were then quantified by measuring the changes in the pooled-difference measures over the initial open-loop period (i.e., over the period of time up to twice the response latency) (Barthelemy, Vanzetta, & Masson, 2006; Gellman, Carl, & Miles, 1990; Takemura et al., 2000). To permit within-subject comparisons across experiments, this measurement window always commenced at the same time after the initial response latency (“stimulus-locked measures”), the actual time being determined by the shortest observed response latency: horizontal DVRs (horDVRs), 71 ms, 70 ms, and 71 ms for subjects AGB, BMS, and WG, respectively; vertical DVRs (vertDVRs), 70 ms and 82 ms for subjects BMS and WG, respectively. The duration of the measurement window was set to 40 ms throughout the entire study. This value was well within the initial open-loop period for each subject and was, in fact, substantially shorter. We chose this value, because in the vast majority of cases with 1D noise stimuli the initial vergence response drive was followed by a secondary one, which developed 40+ ms later. This observation is illustrated by vergence velocity and acceleration profiles shown in Figures 2A and 2B, respectively. We did not want our measures to be affected by any differences between early and later response components (often extending well beyond the initial open-loop period), so we restricted the duration of the measurement window accordingly. Bootstrapping procedures were used for statistical evaluation of the data and to construct 68% confidence intervals of the mean in the figures (these intervals
Figure 2. For Experiment 1, mean eye velocity (A) and eye acceleration (B) profiles over time to unfiltered white noise (dashed green trace), as well as to notch filtered noise, the central spatial frequency of which is noted by grayscale coding of velocity traces (see the insert). Subject BMS. Responses to only eight out of 13 conditions are shown to avoid clutter. Each trace is the mean response to 114–144 repetitions of the stimulus. The abscissa shows the time from the stimulus onset; the horizontal dotted line represents zero velocity, and the horizontal thick black line beneath the traces indicates the response measurement window. (C, D) Enlarged initial portion (50–115 ms from stimulus onset) of mean eye position (upper panel) and velocity (lower panel) profiles to notch and bandpass noise, respectively.
were smaller than the symbol size in many cases and, therefore, not visible on most graphs).

Results

DVRs to white noise stimuli: Data from Experiments 1 and 2

Figure 2C compares mean horizontal vergence position and velocity profiles in response to unfiltered white noise (dashed green trace) with those in response to white noise stimuli after notch filtering (noted by grayscale coding of velocity traces). These profiles were obtained from subject BMS for 0.51° horizontal disparity stimuli. Removing low-to-intermediate SF components produced horDVR reduction, as the light-gray traces fall below the dashed green trace. Conversely, removing high-SF components led to horDVR enhancement, as all dark-gray solid traces lie above the green dashed one. With bandpass filters (Figure 2D), the reversed pattern emerged, as high-SF filters resulted in weak horDVRs, and low-to-intermediate SF filters produced horDVRs that were comparable or even stronger than those to unfiltered white noise. These features are quantified in Figure 3, which shows how the DVR amplitudes change as a function of the central SF of notch and bandpass filters in three subjects. Each column of panels in Figure 3 presents data for one subject. Two upper rows are enclosed in a black-outlined rectangle and display horDVR data: 0.51° horizontal disparity stimuli (subjects AGB and BMS) or 0.68° (subject WG) in the first row, and 1.03° horizontal disparity stimuli (subjects AGB and BMS) or 1.37° (subject WG) in the second row. Two lower rows are enclosed in a green-outlined rectangle and display vertDVR data: 0.51° vertical disparity stimuli in the third row and 1.03° vertical disparity stimuli in the fourth row. Filled green diamonds show DVRs to unfiltered white noise; red open circles to notch filters; and blue filled circles to bandpass filters. Several observations can be readily made:

- Troughs (notch filters) and peaks (bandpass filters) are observed at low to intermediate SFs of 0.2 to 0.5 cpd.
- For each disparity condition (i.e., in every panel of Figure 3), the locations of the trough and the peak are very similar.
- For each subject, troughs and peaks for larger disparity stimuli occur at lower SFs (compare Figure 3 panels in rows 2 and 4 to Figure 3 panels in rows 1 and 3, respectively).

