Do Visual Cues Contribute to the Neural Estimate of Viewing Distance Used by the Oculomotor System?

Min Wei, Gregory C. DeAngelis, and Dora E. Angelaki
Department of Neurobiology, Washington University School of Medicine, St. Louis, Missouri 63110

Perceived shape and depth judgments that require knowledge of viewing distance are strongly influenced by both vergence angle and the pattern of vertical disparities across large visual fields. On the basis of this established contribution of visual cues to the neural estimate of viewing distance, we hypothesized that the oculomotor system would also make use of high-level visual cues to distance. To address this hypothesis, we investigated how compensatory eye movements during whole-body translation scale with viewing distance. Monkeys viewed large-field (85° × 68°) random-dot stereograms that were rear projected onto a fixed screen and simulated either a textured wall or pyramid at different viewing distances. In these stereograms, we independently manipulated vergence angle, horizontal and vertical disparity gradients, relative horizontal disparities, and textural cues to viewing distance. For comparison, random-dot patterns were also projected onto a moveable screen placed at different physical distances from the animal. Several cycles of left–right sinusoidal motion of the monkey at 5 Hz were interleaved with several cycles of motion in darkness, and the relationship between eye movement responses and viewing distance was quantified. As expected from previous work, the amplitude of compensatory eye movements depended strongly on vergence angle. Although visual cues to distance had a statistically significant effect on eye movements, these effects were weaker than the effect of vergence angle. We conclude that sensory and motor systems do not share a common neural estimate of viewing distance and that the oculomotor system relies far less on visual cues than the perceptual system.

Key words: eye movement; binocular; vestibular; vestibuloocular; vergence; stereopsis; distance perception

Introduction
Accurate reconstruction of the three-dimensional (3D) layout of the environment is vital for visually guided behaviors. Although there are many useful cues to depth in natural images, binocular disparities alone can provide a vivid, quantitative impression of 3D structure (Julesz, 1971; Howard and Rogers, 1995). Binocular disparities arise as a result of the lateral separation of the eyes and generally have both horizontal and vertical components. Horizontal disparities provide information about the depth of objects relative to the plane of fixation. Thus, an object at a fixed egocentric location in 3D space will have different horizontal disparities for different viewing distances. A reliable internal (neural) estimate of viewing distance is therefore essential to compute egocentric depth from binocular disparities.

There are two main candidates for the source of a neural estimate of viewing distance: vergence angle and the pattern of vertical disparities across the visual field (Mayhew and Longuet-Higgins, 1982; Bishop, 1989; Cumming et al., 1991). The former could be derived from proprioceptive or efference copy signals (Mays, 1984; Mays and Porter, 1984; Mays and Gamlin, 1995; Gamlin, 1999). The latter is likely to be computed by disparity-sensitive neurons with large receptive fields, although the location of such neurons in the brain is currently unknown. Vertical disparities (defined here as differences in Fick elevation angles) arise whenever an object is closer to one eye than the other, increase monotonically with eccentricity, and decrease in proportion to viewing distance (Bishop, 1989). Although early psychophysical studies argued that vergence angle, but not vertical disparities, contributed to the scaling of horizontal disparities (Cumming et al., 1991; Sobel and Collett, 1991), more recent studies have shown that the contribution of vertical disparities can equal, or even exceed, the contribution of vergence angle when the field of view is large (Rogers and Bradshaw, 1993, 1995; Bradshaw et al., 1996). It should be noted that other visual cues can also contribute to one’s estimate of viewing distance, including the size of familiar objects (O’Leary and Wallach, 1980; Predebon, 1993), texture variations (Gibson, 1950; Sedgwick, 1986; Johnston et al., 1993), the range of disparities in a scene (Glennerster et al., 1998), and eye movements between objects at different depths (Brenner and van Damme, 1998). Thus, in the scaling of horizontal disparities for depth perception, both vergence angle and a variety of visual cues contribute to the neural estimate of viewing distance.

Considerably less is known about the roles that visual and motor cues play in estimating viewing distance for motor behaviors such as eye movements. Two possibilities exist. First, the motor system could use the same distance cues that are used for depth perception. In this case, both the vertical disparity field and

Received April 15, 2003; revised June 10, 2003; accepted June 13, 2003.
This work was supported by National Institutes of Health Grants EY12814, EY13646, and DCO4266 and the McDonnell Center for Higher Brain Function. We are extremely grateful to Michael Morgan who derived the equations that we used to generate our visual stimuli. We also thank Bruce Cumming, Larry Snyder, and Takanori Uka for valuable comments on this manuscript.
Correspondence should be addressed to Dr. Dora Angelaki, Department of Anatomy and Neurobiology, Box 8108, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110. E-mail: angelaki@pcg.wustl.edu.
Copyright © 2003 Society for Neuroscience 0270-6474/03/238340-11$15.00/0
the vergence angle would be expected to contribute to the scaling of motor responses by viewing distance. Alternatively, the motor system could mainly rely on vergence angle and accommodation for estimating viewing distance and could ignore both monocular and binocular visual cues.

