The Initial Disparity Vergence Elicited With Single and Dual Grating Stimuli in Monkeys: Evidence for Disparity Energy Sensing and Nonlinear Interactions

K. Miura, Y. Sugita, K. Matsuura, N. Inaba, K. Kawano, and F. A. Miles

Submitted 7 May 2008; accepted in final form 27 August 2008


In accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

INTRODUCTION

When small binocular misalignments (disparities) are applied to large textured patterns, vergence eye movements are elicited at ultra-short latencies in both humans and monkeys, exactly as expected of negative-feedback mechanisms using binocular disparity to eliminate vergence errors: the initial disparity vergence response (DVR) (Busettini et al. 1996, 2001; Masson et al. 1997, 2002; Takemura et al. 2001; Yang et al. 2003). If the binocular patterns are anticorrelated random noise, a pure second-order disparity stimulus. It is not clear if these responses to second-order attributes occur at short latencies.

In recent experiments on humans, Sheliga et al. (2006b) used a special one-dimensional (1D) broadband stimulus, a square wave with a missing fundamental (cf., Adelson 1982), which was identical at the two eyes except for a phase difference (disparity) of one-quarter wavelength. The initial DVRs associated with this stimulus were always in the opposite direction to those with more usual patterns like random dots, e.g., uncrossed disparities resulted in convergent eye movements. Sheliga et al. argued that the Fourier composition of the missing fundamental was the critical determinant of the vergence responses here, consistent with early spatial filtering of the monocular visual inputs before their binocular combination and mediation by detectors sensitive to disparity energy. In addition, responses to the 3f5f stimulus showed a nonlinear dependence on the relative contrasts of the two sine waves. Thus on average, when the contrast of one sine wave was 2.3 times greater than that of the other, the one with the lower contrast was largely ineffective as though suppressed, and responses were determined almost entirely by the sine wave of higher contrast: Winner-Take-All. The logic is very similar to that described previously on the vertical vergence responses of humans, indicating that the monkey provides a good animal model for studying these disparity vergence responses.
When the gratings were of equal contrast but the interest: the vertical gratings with competing disparities. Their findings in humans when horizontal disparities were applied to two thyroid and Blake (1984) had earlier examined perceived depth. Boothyrd and Blake (1984) argued that this would explain why the initial DVRs were in the opposite direction to those with broadband patterns like random dots. In fact, Sheliga et al. found that the initial DVRs elicited when one-quarter wavelength disparities were applied to missing fundamental stimuli generally approximated those elicited when the two eyes saw pure sine-wave stimuli whose spatial frequency, contrast, and binocular disparity matched those of the third harmonic, consistent with the idea that the earliest DVRs depended mostly on the binocular disparity of this single most prominent harmonic, i.e., the underlying disparity detectors were sensitive to the first-order disparity energy rather than the second-order features. These findings were reminiscent of the earlier psychophysical findings of Boothyrd and Blake (1984), who applied disparities to missing fundamental gratings and reported that, under some conditions, the perceived depth often corresponded to the disparity of the third harmonic (although under other conditions—low spatial frequency and high contrast—the disparity of the overall pattern could determine the perceived depth).

However, the missing fundamental is a broadband stimulus: when one-quarter wavelength binocular disparities are applied, all of the $4n + 1$ harmonics (where $n$ is an integer), such as the 5th, 9th, 13th, etc., will have disparities equal to one-quarter of their wavelengths in the same (forward) direction, whereas all of the $4n - 1$ harmonics, such as the 3rd, 7th, 11th, etc., also have disparities equal to one-quarter of their wavelengths but in the opposite (backward) direction. Sheliga et al. (2007b) subsequently showed, again on humans, that when the missing fundamental stimulus was reduced to just two competing harmonics, the third and fifth (termed the 3f5f stimulus), whose binocular images had one-quarter wavelength phase differences (i.e., disparities of opposite sign), the initial vergence responses showed a nonlinear dependence on the relative contrast of those two sine waves. Thus when the 3f and 5f components differed in contrast by more than about an octave, the one with the higher contrast dominated the vergence responses almost completely and the one with lower contrast lost almost all of its influence as though suppressed: winner-take-all (WTA). These nonlinear interactions were attributed to mutual inhibition between the neural mechanisms responding to the disparities of the two competing harmonics. Interestingly, these nonlinear interactions were generally more powerful for vertical vergence than for horizontal vergence. Boothyrd and Blake (1984) had earlier examined perceived depth in humans when horizontal disparities were applied to two vertical gratings with competing disparities. Their findings when they combined 1f and 3f gratings are of particular interest: the 1f component alone determined perceived depth when the gratings were of equal contrast but the 3f component alone determined the percept when the contrast of the 1f component was selectively reduced, even though that 1f component still gave a robust percept of depth when presented alone. The authors invoked "masking" to explain this effect.

All previous studies on the short-latency DVRs of the monkey (Busettini et al. 1996; Masson et al. 1997; Takemura et al. 2001) used only horizontal disparities with random dot patterns and did not study either the importance of the harmonic composition of the binocular stimuli or the possibility that there were nonlinear interactions between the neural mechanisms sensing competing harmonics. This study on monkeys sought to address some of these gaps in our knowledge by applying vertical disparities of one-quarter wavelength to various 1D horizontal grating patterns. These patterns included pure sinusoidal gratings and 3f5f stimuli whose two component sine waves had spatial frequencies in the ratio 3:5 and disparities of opposite sign. We will show that the monkey's initial vertical vergence responses share many characteristics with those of humans, in particular, a strong dependence on the Fourier composition of the disparity stimulus (rather than its raw features), consistent with mediation by detectors sensitive to first-order disparity energy. In addition, we present evidence that there are powerful inhibitory interactions between the neural mechanisms sensing the disparity of the competing harmonics in the 3f5f stimulus that result in nonlinear vergence responses (WTA) in monkeys very similar to those previously observed in humans. These data provide further evidence that the monkey is a good animal model for the initial disparity vergence eye movements of the human (cf. Miles 1998).

**METHODS**

Data were collected from four adolescent rhesus monkeys (*Macaca mulata*), weighing 6–8 kg. All procedures were approved by the Animal Care and Use Committee of Kyoto University.