In sum, lower SF components appear to provide most of the “drive” for the DVRs, though the SF of Fourier components having the most impact on DVRs appears to depend on stimulus disparity.

Figure 4 shows the results of Experiment 2, in which unfiltered 1D white noise stimuli had horizontal (Figures 4A to 4C) or vertical (Figures 4D and 4E) disparities ranging from 0.09° to 2.73°. In all cases, the dependencies were bell shaped, with maximal DVRs recorded at disparities of 0.3° to 0.7°. In two subjects (BMS and WG), in whom both horizontal and vertical disparities were tested, those dependencies had quite similar disparity tuning.

DVRs to white noise stimuli: Model

To model DVRs to broadband stimuli (such as the white noise stimuli of this study) one would need to have information regarding PD tuning of individual Fourier components. It is often assumed that an interocular phase difference of 90° produces the strongest responses, but this has not been checked empirically. We therefore measured DVRs to sine wave gratings of various SFs, for which the interocular difference in phase (PD) was systematically manipulated (Experiment 3; see Materials and Methods). For each sine wave tested (five different SFs for each subject), the PD dependence was fit by the following equation:

$$DVRSW = A \times \sin(PD) \quad (1)$$

This function is necessarily maximal at 90° and provides a good fit in all subjects for higher SFs (>0.25 cpd). For lower SFs, the PD tuning was asymmetrical, so that the maximal DVRs were recorded at PDs somewhat less than 90°. This asymmetry could be described by the following equation:

$$DVRSW = A \times \sin\left(\left(\frac{PD}{\pi}\right)^n \times \pi\right) \quad (1a)$$

(see Figures A1 and A2 in the Appendix), although there is no easy mechanism we can propose that explains why this should be so; Equation 1a is purely descriptive.
Figure 3. For Experiment 1, dependence of mean DVR amplitude on central SFs of the notch filter (data, open red circles; Equation 3 fits, dashed red lines), bandpass filter (data, filled blue circles; Equation 3 fits, solid blue lines), and unfiltered white noise (data, filled green diamond; Equation 3 fits, open green square). Data for the horizontal disparity conditions are enclosed in a large black-outlined rectangle. The panels in the first row show smaller disparity (0.51°, 0.51°, and 0.68° for subjects AGB, BMS, and WG, respectively); panels in the second row show larger disparity (1.03°, 1.03°, and 1.37° for subjects AGB, BMS, and WG, respectively). Data for the vertical disparity conditions are enclosed in a large green-outlined rectangle; panels in the first row show smaller disparity (0.51° for both subjects); panels in the second row show larger disparity (1.03° for both subjects). The data for each subject occupy a column of panels: subject AGB (A, D; 108–126 trials per condition); subject BMS (B, E, G, I; 114–144 trials per condition); subject WG (C, F, H, J; 148–158 trials per condition).
Given this, it is uncertain if this asymmetry would apply in broadband stimuli. Indeed, subsequent modeling of DVRs to noise stimuli (described below) revealed that using Equation 1a did not improve the fits in any significant way. We, therefore, stick with Equation 1 as a model for Fourier component responses, in which the amplitude ($A$) is the only free parameter. Figure 5 shows how $A$ changes with SF (open circles) in Experiment 3, and those were very well fit by a Gaussian function of log(SF) ($r^2$ range, 0.979–0.998):

$$A = A_{MAX} \times e^{-\frac{[\log(SF) - \mu]^2}{2\sigma^2}}$$  

(2)

This Gaussian has three free parameters: $A_{MAX}$, $\mu$, and $\sigma$. These will be the first three free parameters of the model (presented below) that describes the SF tuning.