The present study was designed to distinguish between these two possibilities. The simplest motor response whose amplitude depends on viewing distance is the translational vestibulo-ocular reflex (TVOR) (Schwarz et al., 1989; Paige and Tomko, 1991). Thus, we studied compensatory eye movements elicited during lateral head motion while we systematically varied different cues to viewing distance. In one condition, images were projected onto a screen that was physically placed at different distances from the monkey, thus providing congruent visual and motor cues to viewing distance. In other conditions, stereoscopic displays were presented to simulate patterns at different viewing distances; these simulated conditions allowed us to dissociate the effects of visual and motor cues on eye movement responses. In a third set of experiments, we examined the contribution of relative horizontal disparities to TVOR amplitude, because relative disparities are known to be critical for depth perception (Westheimer, 1979; Regan et al., 1986; Prince et al., 2000). We find that visual cues (vertical disparities, texture, and relative horizontal disparities) have little or no role in scaling eye movement amplitude with viewing distance, thus indicating that the oculomotor system relies on different cues to viewing distance than the perceptual system.

Parts of this work have been published previously in abstract form (Wei et al., 2002).

Materials and Methods

Animals and experimental setup

Two rhesus monkeys (Macaca mulatta), weighing 5–6 kg, were trained to fixate a central target presented within a random-dot field displayed on a tangent screen. The animals were chronically prepared with skull bolts to restrain the head and were implanted with scleral coils in each eye for binocular eye movement recordings using the magnetic search coil technique. During each experiment, the monkey was comfortably seated in a primate chair that was placed in the center of a three-field magnetic coil system (CNC Engineering, Seattle, WA) and was located on top of a motion simulator (Neurokinetics, Pittsburgh, PA) that could translate laterally. Lateral oscillations of the monkey elicited compensatory, stabilizing eye movements (TVOR) while the monkey fixated a central target at different real or simulated distances. Stimulus presentation and data acquisition used custom-written scripts in the Spike2 software that controlled data acquisition hardware (CED Power 1041; Cambridge Electronic Design, Cambridge, UK). Binocular eye and head movement signals were filtered (six-pole Bessel; DC, 200 Hz), digitized at 833 Hz, and stored for offline analyses. All surgeries, animal treatment, training, and handling were in accordance with National Institutes of Health and institutional guidelines.

Visual stimuli and task

Visual stimuli were random-dot stereograms that were generated by a dedicated computer and rear projected using a stereoscopic projector (100 Hz vertical refresh rate, 1280 × 1024 resolution; Mirage 2000; Christie Digital Systems, Cypress, CA) onto a tangent screen placed in front of the monkey. Stereograms were programmed using OpenGL and were generated by an OpenGL accelerator board (Oxygen GVX1 Pro; 3Dlabs, Milpitas, CA). To simulate the patterns of retinal disparities associated with a plane of dots at different distances from the head, two sets of dots (red and green) were drawn on the screen as animals viewed the display through red–green filters (Kodak Wratten filters 29 and 61; Eastman Kodak, Rochester, NY). Red dots were viewed by the left eye, and green dots were viewed by the right eye. The colored filters were mounted in goggles that were placed in front of the animal’s eyes. The goggles had a rectangular opening for each eye that restricted the field of view to 85 × 68° (horizontal × vertical), and random-dot patterns always filled the entire field of view. Thus, the retinal subtense was held constant for all visual stimuli presented in this study, regardless of whether they were presented on the fixed screen at 32 cm or, in some experiments, on a moveable screen at different distances.

Each trial began with the onset of lateral oscillatory motion, which gradually increased to the desired amplitude (5 Hz, ±0.25 g, 0.5 cm amplitude peak-to-peak) over a period of ~1 sec. At this point, the fixation point appeared at the center of the screen. In some trials, the fixation point was presented with zero disparity; in other trials, the fixation point had a nonzero disparity and the animal was required to converge accurately on the target. Separate electronic windows were used to capture left-eye position, right-eye position, and vergence posture. Once the monkey acquired the fixation target, a full-field random-dot stereogram appeared for 3 sec, depicting either a frontoparallel plane of dots or a pyramid at some distance from the observer (for details, see Experimental protocols). The vergence window was 1° in diameter, and the left and right eye windows were 5° wide during this period. Next, the dots and fixation point were extinguished, and the monkey was required to maintain fixation for 1.5 sec in the dark (the vergence window was increased to 5° during this period of time). This allowed us to compute the amplitude of compensatory eye movements during visual stimulation and in darkness. Animals were rewarded twice during each trial, 1.5 sec after random-dot pattern onset and a second time at the end of the trial, as long as eye position stayed within the specified windows. Although fairly large fixation windows were used in the experiments, the SD of eye position ranged from 0.2 to 0.6° in all stimulus conditions, indicating that the monkeys were fixating well.

Experimental protocols

Three sets of experiments were performed in separate blocks of trials, as described below.