**Animal preparation**

All animals had been previously trained to fixate a small target spot for a liquid reward. Under sodium pentobarbital anesthesia and aseptic conditions, each monkey was fitted with a pedestal, which was secured to the skull through implanted bolts and dental acrylic, to allow the head to be fixed in a standard stereotaxic position. Scleral search coils were also implanted around both eyes for the purpose of monitoring eye movements using an electromagnetic induction technique (Fuchs and Robinson 1966; Judge et al. 1980).

**Visual display and grating stimuli**

The apparatus used to present the visual stimuli was similar to that used in a previous recent study (Hayashi et al. 2008). Two CRT monitors (Eizo T766) were covered with orthogonal polarizing filters (Edmund Techspec linear polarizing laminated film), and the images on the two monitors were superimposed with a half-silvered mirror. Dichoptic stimulus presentation was achieved by matching orthogonal polarizing filters placed in front of the monkey’s eyes. Two crossed polarizing filters with an average transmission with 0.04% were used. The images seen by the left and right eyes were separately derived from the green and blue video outputs of a PC (8-bit resolution). These signals were each split into three lines to provide the RGB signals driving the two CRT monitors. H- and V-sync signals were split into two lines and transmitted to the two CRT monitors. Gamma correction was performed for each monitor, and the luminance of the images seen by each eye through the matched polarizing filters ranged from 0.03 to 5.65 cd/m² (mean, 2.84 cd/m²). During the recording sessions, the monkey sat in a primate chair with its head secured in
place and facing the CRT monitor(s) at a distance of 80 cm. The field size of the stimulus was 18 × 18 deg, and the screen refresh rate was 75 Hz. Each pixel subtended 0.037 deg and the monitor resolution had 640 × 480 pixels.

In a first preliminary experiment, we examined the dependence of the initial vergence responses on spatial frequency. The visual images seen by each eye were 1D horizontal sinusoidal grating patterns (contrast: 32%) and on any given trial could have one of seven spatial frequencies selected randomly from a lookup table: 0.05, 0.1, 0.2, 0.4, 0.8, 1.6, or 3.2 cycles/deg. The images seen by the two eyes were identical except for a (vertical) phase difference that was one quarter of the wavelength, and this defined the binocular disparity of the stimulus.

In a second experiment, we examined the dependence of the initial vergence responses on the contrast of pure sinusoidal stimuli and the relative contrast of two superimposed sinusoidal stimuli with differing spatial frequencies and disparities of opposite sign. The visual images seen by each eye consisted of 1D horizontal grating patterns that filled the stimulus fields (18 × 18 deg) and could have one of three vertical luminance profiles in any given trial: 1) a sum of two sine waves with the relative spatial frequencies and phase of the third and fifth harmonics of the missing fundamental stimulus, creating a pattern with fundamental frequency, f (termed the 3f5f stimulus), 2) a pure sine wave with a spatial frequency identical to that of the 3f component of the 3f5f stimulus (the 3f stimulus), 3) a pure sine wave with a spatial frequency identical to that of the 5f component of the 3f5f stimulus (the 5f stimulus). Again, the images seen by the two eyes were identical except for a (vertical) phase difference that was one quarter of the wavelength, and this defined the binocular disparity of the stimulus. This meant that when the 3f5f pattern seen by the two eyes had a left-hyper disparity that was one quarter of its (fundamental) wavelength, the 3f component had a right-hyper disparity that was one quarter of its wavelength, and the 5f component had a left-hyper disparity that was one quarter of its wavelength. On any given trial, the 3f components of the 3f5f stimuli could have one of nine contrasts selected randomly from a lookup table: 4, 6, 8, 12, 16, 24, 32, 48, and 64% (Michelson contrast), whereas the contrast of the 5f component was fixed at 16%. The selection of contrasts used for the pure 3f and 5f stimuli matched those used for the 3f and 5f components, respectively, of the 3f5f stimuli. Experiment 1 showed a Gaussian dependence on log spatial frequency and the particular spatial frequencies selected for the 3f and 5f stimuli in experiment 2 (0.6 and 1.0 cycles/deg, respectively) occupied roughly symmetrical locations on either side of the peaks of the Gaussians. The net result was that the 3f and 5f stimuli were of similar efficacy, i.e., when of equal contrast they produced responses of similar amplitude. The positions of the stimuli on the two monitor screens were randomized together from trial to trial at intervals of one quarter of the (fundamental) wavelength of the grating pattern.

The changes in the contrast of the 3f component in this second experiment are accompanied by changes in the overall contrast of the 3f5f pattern. This raises the possibility that contrast normalization (Carandini and Heeger 1994; Carandini et al. 1997; Heeger 1992; Heuer and Britten 2002) contributes to the dependence on the contrast ratio. To address this issue directly we did a control experiment on two monkeys using 3f5f stimuli and varied the contrast ratio while keeping the total contrast fixed at 32% (so that increases in the contrast of 1 component were balanced by decreases in the contrast of the other component). The 3f and 5f components of the 3f5f stimuli could have one of nine contrast ratios: 1/4, 1/3, 1/2, 1/1.5, 1.5, 1.5, 2, 3, and 4. This experiment also included pure 3f and 5f stimuli whose contrasts matched those of the 3f and 5f components, respectively, of the 3f5f stimuli. The entries in the lookup table for the 3f stimuli were 6.6, 8.4, 11.3, 13.6, 17.3, 20.8, 23.1, 25.8, and 27.4%. The entries in the lookup table for the 5f stimuli were 26.6, 25.1, 22.6, 20.5, 17.3, 13.8, 11.5, 8.6, and 6.8%.

Procedures

All aspects of the experimental paradigms were controlled by two PCs, that communicated via Ethernet using the TCP/IP protocol. One of the PCs was running a real-time experimentation software package (REX) developed by Hays et al. (1982) and provided the overall control of the experimental protocol as well as acquiring, displaying, and storing the eye movement data. The other PC was running MATLAB subroutines, using the Psychophysics Toolbox extensions (Brainard 1997; Pelli 1997) and generated the visual stimuli on receiving a start signal from the REX machine.