The full model describes DVR amplitudes to broadband stimuli of Experiments 1 and 2 with the following equation:

$$DVR = \frac{C^n}{C^n + C_{50}^{n1}} \times \left[ \frac{\sum_{i=1}^{N} (DVR_i \times (W_i \times C_i)^n)^{\frac{1}{n}}} {\sum_{j=1}^{N} (W_j \times C_j)^n} \right]^{\frac{1}{n}}$$  

(3)

where $DVR_i$ is the response to a given Fourier component, derived from Equation 1:

$$DVR_i = A_i \times \sin(PD_i)$$  

(3a)
Figure 5. For Experiment 3, DVRs were measured to sine-wave gratings of various SFs, for which the interocular difference in phase was systematically manipulated: PD tuning. Open circles indicate amplitudes (parameter $A$) of Equation 1 fits for sine waves of different SFs (abscissa). Thick solid lines indicate Equation 3 fits. Horizontal/vertical PD tuning data are enclosed in black- and green-outlined rectangles, respectively. The data for each subject occupy one column of panels: subject AGB (A; 124–133 trials per condition); subject BMS (B, D; 103–159 trials per condition); subject WG (C, E; 103–156 trials per condition).

where $PD_i$ is a phase disparity of this Fourier component due to the disparity step applied to the stimulus. $A_i$ is given by the Gaussian fit to the SF tuning (Equation 2). $N$ is the number of Fourier components in the image (384 in our case).

$DVR_i$ is multiplied by the contrast of that component in the image ($C_i$) and its weight ($W_i$), normalized by the weighted sum of the contributions of all Fourier components present in the stimulus. This contribution of a single Fourier component is raised to the power $n_1$, the fourth free parameter of the model. The weights of the Fourier components were modeled by a power function of SF:

$$W_i = SF_i^{-K_w}$$

with a single free parameter $K_w$, the fifth free parameter of the model. A competition between contributions of different components is modeled by a power law summation; $n_2$ is the sixth free parameter of the model. $C$ is the overall RMS contrast of the stimulus. $C_{50}$ and $m$ are the last two (seventh and eighth) free parameters of the model. For pure sine wave gratings, Equation 3 simplifies to

$$DVR = \frac{C^m}{C^{m} + C_{50}^m} \times DVR_{MAX}$$

which is a well-known Naka–Rushton equation (Naka & Rushton, 1966), successfully used to describe DVR contrast dependencies to pure sine wave stimuli in the past (Quaia, Optican, & Cumming, 2017; Rambold et al., 2010; Sheliga, FitzGibbon, et al., 2006). $C_{50}$ and $m$ are, therefore, the Naka–Rushton semi-saturation contrast and power term, respectively.
Table 1. Best-fit values of Equation 3 free parameters. The best-fit values of five free parameters were significantly different for the horizontal versus vertical DVRs and are highlighted in bold in the table.

Equation 3 provided quite good fits to the data for horizontal DVRs ($r^2 = 0.796, 0.916,$ and 0.842 for subjects AGB, BMS, and WG, respectively) and vertical DVRs ($r^2 = 0.948$ and 0.918 for subjects BMS and WG, respectively. They are shown by green open squares (unfiltered white noise) and by blue solid (band-pass stimuli) and red dashed (notch stimuli) traces in Figure 3, and by black solid traces in Figure 4 (unfiltered white noise) and Figure 5 (sine-wave gratings). The best-fit values of free parameters are
Figure 7. From the control experiment, dependences of mean DVR amplitudes on central SFs of the notch filter for high-SF absent (data, filled black circles; Equation 3 fits, dashed black lines) and high-SF present (data, open red circles, replotted from Figure 3; Equation 3 fits, solid red lines) white noise stimuli. Unfiltered white noise data and fits are shown on the ordinate. (A) Subject BMS (114–135 trials per condition). (B) Subject WG (128–158 trials per condition).

Equation 3 has five free parameters beyond those that describe the linear stage: $K_w$, $n_1$, $n_2$, $C_{50}$, and $m$. We used a general linear $F$-test to determine if each of these parameters was statistically justified. For these tests, the effects of $K_w$, $n_1$, and $n_2$ were calculated individually, and $C_{50}$ and $m$ were evaluated as a pair (it makes little sense to use a Naka–Rushton equation with only one free parameter). In one subject (WG) for one parameter ($n_1$) this was not significant ($p > 0.55$). All of the other 19 cases were highly significant ($p < 0.01$ in all cases).