“Movable screen” experiments

In these experiments, identical left-eye and right-eye random-dot patterns were projected onto the surface of a moveable screen. The screen was placed at four different distances (12, 18, 32, and 52 cm) from the monkey, and the projector was kept at a constant distance from the back of the screen such that the screen image was identical at all viewing distances (whereas the retinal image varied). No binocular disparities were applied to the random-dot patterns on the moveable screen, such that the dots were always imaged in the plane of the moveable screen. The monkey was required to maintain fixation on a small spot that was also imaged in the plane of the screen. Thus, in this condition, visual (disparity field and texture) and motor cues (vergence and accommodation) were congruent and specified the same viewing distance. Although the retinal subtense of the screen varied with its distance from the monkey, the monkey’s goggles restricted the field of view to be constant at all screen distances. Thus, the peripheral portions of the screen were occluded from view at nearer viewing distances.

Random-dot stereograms with congruent or incongruent vergence, texture, and disparity cues

A frontoparallel “wall” of dots at five different viewing distances (12, 18, 32, 52, and 102 cm) was simulated by projecting random-dot stereograms on the tangent screen, which was fixed at 32 cm. Within a single block of trials, the following five different cue conditions were randomly interleaved.

Disparity Only condition. Random-dot stereograms presented the exact pattern of horizontal and vertical disparities that would be associated with fixation on a plane of dots at the five different simulated distances (see Fig. 2A). Dot size and density (i.e., texture) were fixed, and the central target was always presented at zero disparity such that the monkey’s vergence angle was fixed.

Disparity + Texture condition. This condition was identical to the Disparity Only condition, except that dot size and density varied to simulate the texture cues that would be associated with a plane of dots at different distances (see Fig. 2B). Vergence angle was again fixed.

Parts of this work have been published previously in abstract form (Wei et al., 2002).
Disparity + Vergence condition. A horizontal disparity was applied to all of the dots and the fixation point, thus requiring the monkey to converge at each of the five simulated viewing distances (see Fig. 5A). The pattern of horizontal and vertical disparities across the visual field (retinal image) was identical to that of the Disparity Only condition, and texture cues did not vary with simulated distance.

Disparity + Vergence + Texture condition. Same as the above condition, except that dot size and density also varied to simulate the texture cues associated with stimuli at different distances (see Fig. 5B). Note that this condition reproduces all of the disparity, texture, and vergence variations that occurred in the Moveable Screen condition. As explained in Discussion, there are two main differences between this and the Moveable screen condition. First, despite the close coupling between accommodation and vergence (Semmlow and Hung, 1983), the accommodative state of the animal in this condition would probably not be appropriate for all of the simulated viewing distances because stimuli were rendered in focus on the fixed screen at 32 cm. Thus, any differences in eye movement responses between this condition and the Moveable Screen condition are likely attributable to differences in image blur. A second difference between this and the Moveable Screen condition is the absence of naturalistic motion parallax cues in the stereograms attributable to the fact that the stereo half-images were not updated according to movement of the animal’s head (see Discussion).

Vergence Only condition. In this condition, like the previous two, a full-field horizontal disparity was applied to the dots and fixation target to drive vergence (see Fig. 4). Importantly, however, changes in the disparity field (both vertical disparities and horizontal disparities) that normally accompany vergence changes were cancelled by applying equal and opposite disparities to the screen image. Thus, whereas the monkey’s vergence angle varied in accord with the different simulated distances, the retinal image was maintained to be identical to that of a plane of dots viewed from 32 cm. Dot size and density were fixed. Thus, in this condition, changes in vergence posture were dissociated from changes in the retinal image.

A detailed description of the computations used to generate the stimuli for these five cue conditions is given in Appendix. The general approach is to compute the locations of dots in eye-centered coordinates (azimuth and elevation angles) that would correspond to viewing a texture wall at some simulated distance while fixing at some other distance. To generate the stereo images, we then find the location of dots on the screen that would give the same set of eye-centered coordinates when the eyes are converged on the fixed screen at 32 cm distance. This approach allows us to independently manipulate the monkey’s vergence angle and the visual cues to distance.

Random-dot stereograms simulating either a flat plane or the 3D surface of a pyramid

In a third block of trials, random-dot stereograms depicted either a flat plane at five different viewing distances (12, 18, 32, 52, and 102 cm) or a pyramid with a rectangular base. The base of the pyramid was always presented at zero disparity (i.e., the base appeared in the plane of the screen at a distance of 32 cm). The apex of the pyramid was simulated to be at five different distances (12, 18, 32, 52, and 102 cm), such that the apex pointed toward the subject for near distances and away from the subject for far distances (see Fig. 8). The fixation point was given a horizontal disparity to appear within the flat plane of dots or at the apex of the pyramid, such that the monkey was required to converge accurately at each different viewing distance. Thus, both the plane and pyramid stimuli elicited the same average vergence postures. The key difference between these stimuli was the presence of relative horizontal disparities in the pyramid stimulus, which provided for its 3D structure. The flat plane stimulus was identical to the Disparity + Vergence + Texture condition described above and was included in this experiment as a control to assess the role of relative horizontal disparities in the neural estimate of viewing distance.