At the beginning of each trial, fixation points (diameter, 0.8 deg) appeared at the centers of the otherwise uniform gray screens. The animal was required to fixate within 1 deg of these points for 300–600 ms, after which the fixation points disappeared and were immediately replaced by horizontal grating patterns. These patterns were identical at the two eyes except for a phase difference of one-quarter wavelength (vertical binocular disparity), and filled the screens for 200 ms (12 frames), after which the screens became uniform gray with the same mean luminance, signaling the end of the trial. If a saccadic eye movement was detected during the fixation period, the screens were turned off and the trial was repeated later. If no saccades were detected during the stimulus presentation period, the data were stored on a hard disk, and the animal was given a drop of fruit juice; otherwise, the trial was aborted and fluid was withheld. After an intertrial interval of 500–800 ms, the binocular fixation targets reappeared, starting a new trial. Data were collected in several sessions on consecutive days until each condition had been repeated an adequate number of times to permit good resolution of the responses through averaging.

Eye movement recordings and data analyses

Eye movements were measured using the electromagnetic search coil technique (Fuchs and Robinson 1966). Voltage signals separately encoding the horizontal and vertical components of eye position were passed through an analog low-pass filter (–3 dB at 200 Hz) and were digitized to a resolution of 12 bits, sampling at 1 kHz. All data were stored and transferred to another PC for analysis using computer programs based on MATLAB (The MathWorks). The eye position data were smoothed with a four-pole digital Butterworth filter (3 dB at 25 Hz), and eye velocity was derived from the two-point backward difference. Eye acceleration was derived from the two-point backward difference of the eye velocity data and was used to detect small saccades that went undetected during the experiment. Trials with such saccadic intrusions were discarded. At least 33 trials were available for each stimulus condition.

We used the convention that rightward and upward deflections of the stimuli or the eyes were positive. The vertical vergence angle was computed by subtracting the vertical position of the left eye from the vertical position of the right eye. This meant that left-hyper disparity and left-supravergence had positive signs. The initial vergence responses in each stimulus condition were quantified by measuring the changes in the vergence position measures over the 70-ms time periods starting when the mean vergence velocity first exceeded 0.25 deg/s (response-locked measures). The minimum latency of the vergence responses was generally ~70 ms from the appearance of the disparity stimuli so that most of these vergence-response measures were restricted to the initial open-loop period. We also measured the changes in the vergence position measures over the 70-ms time periods starting 70 ms after the onset of the disparity stimuli (stimulus-locked measures), but these measures were solely for comparison with other studies that used this same methodology and will only be considered in the Discussion.
To improve the signal-to-noise ratio, we calculated the difference between the mean initial vergence response to each left-hyper disparity stimulus and that to the corresponding right-hyper disparity stimulus (referred to as LH-RH or RH-LH response measures, respectively). To permit direct comparisons between the data obtained with single and dual grating stimuli, LH-RH measures were used when the data were obtained with 3f5f and 5f stimuli, whereas RH-LH measures were used when the data were obtained with 3f stimuli. Because left-hyper stimuli and left sursumvergence responses were positive in our sign convention, these difference measures were positive when corrective for the 5f stimuli/components and negative when corrective for the 3f stimuli/components. The LH-RH measures were used to examine the dependence on spatial frequency in the first experiment. The same procedure was applied to the mean vergence velocity temporal profiles (referred to as LH-RH or RH-LH vergence velocity profiles).

RESULTS

Pure sine-wave stimuli: dependence on spatial frequency

The first experiment applied one-quarter wavelength disparities to pure sine-wave stimuli and examined the dependence of the initial vergence responses on spatial frequency. The direction of the vergence responses was always as expected of a negative-feedback mechanism operating to eliminate the one-quarter wavelength phase differences, consistent with a sensing mechanism that gives greatest weight to the nearest neighbor binocular matches (Sheliga et al. 2006b). Thus when the sign of the disparity stimulus was defined by the one-quarter wavelength phase differences, left-hyper disparities resulted in left sursumvergence and right-hyper disparities resulted in right sursumvergence, with minimum onset latencies ~70 ms. Figure 1A shows the mean vergence velocity profiles over time for monkey K, indicating that there was generally a large initial transient response lasting 40–50 ms before the velocity declined to a lower level. As the spatial frequency increased from 0.05 to 0.8 cycles/deg (continuous traces), this initial transient increased in amplitude and decreased in latency; further increases in spatial frequency (dotted traces) had the reverse effects, the latency increasing and the amplitude decreasing. These data were quantified by plotting the latency of response onset and the mean LH-RH changes in vergence position (response-locked measures) as functions of spatial frequency (on logarithmic scales): see the black upright triangles plotted in Fig. 1, B and C. The latency on spatial frequency was clearly band-pass for the amplitude measures and U-shaped for the latency estimates, so that the former were close to maximal when the latter were close to minimal. Similar data were obtained from the other three monkeys and these are also included in Fig. 1, B and C. The amplitude data in Fig. 1B were well represented by a Gaussian function with $r^2$ values $\geq 0.958$, peaks $(f_{0})$ at 0.84, 0.67, 0.62, and 0.89 cycles/deg, and SD (or) of 0.64, 0.91, 0.67, and 0.73 log units for monkeys B, C, K, and T, respectively (colored curves). These parameters of the best-fit Gaussian functions were used to derive a low-frequency cut-off $(f_{lo})$ and a high-frequency cut-off $(f_{hi})$, defined as the spatial frequencies at which the tuning curve was half its maximum, using the following expression from Read and Cumming (2003): $a exp(\pm \sigma ln 4)$. The computed values of $f_{0}$ were 0.15, 0.06, 0.10, and 0.12 cycles/deg, and the computed values of $f_{hi}$ were 4.70, 7.95, 3.76, and 5.71 cycles/deg for monkeys B, C, K, and T, respectively. The latency data in Fig. 1C were well described by the expression, $a \log_{10} (b - c) + c$, with $r^2$ values $\geq 0.931$ (colored curves). The mean parameter values $\pm$ SD were $a = 32.58 \pm 3.17$, $b = -0.21 \pm 0.14$, and $c = 72.51 \pm 4.87$ (ms), yielding minimum latencies at a spatial frequency of 0.64 $\pm$ 0.18 cycles/deg (range, 0.38–0.79 cycles/deg).