Figures 6A and 6B and Figures 6C and 6D show the best-fit $A_i$ function, $A_i = F(SF_i)$, and $W_i$ function, $W_i = F(SF_i)$, respectively, for each subject. Although there is some inter-subject variability, a universal feature is that high SF components have less influence on the response in broadband stimuli than expected from the responses to single gratings (Figures 6C and 6D).

Control experiment: Multiplicative scaling

The model developed above makes a specific prediction about the effect of removing high-spatial-frequency components from the stimulus. DVR is very small for frequencies greater than 1 cpd (see Figure 5), so removing these components has little effect on the DVR drive. Nonetheless, these components make a significant contribution to the denominator inside the summation term of Equation 3 (they account for most of the stimulus contrast; see Figure 1). So, removing these components makes the denominator smaller and should therefore make the responses larger. Furthermore, this response increase should be multiplicative, which would not be the case if, for example, the suppressive effect simply reflected aliasing at a high SF that introduced a response component with the wrong sign. (There is some evidence of aliasing in Figures 3I and 3J, so it is important to exclude this explanation.) We therefore measured responses to a stimulus with SFs > 1 cpd removed. Although this lowpass filter reduced RMS contrast from 32% to 10%, DVR response magnitudes increased (Figure 7). By then applying notch filters (as in Experiment 1) to this lowpass-filtered stimulus, we also vary the strength of DVR drive. Figure 7 shows that all stimuli lacking high SFs (black filled circles and filled diamond) resulted in stronger DVRs than the ones with the high-SF components (red open circles and open diamond; replotted from Figure 3). Most importantly, the effect is clearly multiplicative, as the two curves are related by a scaling. These data were well described by our model. In fact, the horizontal DVRs of subjects BMS and WG shown in Figures 3 to 5 are fit using a single set of parameters (see Table 1) to explain all of the data for Experiments 1, 2, and 3 and the control experiment. In Figure 7, the fits are shown by dashed black (control experiment) and solid red lines (Experiment 1).

Discussion

Our simple model (Equation 3) reproduced DVRs to broadband stimuli quite well. Two mechanisms (beyond linear summation) were incorporated into the model: a weighted excitatory drive and a contrast normalization. The weighted excitation is not just a linear weighting but also contains a nonlinearity modeled by the exponent $n_1$. Values of $n_1 > 1$ produce winner-take-all interactions. However, the best-fitting value of $n_1$ here was sometimes less than 1 (see Table 1). At first sight this may seem at odds with the winner-take-all
interaction shown by Sheliga, FitzGibbon et al. (2006). However, that study used two sinusoidal gratings with similar SFs. A subsequent study (in the motion domain, measuring short-latency ocular-following responses) found that exponents $>1$ are only required to describe the interaction for nearby spatial frequencies (Sheliga, Quaia, FitzGibbon, & Cumming, 2020). In broadband stimuli like those used here, there are many possible interactions between component SF pairs, so the value of $n1$ we find presumably reflects an average over all of these. Nonetheless, in both studies, the weight of a given component was not related to the response amplitude observed when that component was presented in isolation. The weighting of component contributions was not necessary in an earlier study, which modeled ocular following responses (OFRs) to white noise stimuli (Sheliga, Quaia, FitzGibbon, & Cumming, 2016b). That study, however, did not use OFRs to pure sine-wave gratings to estimate the contribution of Fourier components; instead, it used bandpass stimuli. It is possible that the impact of component weighting was already reflected in the OFRs to bandpass stimuli themselves and, therefore, was no longer necessary as a separate term in the equations of the model.