Data analyses

All data analyses were performed offline using custom-written scripts in Matlab (MathWorks, Natick, MA). Eye movements were calibrated using a daily fixation task and then differentiated using a polynomial filter (Sivitsky and Golay, 1964). Sinusoidal modulations in eye velocity were quantified by fitting a sum-of-sinusoids (first and second harmonics) to the data using a nonlinear, least-squares algorithm based on the Levenberg-Marquardt method. All data reported here are taken from right-eye measurements; data from the left eye look almost identical. The magnitude of the compensatory eye movement during translation, which we will refer to as the TVOR response, was computed by taking the ratio of peak eye velocity (in degrees per second) to peak head velocity (in centimeters per second) for each individual cycle of lateral motion. Thus, the resultant response (as plotted in Figs. 3, 6, 9) has units of degrees per centimeter. For each cycle, a vergence angle was also computed as the difference between mean right- and left-eye positions. Data were analyzed separately for motion in the presence of visual stimuli and in darkness. For motion during visual stimulation, data from the last 1.5 sec of the visual stimulus period were included in the quantitative analyses. During motion in the dark, when the vergence window was typically increased from 1 to 5°, we only analyzed data from movement cycles during which the vergence angle deviated by <10% from the ideal value for a particular viewing distance.

The dependence of compensatory eye movements on vergence angle was quantified by examining the relationship between TVOR response and the inverse of a “vergence-defined viewing distance.” The latter was estimated from the instantaneous vergence angle and interocular distance and expressed as meter angles in units of m° (Paige and Tomko, 1991). One meter angle corresponds to a vergence state in which the two gaze directions intersect 1 m away from the subject. The relationship between TVOR response and vergence-defined inverse distance was quantified using linear regression. To properly stabilize the target, TVOR response amplitude should increase as a function of inverse distance with a slope of 0.57 (Figs. 3, 6, 9, dashed lines) (Paige and Tomko, 1991; Telford et al., 1997).

All statistical analyses used analysis of covariance with the different random-dot stereogram conditions as ordinal factor levels and vergence-defined viewing distance as a continuous covariate (Statistica; StatSoft, Tulsa, OK). For the comparisons between the Disparity Only and Disparity + Texture conditions in which vergence angle was not varied, the simulated viewing distance was used as the covariate.

Results

Horizontal eye movement responses to lateral whole-body motion (5 Hz oscillations in the interaural plane, 0.5 cm peak-to-peak) were recorded while two monkeys fixated a central target within a random-dot visual display. We shall refer to the resulting compensatory eye movements as the TVOR, although there could be contributions from other sources (e.g., ocular following) during the time period in which visual patterns were presented. To control for this possibility, TVOR responses were also studied during a period of darkness (Fig. 1, Pattern Off). Whole-body motion at 5 Hz was used because the contributions of visual feedback-based tracking mechanisms to eye stabilization are minimal at high frequencies (Paige and Tomko, 1991; Paige et al., 1998).

The main experimental tack was to compare TVOR responses during fixation on a surface at different physical distances (for which visual and motor cues to distance are congruent) with responses during viewing of surfaces at different simulated distances such that visual and motor cues can be dissociated. In some experiments, random-dot stimuli were rear projected onto a moveable screen that was placed at different distances from the animal. In other experiments, the screen was fixed at 32 cm from the animal, and stereoscopic displays simulated a frontoparallel plane of dots at different distances from the animal. To stabilize the retinal image during head movements, the amplitude of compensatory eye movements (TVOR response) should decrease with the subject’s viewing distance (Schwarz et al., 1989). An
To maintain fixation on the extinguished target in the dark, the random-dot pattern was turned off, because animals were less sensitive to eye velocity modulation with viewing distance persisted even when the eye movements were more oscillatory than the far target (52 cm). This change in eye movements was associated with changes in physical distances. As shown by the eye position and eye velocity traces in Figure 1, the near (12 cm) target was associated with both a larger change in eye position and velocity and an increase in the oscillatory eye movements than the far target (52 cm). This change in eye velocity modulation with viewing distance persisted even when the random-dot pattern was turned off, because animals were trained to maintain fixation on the extinguished target in the dark (Fig. 1, bottom traces, Pattern Off).

To scale the amplitude of the TVOR response correctly, the brain must construct an accurate estimate of the subject’s viewing distance. Under natural viewing conditions, a variety of sensory and motor signals (cues) are available for the computation of viewing distance. The most important motor cues are vergence angle and accommodation (Cumming et al., 1991; Sobel and Collett, 1991). Among the most important visual cues are the binocular disparity field (Rogers and Bradshaw, 1993, 1995; Bradshaw et al., 1996) and texture gradients (Gibson, 1950; Sedgwick, 1986). To investigate the relative roles of these visual and motor cues in the neural estimation of viewing distance for the TVOR, we performed a series of experiments in which visual and motor cues were manipulated separately and in various combinations. Results of these experiments will be presented in three parts: (1) effects of visual cues while vergence is fixed, (2) effects of visual cues in combination with changes in vergence angle, and (3) effects of relative horizontal disparities (3D structure).