3f5f stimulus and its components: dependence on contrast and relative contrast

The second experiment examined the initial vergence responses to the 3f5f dual grating stimuli as well as to their two components when each was presented in isolation. Figure 2A shows the mean RH-LH velocity profiles over time obtained
from monkey $K$ with the pure $3f$ stimulus (0.6 cycles/deg), indicating that as the contrast increased (over the limited range examined) the initial transient response increased in amplitude and decreased in latency. The RH-LH response measures for these data show a monotonic dependence on contrast with a slight leveling off when plotted on a logarithmic abscissa: see the black upright triangles in Fig. 2B. The dependence of latency on contrast is roughly linear when plotted on a logarithmic abscissa: see the black upright triangles in Fig. 2C. Similar data were obtained from the other three monkeys and these are also included in Fig. 2, B and C. The response dependence on contrast was characterized with the following expression based on the Naka-Rushton equation: $R_{\text{max}} = C^n / (C^n + C_{50}^n)$, where $R_{\text{max}}$ is the maximum attainable response, $C$ is the contrast, $C_{50}$ is the semisaturation contrast (at which the response has half its maximum value), and $n$ is the exponent that sets the steepness of the curves (Naka and Rushton 1966). This expression has been shown to provide a good fit to the contrast dependence curves of 1) neurons in the LGN, V1, and middle temporal area of monkeys (Albrecht and Hamilton 1982; Albrecht et al. 2002; Heuer and Britten 2002; Sclar et al. 1990), 2) the initial ocular following responses (OFRs) of monkeys and humans (Masson and Castet 2002; Miura et al. 2006; Sheliga et al. 2005), and 3) the initial DVRs of humans (Sheliga et al. 2006b). The curves in Fig. 2B show the best-fit Naka-Rushton functions and clearly provide a good description of the data, with $r^2$ values $\geq 0.967$, exponents ($n$) of 0.86, 1.23, 0.96, and 0.50, and $C_{50}$ values of 5.39, 3.49, 3.83, and 3.70% for monkeys $B$, $C$, $K$, and $T$, respectively. The latency data in Fig. 2C were well described by the expression, $a + b \times c$, with a mean $r^2$ value of 0.923 (colored curves). The mean parameter values $\pm$ SD were $a = 27.76 \pm 5.2$, $b = 0.09 \pm 0.01$, and $c = 70.41 \pm 1.75$ (ms).

Figure 3 shows a subset of the mean vergence velocity temporal profiles obtained with the $3f/5f$ stimuli and with each of their components in isolation for each of the four monkeys. The contrast of the $5f$ stimuli/components was always 16% and the contrast of the $3f$ stimuli/components was 4, 8, 16, 32, or 64%. Each column in Fig. 3 shows the data for a given monkey and each row the data for a given ($3f$) contrast. For the responses to the $5f$ stimuli (green dash-dot traces) and the $3f/5f$ stimuli (blue continuous traces), the LH-RH vergence velocity profiles are plotted, whereas for the responses to the $3f$ stimuli (red dashed traces), RH-LH vergence velocity profiles are plotted. The reason for this difference is that the one-quarter wavelength disparity applied to the $3f$ component of the $3f/5f$ stimulus was opposite in sign to that applied to the $3f/5f$ pattern (and the $5f$ component).

The initial vergence responses elicited by the $3f/5f$ stimuli depended critically on the contrast of the $3f$ components, and we will start by describing the data when the contrast of the $3f$ components was 4%, i.e., one quarter of the contrast of the $5f$ component: see the top row of traces in Fig. 3. The mean vergence velocity profiles obtained with the $3f/5f$ stimuli here were almost identical to those obtained with the pure $5f$ stimuli, even though pure $3f$ stimuli with a contrast of 4% always generated robust responses. However, the initial 10–20 ms of the profiles predicted by the vector sum of the responses to the pure $3f$ and $5f$ stimuli (cyan dotted traces) closely follow the response profiles obtained with the $5f$ stimuli alone. This indicates that, initially, the $5f$ components dominated the responses to the $3f/5f$ stimuli and were unopposed by $3f$ components because of a latency difference: the initial responses to the low-contrast $3f$ stimuli had slightly longer latencies—and/or were much weaker—than the initial responses to the higher-contrast $5f$ stimuli. That the later responses to the $3f/5f$ stimuli exceed the summed responses to the pure $3f$ and $5f$ stimuli—even though the responses to the former were now significant—indicates that the responses to the $3f$ components were being actively suppressed. That these later responses to the $3f/5f$ stimuli often come close to matching the responses to the pure $5f$ stimuli indicates that this suppression often blocked all responses to the $3f$ component: winner-take-all (WTA).

The opposite situation prevailed when the $3f$ components of the $3f/5f$ stimuli had a contrast of 64%, i.e., had four times the contrast of the $5f$ components: see the bottom row in Fig. 3. In
this case, for the first 5–10 ms of the responses, the response profiles obtained with the \(3f5f\) stimuli closely follow those obtained with the pure \(3f\) stimuli as well as the profiles predicted by the vector sum of the responses to the pure \(3f\) and \(5f\) stimuli. Thus initially, the component of higher contrast once more dominates the responses to the \(3f5f\) stimuli because that component has a slightly shorter latency and/or the initial response to it was very weak. Once more, the subsequent responses to the \(3f5f\) stimuli exceeded the summed responses to the pure \(3f\) and \(5f\) stimuli indicating that the responses to the component of lower contrast were being suppressed and, once again, this suppression was often almost total: WTA.

When the contrasts of the two components of the \(3f5f\) stimuli were more similar (Fig. 3, 2nd and 4th rows), the responses to the combined stimuli were still greater than the sum of the responses to the pure \(3f\) and \(5f\) stimuli but were now often less than to the component stimulus of higher contrast when applied alone. Thus the responses to the component with the lower contrast were now only partially suppressed. When the two components of the \(3f5f\) stimulus had the same contrasts (Fig. 3, middle row), the responses to the combined stimuli were often irregular and not always weak, at first sometimes veering toward the \(3f\) data and later veering toward the \(5f\) data.