The contrast normalization probably reflects divisive inhibition among populations of cortical neurons sensitive to different SF components of visual stimuli (Britten & Heuer, 1999; Carandini & Heeger, 1994; Carandini, Heeger, & Movshon, 1997; Heeger, 1992; Heuer & Britten, 2002; Simoncelli & Heeger, 1998). Our models suggest that, in the context of the short-latency reflexive eye movements, the contributions of neural populations tuned to different SFs are weighted during summation for both motion-driven and disparity-driven eye movements. In both cases, this assigned larger weights to lower SFs. Whether this weighted summation is a more general principle deserves further investigation.

Different components of Equation 3 were used in models of our previous studies. The power law summation $(\sum R_{ij}^k)^{1/k}$ was used previously to model eye responses (OFRs) to white noise motion stimuli (Sheliga et al., 2016b). The term $(W_i \times C_j)^{n}$ was used to successfully model the OFRs to stimuli comprised of two or three sine waves moving in the same or opposite directions (Sheliga et al., 2020). Equation 3 also fits very well our old sine-wave data for both DVRs (Sheliga et al., 2007) and OFRs (Sheliga, Kodaka, FitzGibbon, & Miles, 2006), accounting for responses to a sum of two sine waves and to pure sine-wave stimuli, as well. With compound stimuli, the PD of the individual components might not be the only factor that could impact the amplitude of DVRs. The monocular phase alignment of the components (which affects monocular contrast), the PD of contrast modulations at the beat frequencies (so-called second-order disparity), and potentially even the position disparity of “features” (so-called third-order disparity) could also have an effect. Some of these factors are indeed known to affect the DVRs (Rambold et al., 2010; Quaia, personal communication). However, in this study, we used broadband stimuli in which many frequencies are present, and the contributions of these other factors will be incongruent and tend to wash out. This is further assured by the fact that we were averaging the responses to presentations of many different noise samples (50), so that small net contributions in one direction in one sample will be canceled out by small net contributions in the opposite direction in another.

We measured both horizontal and vertical DVRs in only two subjects (BMS and WG), so it might be premature to draw any definite conclusions comparing horizontal and vertical responses. Nevertheless, for both subjects the best-fit values of five free parameters were significantly different for the horizontal versus vertical DVRs; they are highlighted as bold numbers in the vertical-response cells of Table 1. Three of five are free parameters for SF tuning: amplitude, offset, and sigma of the Gaussian fit. For horDVRs, the amplitude is smaller and the sigma is larger, but the offset is shifted toward higher SFs. In our 2006 paper, we used pure sine-wave gratings (Sheliga, FitzGibbon, et al., 2006) and observed similar trends for the offset and sigma free parameters, whereas the differences in horizontal/vertical amplitude appeared to be idiosyncratic (see also Rambold & Miles, 2008; Quaia, personal communication). Also, free parameter $n1$ was close to 1 for horDVRs, but it was much higher for vertDVRs ($n1 > 1.5$). Finally, differences in SF weights (free parameter $K_W$) are more accentuated for horDVRs than for vertical ones. More noteworthy, however, may be the fact that, despite quite different roles of horizontal and vertical disparity signals in scene depth analyses, the earliest processing stages of both signals appear to be very similar: The same model successfully captures the characteristics of horizontal as well as vertical DVRs.

**Keywords:** broadband visual stimuli, phase disparity tuning, horizontal and vertical disparity

**Acknowledgments**

Supported by the Intramural Research Program of the National Eye Institute at the National Institutes of Health.

Commercial relationships: none.
Corresponding author: Boris M. Sheliga.
Email: bms@lsr.nei.nih.gov.
Address: Laboratory of Sensorimotor Research, National Institutes of Health Building 49 Room 2A50, 49 Convent Drive, Bethesda, MD 20892-4435, USA.
Footnotes

1The horizontal DVRs of subject WG to ±0.51° and ±1.03° were considerably smaller than those of the other two subjects. Using ±0.68° and ±1.37° horizontal disparities instead boosted the amplitude of horizontal DVRs in this subject. This was the only reason for using larger horizontal disparity values in subject WG.

2Sursumvergence is the upward deviation of one eye in relation to the other. Thus, left sursumvergence means that the left eye moved upward more than the right one. In our sign convention, left/right sursumvergence (vertical vergence) is analogous to convergence/divergence (horizontal vergence).