Effects of visual cues only

To investigate whether visual cues by themselves could modulate eye movement amplitude during translation, we measured TVOR responses while monkeys viewed random-dot patterns that reproduced the binocular disparity and texture variations that the monkey would experience by fixating on a plane of dots at different distances (12, 18, 32, 52, and 102 cm). All stimuli were rear projected onto a screen at 32 cm, such that the vergence demand was identical at all simulated distances (Fig. 2) (see Materials and Methods). Hence, these stimuli isolated visual cues to viewing distance. In the Disparity Only condition (Fig. 2A), the peripheral horizontal and vertical disparities in the image varied while dot size and density were fixed. In the Disparity + Texture condition (Fig. 2B), dot size and density varied along with the disparity field to simulate different viewing distances. The latter stimulus condition is more realistic, because perception of distance and depth can be strongly influenced by texture cues (Newman, 1972) (for review, see Sedgwick, 1986; Collett et al., 1991; Johnston et al., 1993).

Data from monkey K are shown in Figure 3. These data were taken from cycles of lateral motion during which the monkey viewed the random-dot stimuli, but similar results were obtained during motion in the subsequent period of darkness as discussed below (see Fig. 7). For the Disparity Only condition, the slope of the relationship between TVOR response and simulated inverse distance was 0.008 ± 0.04 SD, which was statistically significant (F(1,1613) = 89.2; p < 0.0001) because of the large number of samples (1613) in the analysis. Note, however, that the slope of this relationship is 80-fold shallower than the ideal slope for perfect image stabilization (dashed line), indicating that variations in the disparity field alone had very little effect on eye movement amplitude. The slope of the data was slightly, but significantly, steeper (0.02 ± 0.03 SD) in the Disparity + Texture condition (F(1,3175) = 64.0; p < 0.0001). However, even with congruent variations in both binocular disparities and texture, the observed effects were more than an order of magnitude weaker than the expectation for ideal performance based solely on visual cues. Thus, the data of Figure 3 suggest that visual cues, by themselves, make a negligible contribution to scaling the amplitude of the TVOR response with viewing distance.

Combined effects of visual cues and vergence angle

The lack of a large effect of visual cues on the TVOR response could have been a consequence of the fact that these cues were in direct conflict with viewing distance specified by vergence angle (and possibly also accommodation). To evaluate the contributions of vergence, we conducted experiments in which vergence angle was systematically varied by adding a full-field horizontal disparity component to the random-dot pattern and fixation point (Figs. 4, 5). Trained animals fused the central target and maintained their vergence within ±0.5° of the desired angle corresponding to each of five simulated viewing distances (12, 18, 32, 52, and 102 cm). Three stimulus conditions were examined: one in which vergence angle varied but the retinal image was always identical to that of a plane of dots viewed from 32 cm (Fig. 4, Vergence Only condition), one in which both vergence angle and binocular disparities specified viewing distance (Fig. 5A, Vergence + Disparity condition), and one in which vergence angle, disparities, and texture all consistently specified one of the five simulated distances (Fig. 5B, Vergence + Disparity + Texture condition).

Data from monkey K for each of these stimulus conditions are shown in Figure 6, along with data collected while the monkey viewed random-dot patterns presented on a moveable screen that was physically placed at different distances (gray circles) (for details, see Materials and Methods). Three main points should be taken from the data of Figure 6. First, scaling of TVOR response with distance was much stronger for the Vergence Only condition (slope of 0.14 ± 0.06 SD; black circles) than for either of the purely visual conditions in Figure 3. Second, adding peripheral disparities (Disparity + Vergence condition; black triangles) or...
disparity and texture cues (Disparity + Vergence + Texture condition; white squares) to the pure vergence condition produced only a small increase in the slope of the relationship between TVOR response and viewing distance (0.16 ± 0.07 SD for both).
Thus, visual cues contribute very little to distance scaling of the TVOR, even when superimposed onto the appropriate vergence eye movements. Third, the slope of the data for the Moveable screen condition (0.35 ± 0.06 SD) was substantially larger than that for any of the other conditions in Figure 6 and began to approach the ideal slope (dotted line) for perfect image stabilization.

Figure 7 summarizes the data from both monkeys. Slopes of the relationship between TVOR response and inverse distance are plotted for each of the stimulus conditions shown previously in Figure 6 and began to approach the ideal slope (dotted line) for perfect image stabilization.