In summary, it is evident that when the contrast of one component of the \(3f5f\) stimulus exceeded that of the other component by a sufficient margin then vergence responses were dominated by the one of higher contrast. However, two factors seem to be involved here: the first was the longer latency of the low-contrast component, which meant that the higher contrast component was initially dominant because it was unopposed, and the second was active suppression of the responses to the component of lower contrast, i.e., there was a nonlinear interaction. We estimated the impact of these nonlinear interactions using response measures that, for each data set, were all synchronized to the onset of the mean responses to either the pure \(3f\) stimulus or the pure \(5f\) stimulus, whichever

---

**FIG. 3.** The initial vertical vergence responses to \(5f\) stimuli (green dash-dot traces), \(3f\) stimuli (red dashed traces), and \(3f5f\) stimuli (dark blue thick traces), together with the responses predicted by the vector sum of the responses to the \(5f\) and \(3f\) stimuli (light blue dotted traces): mean vergence velocity temporal profiles of 4 monkeys. Each column shows data for one monkey and each row shows data for \(3f\) stimuli/components of one contrast (indicated by labels in leftmost column). For the responses to the pure \(5f\) and to the \(3f5f\) stimuli, the LH-RH mean eye velocity profiles (in deg/s) were used, whereas for the responses to pure \(3f\) stimuli, RH-LH mean eye velocity profiles (in deg/s) were used. Horizontal dashed lines, zero eye velocity. The abscissas denote the time elapsed (in ms) since stimulus onset.
was later. For example, the response measures for the data set of monkey K in the top row of Fig. 3 were all synchronized to the onset of the response to the 3f stimulus, which clearly had a longer latency than the response to the 5f stimulus and determined the onset of the nonlinear interactions. Thus in this case, the response measures were based on the changes in vergence position during the 70-ms time period starting when the vergence velocity response to the 3f stimulus first exceeded 0.25 deg/s. Figure 4 plots the resultant mean LH-RH response measures for the data obtained with the 3f5f (dark blue squares) and 5f (green triangles) stimuli, together with the mean RH-LH response measures for the data obtained with the 3f stimuli (red circles), as a function of the contrast of the 3f stimuli/components for each of the four monkeys. As expected from the vergence velocity profiles in Fig. 3, when the contrast of the 3f component of the 3f5f stimulus was at its lowest, the response measures for the dual grating data were very similar to those for the pure 3f data, and when the contrast of the 3f component was at its highest, the response measures for the dual grating data were very similar to those for the pure 3f data. The clear suggestion here is that the component with the lower contrast had little or no influence on the responses to the dual gratings. Of course, when the contrasts of the two components were more similar, responses were intermediate as the influence of one component waxed while the other waned. These data clearly deviate substantially from the vector-sum predictions, which are plotted in light blue line in Fig. 4.

To quantify the transition from dominance by one component to dominance by the other, we computed the response ratio to the linear (i.e., vector-sum) predictions, which are shown in light blue lines in Fig. 5, as well as the vector-average prediction (for which the response ratio is always 0.5).

To obtain a quantitative estimate of the abruptness of the transitions in the response ratios in Fig. 5, the data for each monkey were fitted with a cumulative Gaussian function using a least squares criterion: see the dark blue smooth curves in these graphs. The $r^2$ values for these fits ranged from 0.996 to 0.999, indicating that they provide a very adequate description of these data. The amplitudes of the cumulative Gaussians were always slightly less than unity (mean, 0.91) and their SD ranged from 0.10 to 0.25 (mean, 0.17). We also obtained a quantitative estimate of how different the contrasts of the two components of the 3f5f stimuli had to be for one of the components to effectively lose its influence. For this we used the cumulative Gaussian to determine a transition zone, which we defined as the range of contrast ratios over which the response ratio ranged from 0.05 to 0.95. On average—based on the mean cumulative Gaussian for all data from all monkeys—this transition zone extended from 0.65 to 2.31. Thus if the two sine waves were of similar efficacy when of equal contrast and applied singly, when combined as in our dual gratings, on average, a 2.3-fold difference in contrast sufficed for the one with the lower contrast to almost totally lose its influence (based on the average of the 5% values and the reciprocals of the 95% values).

![Graphs](image-url)

**FIG. 4.** The initial vertical vergence responses to 5f stimuli (green triangles), 3f stimuli (red circles), and 3f5f stimuli (dark blue squares), together with the responses predicted by the vector sum of the responses to the 5f and 3f stimuli (light blue line): dependence on the contrast of the 3f stimuli/components for each of the four monkeys. LH-RH measures (in deg) were used for the responses to 5f and 3f5f stimuli, whereas RH-LH measures (in deg) were used for the responses to 3f stimuli. Note all measures in any given response set (such as those for any one of the 20 plots in Fig. 3) were given by the mean changes in vergence position (in deg) during the 70-ms time period starting with the onset of the response to the 5f or 3f stimulus, whichever took longer to 1st exceed the onset threshold of 0.25 deg/s. Note that the abscissas all have a logarithmic scale. The discontinuities (indicated in dotted line) in the otherwise continuous lines linking the data points denote the transition from measures time-locked to the responses to the 3f stimuli (generally, when the contrast of the latter was small) and the responses to the 5f stimuli (generally, when the contrast of the 3f stimulus was high). Error bars, 95% CIs.
Control experiment in which the total contrast of the 3f5f pattern was kept constant

We did a control experiment on monkeys B and T using 3f5f stimuli and varied the contrast ratio while keeping the total contrast constant at 32% (see Methods). The data were very similar in all essentials to those obtained in the main experiment using 3f5f stimuli. Thus the plots of response ratio against contrast ratio were well fit by cumulative Gaussians ($r^2 \geq 0.99$) whose amplitudes (0.98 for monkey B, 0.93 for monkey T) and SDs (0.25 for monkey B, 0.14 for monkey T) were not significantly different from those in the main experiment ($P > 0.05$, t-test).