3We could use instantaneous velocity at a single time point to quantify the responses. As Figure 2 shows, over our analysis window the velocity profiles are scaled versions of one another, so the results would be extremely similar; there is a largely fixed relationship between the instantaneous DVR velocity and the total displacement. Our displacement measure is more robust to noise, as it averages the velocity over the entire time window.

4In contrast, the effects of the removal of low and/or intermediate SFs would be problematic to interpret, as both the numerator and denominator of Equation 3 become smaller.

References

Barthelemy, F. V., Vanzetta, I., & Masson, G. S. (2006). Behavioral receptive field for ocular following in humans: Dynamics of spatial summation and center-surround interactions. *Journal of Neurophysiology, 95*(6), 3712–3726.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Britten, K. H., & Heuer, H. W. (1999). Spatial summation in the receptive fields of MT neurons. *The Journal of Neuroscience, 19*(12), 5074–5084.

Busettini, C., FitzGibbon, E. J., & Miles, F. A. (2001). Short-latency disparity vergence in humans. *Journal of Neurophysiology, 85*(3), 1129–1152.

Busettini, C., Miles, F. A., & Krauzlis, R. J. (1996). Short-latency disparity vergence responses and their dependence on a prior saccadic eye movement. *Journal of Neurophysiology, 75*(4), 1392–1410.

Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science, 264*(5163), 1333–1336.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of Neuroscience, 17*(21), 8621–8644.

Collewijn, H., Van Der Mark, F., & Jansen, T. C. (1975). Precise recording of human eye movements. *Vision Research, 15*, 447–450.

Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience, 24*, 203–238.

Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature, 389*(6648), 280–283.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*(5), 545–559.

Gellman, R. S., Carl, J. R., & Miles, F. A. (1990). Short latency ocular-following responses in man. *Visual Neuroscience, 5*(2), 107–122.

Hays, A. V., Richmond, B. J., & Optican, L. M. (1982). UNIX-based multiple process system for real-time data acquisition and control. In: *WESCON Convention Record* (pp. 1–10). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*(2), 181–197.

Heuer, H. W., & Britten, K. H. (2002). Contrast dependence of response normalization in area MT of the rhesus macaque. *Journal of Neurophysiology, 88*(6), 3398–3408.

Howard, I. P., Allison, R. S., & Zacher, J. E. (1997). The dynamics of vertical vergence. *Experimental Brain Research, 116*(1), 153–159.

Howard, I. P., Fang, X., Allison, R. S., & Zacher, J. E. (2000). Effects of stimulus size and eccentricity on horizontal and vertical vergence. *Experimental Brain Research, 130*(2), 124–132.

Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature, 389*(6648), 283–286.

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.

Naka, K. I., & Rushton, W. A. (1966). S-potentials from colour units in the retina of fish (Cyprinidae). *Journal of Physiology, 185*(1), 153–159.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science, 249*(4972), 1037–1041.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Prince, S. J., Cumming, B. G., & Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology, 87*(1), 209–221.

Prince, S. J., Pointon, A. D., Cumming, B. G., & Parker, A. J. (2002). Quantitative analysis of the
responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *Journal of Neurophysiology, 87*(1), 191–208.

Quaia, C., Optican, L. M., & Cumming, B. G. (2017). Combining 1-D components to extract pattern information: It is about more than component similarity. *Journal of Vision, 17*(3):21, 1–13, https://doi.org/10.1167/17.3.21.

Rambold, H. A., & Miles, F. A. (2008). Human vergence eye movements to oblique disparity stimuli: Evidence for an anisotropy favoring horizontal disparities. *Vision Research, 48*(19), 2006–2019.

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, https://doi.org/10.1167/10.14.31.

Robinson, D. A. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Transactions on Bio-medical Electronics, 10*(4), 137–145.

Sheliga, B. M., FitzGibbon, E. J., & Miles, F. A. (2006). Short-latency disparity vergence eye movements: A response to disparity energy. *Vision Research, 46*(21), 3723–3740.