Figure 6. Scaling of compensatory eye movement amplitude (TVOR response) as a function of inverse viewing distance. To account for potential inaccuracies in vergence posture at the different simulated distances, the horizontal axis plots the inverse viewing distance as defined by the vergence posture during each cycle of whole-body motion. Black circles show data from the Vergence Only condition (Fig. 4), black triangles show data from the Disparity + Vergence condition (Fig. 5A), and white squares show data from the Disparity + Vergence + Texture condition (Fig. 5B). For comparison, gray circles show data for stimuli presented on a Moveable Screen placed at different physical distances. Data are shown from animal K. Solid lines indicate the best linear fits to the data.
Figures 3 and 6. Error bars denote the 95% confidence interval for each slope, so that the statistical significance of differences between conditions can be assessed visually by noting whether or not the confidence intervals overlap. Data obtained during presentation of the random-dot patterns are presented on the left (Pattern ON); data obtained during the subsequent period of darkness are shown on the right (Pattern OFF). For both animals, purely visual cues (D and DT conditions) yielded small but statistically significant slopes during presentation of the random-dot pattern, and these effects mostly became nonsignificant in darkness. For both monkeys, during both pattern viewing and in darkness, the slopes of the response–distance functions were substantially steeper for random-dot patterns presented on the Moveable Screen (MScr condition) than for any of the disparity-driven vergence conditions (DVT, DV, and V) that were presented on the fixed screen at 32 cm.

The improved distance scaling of the TVOR in the Moveable Screen condition cannot be attributable to differences in vergence posture because the monkeys were forced to converge accurately in each condition. Moreover, the data are analyzed (Fig. 6) as a function of the measured vergence angle to determine the slopes. Random-dot patterns were computed such that the retinal image in the Disparity + Vergence + Texture (DVT) condition was identical to the retinal image in the Moveable Screen (MScr) condition. Thus, the steeper slopes in the MScr condition cannot be attributable to visual cues. It is worth noting that different distances were presented in separate blocks in the MScr condition, whereas they were interleaved in the DVT condition. To rule out this possible confound, we repeated the D, DT, DV, and DVT conditions using a blocked design and found an identical pattern of results (data not shown). Together, these findings imply that other nonvisual cues besides vergence angle must have a strong effect on distance scaling of the TVOR, with accommodation being the most likely candidate (Schwarz et al., 1989; Schwarz and Miles, 1991) (see Discussion).

Effects of 3D structure (relative disparities)
In the experiments described above, our random-dot stimuli (Disparity + Vergence + Texture condition) accurately simulated the retinal images that would arise from a frontoparallel plane of dots viewed at different distances. This is still a fairly impoverished visual scene, however, because it contains no depth variations. Psychophysical studies have shown that perception of depth from horizontal disparities requires the presence of relative disparities within the field of view (Regan et al., 1986). Thus, we sought to test whether adding 3D structure based on relative disparities would have an effect on TVOR amplitude.

In a separate block of trials, we presented random-dot stereograms that depicted a three-dimensional pyramid whose base lay in the plane of the screen at 32 cm and whose apex pointed either toward or away from the observer (Fig. 8). The fixation target was located at the apex of the pyramid and appeared at simulated viewing distances of 12, 18, 32, 52, and 102 cm. Frontoparallel planar stimuli (as in the Disparity + Vergence + Texture condition) at the same five simulated distances were interleaved with the pyramids. Results from monkey K are illustrated in Figure 9. The slope (±95% confidence interval) of the best-fitting line was 0.153 ± 0.004 for the pyramid stimuli (Pr; white squares) and 0.146 ± 0.004 for the plane stimuli (Pl; black triangles). Corresponding values for monkey P were 0.173 ± 0.005 and 0.172 ± 0.004, respectively. The difference in slope between the pyramid and plane stimuli was not significant for either animal (p > 0.05). For both animals, however, the slope of the data for the Movable Screen condition (Fig. 7, MScr) was significantly larger (p < 0.0001) than that for the pyramid and plane conditions.

Thus, the presence of 3D structure based on relative horizontal disparities, like the vertical disparity and texture cues, had little or no effect on the neural estimate of viewing distance used to scale the TVOR response. Similar to the experiments of Schwarz and Miles (1991), the ineffectiveness of visual cues was further verified by having one of the monkeys fixate on light-emitting diodes (LEDs) presented at different distances in a softly lit room (black squares) or fixated on a plane of dots presented on a moveable screen (gray circles). Data are shown from animal K.

Figure 9. Effects of 3D structure defined by relative horizontal disparities on distance scaling of the TVOR. TVOR response is plotted as a function of vergence-defined inverse distance for the pyramid (white squares) and plane (black triangles) stimuli of Figure 8. For comparison, data are shown for conditions in which the monkey fixated on LED targets at different distances in a softly lit room (black squares) or fixated on a plane of dots presented on a moveable screen (gray circles). Data are shown from animal K.
the Moveable Screen experiments (such as perspective and familiar size), but these visual cues were also not effective in scaling the amplitude of TVOR responses with viewing distance.

**Discussion**

Knowledge of one’s viewing distance (the distance from the eyes to the point of fixation) is important for a variety of perceptual and motor processes. For depth perception, the pattern of horizontal disparities between the two retinal images must be scaled by viewing distance to compute depth (Ono and Comerford, 1977; Foley, 1980; Rogers and Bradshaw, 1993; Glennerster et al., 1998). Previous psychophysical work with humans has shown that visual cues (vertical disparities) and motor cues (vergence angle) both make substantial contributions to this process (Cumming et al., 1991; Rogers and Bradshaw, 1993, 1995; Bradshaw et al., 1996). In the oculomotor system, compensatory eye movements that stabilize the retinal image during head motion must also be scaled by viewing distance to work properly (Schwarz et al., 1989; Paige and Tomko, 1991; Schwarz and Miles, 1991). Previous work with monkeys has shown that motor cues (both vergence and accommodation) have a very weak effect on scaling the amplitude of these compensatory eye movements according to viewing distance (Schwarz and Miles, 1991), but the effects of visual cues in this process have remained unclear.