DISCUSSION

The sudden appearance of 1D horizontal sine-wave grating patterns with vertical disparities elicited vertical vergence responses in all monkeys at short latencies (minimum, ~70 ms), with the direction expected of a negative feedback mechanism operating to eliminate the one-quarter wavelength phase difference, consistent with a sensing mechanism that gives greatest weight to the nearest neighbor binocular matches. The change in vergence position measured over the initial open-loop period (the 70-ms interval starting with response onset) showed a Gaussian dependence on log spatial frequency (on average: peak = 0.75 cycles/deg; SD = 0.74; $r^2 = 0.980$) and a monotonic dependence on log stimulus contrast with a gradual saturation that was well fit by the Naka-Rushton equation (on average: $n = 0.89$; $C_{50} = 4.1\%$; $r^2 = 0.978$). The initial vergence responses elicited by the 3f5f stimulus—two superimposed sine waves that had spatial frequencies in the ratio 3:5 and binocular disparities of opposite sign—showed a nonlinear dependence on the relative contrasts of the two competing components: on average, when the contrast of one was 2.3 times greater than that of the other, the one with the lower contrast became almost ineffective as though suppressed and the vergence response was almost entirely determined by the sine wave of higher contrast (WTA). A control experiment, in which the overall contrast of the 3f5f stimulus was kept constant, indicated that this nonlinear behavior was not caused by a nonspecific contrast-normalization process.

Energy-based or feature-based disparity detection?

The strong dependence of the initial vergence responses on the relative contrasts of the two components of the 3f5f gratings occurred despite the fact that the binocular disparity of the overall stimulus pattern and its features (peaks and troughs) was always the same—one quarter of the wavelength of the fundamental—throughout the experiment. In fact, when the 3f component had the higher contrast then the initial vergence responses worked to decrease the binocular disparity of that component but to increase the binocular disparity of the 3f5f pattern. This indicates that the Fourier composition of the disparity stimulus was the critical determinant of the direction and amplitude of the initial vergence response. This is consistent with mediation by a disparity sensing mechanism that matches spatially filtered versions of the monocular inputs—rather than the raw images, overall pattern, or features—and senses the first-order disparity energy in the stimulus as postulated for the disparity selective complex cells in striate cortex (Ohzawa et al. 1990). These findings and conclusions are essentially the same as those in recent studies on humans in which broadband and dual-grating disparity stimuli were used to elicit horizontal and vertical vergence eye movements at short latency (Sheliga et al. 2006b). The apparent lack of any major second-order influence in our vertical vergence data are also consistent with the findings of Stevenson (2002) in humans.

Nonlinear interactions between the neural mechanisms sensing competing disparities

The nonlinear dependence of the initial vergence responses on the relative contrasts of the two components of the 3f5f disparity stimulus also mirrors recent findings on humans (Sheliga et al. 2007b), which were well described by a contrast-weighted-average (CWA) model with two free parameters (cf. Krompenhoek and Wiergerink 1998; McGowan et al. 1998; Port and Wurtz 2003; Recanzone and Wurtz 1999)

\[
R_{3f5f} = \frac{C_{3f}}{C_{5f} + C_{3f}} R_{3f} + \frac{C_{5f}}{C_{5f} + C_{3f}} R_{5f}
\]

(1)

where $C_{3f}$ and $C_{5f}$ are the contrast levels of the 3f and 5f components, respectively; the two exponents, $n_{3f}$ and $n_{5f}$.
which are the only free parameters, reflect the efficacies of the 3f and 5f components, respectively, of the given 3f5f stimulus and were determined by the optimization procedure. We fitted this CWA model to the differential measures obtained in the present study (LH-RH response measures for R3f5f and R5f, RH-LH response measures for R3f3f), and the least-squares best-fit values of the n3f and n5f parameters, together with the r² values indicating the goodness of the fits are listed in Table 1. For all monkeys, the CWA model provided an excellent fit to the data over the full range of contrasts of the 3f component, the r² values ranging from 0.994 to 0.999.

It is common to invoke mutual inhibition between the neural mechanisms sensing the two competing stimuli to account for nonlinear interactions like those here that can result in complete dominance by the stimulus of higher contrast, i.e., WTA (Dearworth et al. 2005; Ferrera 2000; Ferrera and Lisberger 1995, 1997; Kodaka et al. 2007; Recanzone and Wurtz 1999; Sheliga et al. 2006c, 2007b). The article of Dearworth et al. (2005) is of particular interest because it describes the vergence responses to two competing disparity targets. The exponents in the CWA model, which provide an estimate of the strengths of this postulated mutual inhibition, averaged 3.88 (n3f) and, 4.09 (n5f), suggesting that they were of very similar efficacy. For the inhibition generated by the higher contrast component to totally suppress even the earliest vergence responses generated by the lower contrast component, the former must have the shorter latency. That this condition was met in the present case is evident from the inverse dependence of latency on contrast in Fig. 2C. In fact, this meant that when there was an appreciable difference in the contrasts of the two components the earliest responses to the 3f5f stimulus were the same as to the component of higher contrast alone, initially because of the latency difference and only later because of nonlinear interactions: see the top and bottom rows in Fig. 3. This inverse dependence of latency on contrast has often been reported in studies of neuronal responses in striate cortex (Albrecht 1995; Albrecht et al. 2002; Carandini and Heeger 1994; Carandini et al. 1997; Gawne et al. 1996).

A major functional consequence of these nonlinear interactions is that the initial DVR will tend to discriminate in favor of those disparate images that have the higher contrast. This suppression of low-contrast signals would provide some noise immunity. Recent studies on the initial DVRs of humans (Sheliga et al. 2007b), as well as the initial OFRs of both humans (Sheliga et al. 2006c) and monkeys (Matsuura et al. 2008), showed similar nonlinear interactions and argued that one consequence of them might be that objects in the plane of fixation would tend to be preferred over objects in other depth planes: because of accommodation, the retinal images of objects in the plane of fixation tend to be better focused—and hence tend to have higher contrasts—than those of objects in other depth planes. This would also be in line with the finding (in humans only, to date) that the vertical vergence mechanism responds best to vertical disparities when horizontal disparity is minimal and vice versa (Yang et al. 2003). However, recent studies on the horizontal OFR of humans indicate that robust initial responses can be elicited by moving stimuli that occupy a single horizontal band that extends the full width of the display but is only 1 degree high (Sheliga et al. 2008) and that when the two competing components of the 3f5f stimulus each occupies such a strip then the nonlinear dependence on their relative contrasts is seen only when they are overlapping: the nonlinear interaction was eliminated when the competing components were separated by a vertical gap of 1 degree or more (Sheliga et al. 2007a). If this were also the case for the initial DVRs under study here then the supposed benefit of the WTA mechanism and the postulated mutual inhibition would be very local.