Sheliga, B. M., FitzGibbon, E. J., & Miles, F. A. (2007). Human vergence eye movements initiated by competing disparities: Evidence for a winner-take-all mechanism. *Vision Research, 47*(4), 479–500.

Sheliga, B. M., Kodaka, Y., FitzGibbon, E. J., & Miles, F. A. (2006). Human ocular following initiated by competing image motions: Evidence for a winner-take-all mechanism. *Vision Research, 46*(13), 2041–2060.

Sheliga, B. M., Quaia, C., FitzGibbon, E. J., & Cumming, B. G. (2016a). High spatial frequency components of white noise stimuli suppress initial disparity-vergence responses in humans. 2016 *Neuroscience Meeting Planner, Program No. 328.17*. San Diego, CA: Society for Neuroscience.

Sheliga, B. M., Quaia, C., FitzGibbon, E. J., & Cumming, B. G. (2016b). 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, https://doi.org/10.1167/16.1.8.

Sheliga, B. M., Quaia, C., FitzGibbon, E. J., & Cumming, B. G. (2020). Short-latency ocular-following responses: Weighted nonlinear summation predicts the outcome of a competition between two sine wave gratings moving in opposite directions. *Journal of Vision, 20*(1):1, 1–23, https://doi.org/10.1167/jov.20.1.1.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research, 38*(5), 191–208.

Stevenson, S. B., Lott, L. A., & Yang, J. (1997). The influence of subject instruction on horizontal and vertical vergence tracking. *Vision Research, 37*(20), 2891–2898.

Takemura, A., Inoue, Y., & Kawano, K. (2000). The effect of disparity on the very earliest ocular following responses and the initial neuronal activity in monkey cortical area MST. *Neuroscience Research, 38*(1), 93–101.

Takemura, A., Inoue, Y., & Kawano, K. (2002). Visually driven eye movements elicited at ultra-short latency are severely impaired by MST lesions. *Annals of the New York Academy of Sciences, 956*, 456–459.

Takemura, A., Inoue, Y., Kawano, K., Quaia, C., & Miles, F. A. (2001). Single-unit activity in cortical area MST associated with disparity-vergence eye movements: Evidence for population coding. *Journal of Neurophysiology, 85*(5), 2245–2266.

Takemura, A., Kawano, K., Quaia, C., & Miles, F. A. (2002). Population coding in cortical area MST. *Annals of the New York Academy of Sciences, 956*, 284–296.

Takemura, A., Murata, Y., Kawano, K., & Miles, F. A. (2007). Deficits in short-latency tracking eye movements after chemical lesions in monkey cortical areas MT and MST. *The Journal of Neuroscience, 27*(3), 529–541.

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.
Appendix

Figure A1. In Experiment 3, horizontal DVRs were measured to sine-wave gratings of various SFs for which the horizontal interocular difference in phase was manipulated: horizontal PD tuning. (A, D, G) PD tuning for sine waves of different SFs (coded by different symbols; see rectangular inserts). Solid lines indicate Equation 1a fits to the data. (B, E, H) Symbols indicate Equation 1a best-fit exponent (n) values, and lines indicate cumulative Gaussian fits of exponent (n) dependencies on sine-wave SFs. (C, F, I) Symbols indicate Equation 1a best-fit amplitude (A) values, and lines indicate Gaussian fits of amplitude (A) dependencies on sine-wave SFs.
Figure A2. In Experiment 3, vertical DVRs were measured to sine-wave gratings of various SFs for which the vertical interocular difference in phase was manipulated: vertical PD tuning. (A, D) Phase PD tuning for sine waves of different SFs. Solid lines indicate Equation 1a fits to the data. (B, E) Symbols indicate Equation 1a best-fit exponent (n) values, and lines indicate cumulative Gaussian fits of exponent (n) dependencies on sine-wave SFs. (C, F) Symbols indicate Equation 1a best-fit amplitude (A) values, and lines indicate Gaussian fits of amplitude (A) dependencies on sine-wave SFs. All conventions are as in Figure A1.