One might expect that the brain would integrate a variety of visual and motor cues to compute a robust estimate of viewing distance, and previous psychophysical work on human depth perception supports this notion. On the basis of these psychophysical results, we expected that the amplitude of TVOR responses would be scaled by manipulation of visual cues to viewing distance. Our findings, however, are not consistent with this expectation and can be summarized as four main points. (1) Visual cues to distance, either by themselves or in conjunction with changes in vergence angle, had a very small effect on the amplitude of TVOR responses. (2) Stimulus conditions that simulated fixation on a plane of dots at different distances (either via disparity-driven vergence alone or by vergence in combination with visual cues) produced much weaker scaling of eye movements than fixation on a plane of dots that was presented on a moveable screen. Because the retinal images and vergence angles were comparable in these two conditions, the difference in results between the fixed and moveable screens must be attributable to differences in accommodative demand. (3) Adding relative horizontal disparities to create 3D structure in the visual displays had no effect on scaling of compensatory eye movements. (4) Distance scaling never reached the ideal relationship that would be needed to stabilize the eyes on the target at all distances, even in the Moveable Screen and real LED conditions for which vergence posture, accommodative state, and visual cues were all in agreement. Implications of this finding will be discussed further below.

Given that at least some of the visual cues used in our study (e.g., vertical disparities) are known to contribute to distance scaling for depth perception (Rogers and Bradshaw, 1993, 1995; Bradshaw et al., 1996), our results suggest that the perceptual and oculomotor systems do not rely on a common (neural) estimate of viewing distance. This conclusion relies on a comparison between human psychophysics and monkey eye movement studies, thus some caveats should be noted. First, we assume that visual cues to viewing distance would have the same scaling effects on monkeys’ judgments of depth as they have on humans. This has not been tested directly, but disparity-driven vergence movements and stereoaucuity thresholds have been shown to be very similar for humans and monkeys once the species difference in interocular separation is taken into account (Harwerth et al., 1995). Thus, we think it is likely that monkeys would make use of visual cues to viewing distance for depth perception. Second, we implied that the oculomotor and perceptual systems rely on different estimates of viewing distance that are computed separately. Alternatively, a single locus for this computation could exist in the brain, but the weights assigned to different visual and motor cues could be task dependent, such that visual cues to viewing distance are used for depth judgments but gated out of the computation when the animal is simply fixating (as in our task). The very weak effects of visual cues that we observed are consistent with this latter possibility.

One possible explanation for the weak effects of visual cues in our experiments involves processing speed. Computation of viewing distance from visual cues would require at least several tens of milliseconds as a result of visual latencies and processing time. In contrast, vergence and accommodation signals arising from efference copy or proprioception would be relatively fast. Because distance scaling of the TVOR must vary rapidly during natural vision, visual cues to viewing distance may be too slow to contribute to image stabilization. Although there was ample time for visual cues to distance to be used by the monkey in our experiments, the oculomotor system does not appear to make use of these signals. It is worth noting, however, that relative disparity cues and depth perception were reported previously to affect the amount that the 3D coordinate system for saccades and fixations (Listing’s plane) rotates with coverage (Kapoula et al., 1999), suggesting that some oculomotor functions may respect visual cues to distance.

**Relationship to previous studies**

Miles and colleagues (Schwarz et al., 1989; Schwarz and Miles, 1991) trained monkeys to fixate on small target lights located at different distances from the animal and observed that the amplitude of the TVOR response scaled approximately linearly with inverse viewing distance. Using prisms to dissociate vergence and accommodative demands, they further demonstrated that vergence and accommodation make approximately equal contributions to scaling the TVOR with viewing distance. Our findings are quite consistent with theirs in this regard. The difference between our Disparity + Vergence + Texture and Moveable Screen conditions is likely attributable to accommodative demand because the retinal images and vergence angles were equated in these two conditions. Despite the close link between vergence angle and accommodation (Semmlow and Hung, 1983), we would expect accommodation to be at least partly in conflict with vergence in the Disparity + Vergence + Texture condition because all stimuli were presented on a fixed screen at 32 cm in this case. We found that the slope of the data was approximately twice as steep for the Moveable Screen condition as it was for the Disparity + Vergence + Texture condition (Fig. 7), consistent with approximately equal contributions of vergence and accommodation to distance scaling.