### Is the monkey a good animal model?

Earlier work on the initial vergence responses to large-field disparity stimuli such as random dots indicated that monkeys and humans share many fundamental properties (e.g., ultrashort latency, similar dependence on a prior saccade, and the magnitude of the disparity) and this led to the suggestion that the monkey might provide a good animal model for the human (Miles 1998). The current data indicate that this similarity between the initial vergence responses of monkeys and humans also extends to their dependence on spatial frequency and contrast when simple 1D sinusoidal stimuli are used, as well as to their dependence on the Fourier composition and relative contrast when dual gratings like the 3f5f stimulus are used. Indeed, the simple mathematical functions that described the present data from monkeys so successfully—the Gaussian function to describe the dependence on (log) spatial frequency, the Naka-Rushton equation to describe the dependence on contrast, the cumulative Gaussian (and the CWA model) to describe the nonlinear dependence on relative contrast—had all been similarly successful in describing human DVRs. However, one problem with attempting a direct quantitative comparison of the best-fit parameters is that the published data for humans all relied on response measures that were synchronized to the onset of the stimulus, whereas the measures in this study on the monkey were all synchronized to the onset of the response (because of latency effects). We addressed this issue by reanalyzing the present data using a stimulus-locked window comparable with that used on the human data, i.e., we measured the changes in vergence position over the 70-ms time period starting 70 ms after stimulus onset. Importantly, these stimulus-locked measures were still well described by the same mathematical functions, and the best-fit parameters, averaged for all four monkeys for each of the two sets of measures, are summarized in the first two columns of Table 2. The impact of using the stimulus-locked response window was quite minor; the only significant difference was a small increase in the C50 parameter (t-test,  P < 0.05). The equivalent human data from Sheliga et al. (2006b, 2007b), which are listed in the third column of Table 2, are significantly different from the monkey data in a number of respects (Table 2, *): the spatial-frequency dependence peaks at a much lower frequency (0.36 vs. 0.72 cycles/deg) and has a narrower bandwidth (indicated by σ), mainly because the high-frequency cut-off (fhi) is lower; the

### Table 1. Best-fit exponents of the CWA model

<table>
<thead>
<tr>
<th>Monkey</th>
<th>n3f</th>
<th>n5f</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>2.36</td>
<td>2.50</td>
<td>0.994</td>
</tr>
<tr>
<td>C</td>
<td>4.13</td>
<td>4.47</td>
<td>0.994</td>
</tr>
<tr>
<td>K</td>
<td>3.30</td>
<td>3.29</td>
<td>0.999</td>
</tr>
<tr>
<td>T</td>
<td>5.71</td>
<td>6.08</td>
<td>0.996</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>3.88 ± 1.42</td>
<td>4.09 ± 1.56</td>
<td>0.996 ± 0.002</td>
</tr>
</tbody>
</table>

Downloaded from http://jn.physiology.org/ by 10.220.33.4 on June 26, 2017
contrast dependence has a higher $C_{50}$, by a factor of two; the dependence on the relative contrast of the competing components of the $3f/5f$ stimuli shows a transition zone that is shifted toward slightly lower contrast ratios in humans, but its width—as indicated by the SD—and the exponents, $n_{5f}$ and $n_{3f}$, are very similar, indicating that the strengths of the postulated mutual inhibition are comparable. These differences might be caused by some quantitative differences in the experimental conditions, involving the size of stimuli ($18 \times 18$ deg in the present study, $46 \times 35$ deg in the previous study on humans), the viewing distance (80 cm in the present study, 47.1 cm in the previous study), and the mean luminance level (2.84 cd/m$^2$ in the present study, 38.7 cd/m$^2$ in the previous study). The general picture that emerges here is that, despite quantitative differences, the initial DVRs of the monkey display all the major features of the human responses and can be described by the same mathematical functions.

**Neuronal mediation**

All studies concerned with the neuronal mediation of the initial DVR were done on monkeys and considered only the responses to horizontal disparities. The published behavioral studies on the initial DVRs of humans have generally included the responses to both horizontal and vertical disparities and indicate that differences between them are relatively minor—quantitative rather than qualitative. For example, the shapes of the disparity tuning curves with large-field random dot stimuli are very similar (Busettini et al. 2001), showing a very similar sensitivity to orthogonal offsets (Yang et al. 2003), both rely on first-order disparity energy and show very similar dependencies on spatial frequency and contrast with 1D sine-wave grating stimuli (Sheliga et al. 2006b), and also show very similar nonlinear dependencies on the relative contrast of two sine waves with competing disparities, (Sheliga et al. 2007b). This suggests that the neuronal mechanisms mediating the initial horizontal and vertical DVRs of humans are very similar, although it says nothing about where those mechanisms are located in the brain. The middle temporal and medial superior temporal areas of the cortex (MT/MST) seem to play a critical role in the generation of the earliest horizontal DVRs in macaques; bilateral lesions of the MT/MST area result in major impairments of these eye movements (Takemura et al. 2001) and single unit studies indicate that the summed activity of the disparity-selective neurons in MST encodes the magnitude, direction, and time course of these eye movements, even reflecting idiosyncratic differences between animals (Takemura et al. 2001). In addition, reversible inactivation of areas V2 and V3 in the macaque by cooling selectively attenuates these eye movements as well as the disparity-selective responses in area MT (Ponce et al. 2008), a region that contains cells sensitive to vertical disparity (Maunsell and Van Essen 1983b) and has a strong projection to MST (Maunsell and Van Essen 1983a; Ungerleider and Desimone 1986). However, many of the spatiotemporal properties of the earliest DVRs like those described in the present study have been attributed to disparity selective neurons at earlier stages of cortical processing, perhaps as early as striate cortex (Sheliga et al. 2006b, 2007b), where many neurons are sensitive to vertical disparity (Durand et al. 2002, 2007) and where the properties of many disparity-selective neurons are well captured by the disparity-energy model (Fleet et al. 1996; Ohzawa et al. 1990; Parker and Cumming 2001; Qian 1994; Read and Cumming 2003; Read et al. 2002). Of possible relevance here is the finding that horizontal disparities applied to dense anticorrelated random dot patterns elicit horizontal DVRs at short latency that are in the reverse direction of those elicited when disparities are applied to correlated stimuli (Masson et al. 1997; Takemura et al. 2001), a feature of many disparity-selective neurons in striate cortex (Cumming and Parker 1997; Ohzawa et al. 1990) and MST (Takemura et al. 2001). Furthermore, such dense anticorrelated patterns do not support stereopsis (Cogan et al. 1993; Cumming and Parker 1997; Cumming et al. 1998; Masson et al. 1997), consistent with the idea that the earliest horizontal DVRs derive their sensory inputs from an early

**TABLE 2. Parameters of the best-fit functions in the present study on monkeys and in the studies of Sheliga et al. (2006b, 2007b) on humans**

<table>
<thead>
<tr>
<th>Monkey (n = 4)</th>
<th>Human (n = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response-Locked</td>
<td>Stimulus-Locked</td>
</tr>
<tr>
<td>$f_c$, cycles/deg</td>
<td>$0.75 \pm 0.13^*$</td>
</tr>
<tr>
<td>$\sigma_c$, cycles/deg</td>
<td>$0.74 \pm 0.12^*$</td>
</tr>
<tr>
<td>$f_r$, cycles/deg</td>
<td>$0.11 \pm 0.04$</td>
</tr>
<tr>
<td>$r_f^2$</td>
<td>$5.71 \pm 1.86^*$</td>
</tr>
<tr>
<td>n</td>
<td>$0.980 \pm 0.019$</td>
</tr>
<tr>
<td>Contrast (Naka-Rushton)</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>$0.89 \pm 0.30$</td>
</tr>
<tr>
<td>$C_{50}$</td>
<td>$4.10 \pm 0.87^*$</td>
</tr>
<tr>
<td>$r_f^2$</td>
<td>$0.978 \pm 0.014$</td>
</tr>
<tr>
<td>Relative contrast (cumulative Gaussian)</td>
<td></td>
</tr>
<tr>
<td>SD, log unit</td>
<td>$0.17 \pm 0.07$</td>
</tr>
<tr>
<td>5%</td>
<td>$2.31 \pm 0.65$</td>
</tr>
<tr>
<td>95%</td>
<td>$0.65 \pm 0.19^*$</td>
</tr>
<tr>
<td>$r_f^2$</td>
<td>$0.997 \pm 0.002$</td>
</tr>
<tr>
<td>Relative contrast (CWA model)</td>
<td></td>
</tr>
<tr>
<td>$n_{5f}$</td>
<td>$3.88 \pm 1.42$</td>
</tr>
<tr>
<td>$n_{3f}$</td>
<td>$4.09 \pm 1.56$</td>
</tr>
<tr>
<td>$r_f^2$</td>
<td>$0.996 \pm 0.003$</td>
</tr>
</tbody>
</table>

Differences that are significant on the t-test ($P < 0.05$). *Monkey data vs. human data. †Response-locked vs. stimulus-locked measures of monkey.
One of a family of oculomotor reflexes

The DVR is one of three types of oculomotor response that can be elicited at short latencies in both monkeys and humans by large-field visual stimuli; the other two are the OFR, which generates version eye movements in response to linear motion orthogonal to the line of sight (Gellman et al. 1990; Miles et al. 1986), and the radial-flow vergence response (RFVR), which generates vergence eye movements in response to radial optic flow (Busettini et al. 1997; Inoue et al. 1998; Yang et al. 1999). There is strong evidence from unit recordings and lesion studies in monkeys that the OFR and RFVR are also mediated by MT/MST (Kawano et al. 2000; Takemura and Kawano 2002; Takemura et al. 2007). Earlier studies using random dot stimuli led to the idea that the three types of oculomotor response constitute a family of complementary oculomotor reflexes that work in harmony to stabilize the gaze of the translating observer, whether human or monkey: see Miles (1998) and Miles et al. (2004) for review of this earlier work. More recent studies using stimuli like those in the present experiments—1D sine-wave gratings and dual (or broadband) gratings with competing harmonics—have concentrated on the fundamental spatiotemporal properties and have uncovered a remarkable degree of uniformity among the three very different kinds of eye movements in humans and monkeys, e.g., all respond to the local energy in the stimulus rendering them sensitive to its harmonic structure (rather than its raw features) and opponent stimuli reveal the existence of competitive interactions that can result in a WTA situation. In fact, in both humans and monkeys, the same simple mathematical functions as were used to describe the DVRs in the present study—each with only a couple of free parameters—can be used to describe the fundamental properties for all three reflexes (Kodaka et al. 2007; Matsuura et al. 2008; Miura et al. 2006; Sheliga et al. 2005, 2006b,c, 2007a,b, 2008). Interestingly, recent studies on humans indicate that, although the RFVR and OFR respond to very different patterns of global motion—radial versus linear—and result in very different types of eye movement—vergence versus version—they have very similar local spatiotemporal properties (even quantitatively) as though mediated by the same local motion-energy detectors, which it was suggested were in the striate cortex where such detectors first occur in the cortical visual projection (Kodaka et al. 2007). In addition, the OFRs and RFVRs of humans to apparent-motion stimuli consist of “two-frame movies” could be reversed by a brief interstimulus interval, consistent with the idea that the visual input to the underlying motion detectors has a biphasic temporal impulse response (Kodaka et al. 2007; Sheliga et al. 2006a). These observations reinforce the idea that the three reflex eye movements provide a powerful probe with which to access the earliest cortical processing of motion and disparity in monkeys and humans.

Acknowledgments

The authors thank Dr. R. Hayashi for help in the course of experiments.

Grants

This work was supported by Grant-in-Aid for Scientific Research 16GS0312 from the Japan Society for the Promotion of Science, the Japan-U.S. Brain Research Cooperative Program, and the Intramural Program of the National Eye Institute at the National Institutes of Health.

References