An additional difference between the Disparity + Vergence + Texture and Moveable Screen conditions involves motion parallax. Because we did not update our random-dot stereograms based on the actual position of the animal’s head during the ±0.25 cm sinusoidal trajectory, the Disparity + Vergence + Texture condition lacks the naturalistic motion parallax cues of the Moveable Screen condition. A lack of motion parallax cues explains why moving observers commonly experience a powerful illusory rotation of 3D structure in stationary stereograms. Be-
cause it is known that motion parallax can influence the direction of motion perceived by the observer (Warren and Hannon, 1990; Roy et al., 1992), we cannot rule out the possibility that lack of motion parallax cues contributed to the weaker distance scaling of TVOR responses seen in the Disparity + Vergence + Texture condition (although we consider this possibility unlikely).

Although TVOR amplitude scaled linearly with distance in the studies of Miles and colleagues, the scaling was insufficient to allow perfect image stabilization (Telford et al., 1997; Paige et al., 1998; Angelaki and Hess, 2001; Angelaki, 2002). Schwarz and Miles (1991) suggested that this shortfall could be attributable to a lack of rich visual cues to viewing distance in their experiments. However, we provided a variety of visual cues in our experiments, and we still find that distance scaling of the TVOR is incomplete. Because of this apparent discrepancy, it is worth clarifying some of the differences between the two sets of experiments. In the experiments of Schwarz and Miles (1991), room lights were on while the monkey fixated and were extinguished just before the brief, transient movement of the animal. Thus, visual cues to viewing distance were available to the animals, although the nature and salience of the visual cues (content of the room) is unclear. The similarity of the results of Schwarz and Miles (1991) to our Moveable Screen and LED conditions suggests strongly that visual cues were not responsible for the shallower-than-ideal slopes that they observed. Schwarz and Miles (1991) used prisms to vary vergence angle while keeping retinal images fixed, similar to what we did in our Vergence Only condition. On the other hand, they could not vary visual cues to viewing distance while keeping vergence and accommodative demands fixed. In this study, we fully dissociated visual and motor cues to viewing distance, and this reveals clearly that visual cues do not contribute substantively to distance scaling of the TVOR.

Because visual cues cannot account for the imperfect distance scaling in both our experiments and those of Miles and colleagues, what factors could be lacking in these experiments to prevent optimal distance scaling? Previous studies with human subjects actively moving through their environment have reported nearly perfect image stabilization (retinal slip of <4°/sec). In these studies, involving walking and running, translation of the head in one direction was always associated with a head rotation in the other direction (Demer and Viirre, 1996; Crane and Demer, 1997). Whether the nearly perfect image stabilization with near target viewing during natural activities is attributable to rotation–translation active synergies (Bloomberg et al., 1992; Crane and Demer, 1997; Moore et al., 1999; Medendorp et al., 2002) and/or to nonlinear semicircular canal–otolith interactions (Telford et al., 1998; Fuhry et al., 2002) remains unknown at present.

Snyder et al. (1992) have reported that scaling of the TVOR by viewing distance precedes the actual change in a monkey’s vergence angle by at least 50 msec. This result implies that central motor commands, rather than low-level efference copy or proprioceptive information, may be the signals used to scale the TVOR with viewing distance. Our results demonstrate that these command signals, whatever their precise nature, do not incorporate the available visual cues to viewing distance. The differences in use of egocentric distance cues between the monkey’s oculomotor system and human depth perception are consistent with the idea of separate visual representations for perception and action (Goodale and Milner, 1992; Goodale et al., 1994; Burr et al., 2001).

Appendix

Here, we describe the computations used to generate the random-dot stereograms (Figs. 2, 4, 5) that were used to simulate a textured plane at different viewing distances. The basic task is to generate dots (on a fixed screen) that recreate the pattern of binocular disparities and vergence angle that a subject would experience if she were viewing a plane of dots at a distance $z_P$ while fixating at a distance $z_F$. We accomplish this task by computing the eye-centered coordinates (azimuth and elevation angles) of the simulated plane of dots viewed from the simulated viewing distance. We then apply a coordinate transformation to rotate the eyes to the actual fixation position, and we compute (by calculations that are equivalent to ray tracing) the set of points on the display screen that correspond to the eye-centered coordinates of the simulated pattern. The only assumption that we make in this analysis is that there are no torsional eye movements involved in changing fixation from one viewing distance to another. This assumption is valid as long as the fixation points are directly in front of the animal (Ferman et al., 1987; Tweed and Vilis, 1987, 1990), which was always the case in our experiments.

Coordinate frame

We start by defining a head-centered coordinate system in which the $x$-axis points rightward, the $y$-axis points upward, and the $z$-axis points backward (to have a right-handed system). Thus, objects in front of the observer have negative $z$ coordinates. The origin, $(x, y, z) = (0, 0, 0)$, is taken to be the location of the nodal point of the cyclopean eye. In this coordinate frame, the location of the left eye is given by

$$L = (-IO/2, 0, 0),$$

and the location of the right eye is given by

$$R = (IO/2, 0, 0),$$

where $IO$ denotes the interocular distance, which is typically ~3.5 cm for a rhesus monkey. We define the location of the simulated fixation point by the vector $F$,

$$F = (xF, yF, zF),$$

and we define a point on the simulated plane of dots by the vector $P$,

$$P = (xP, yP, zP),$$

where $zP = -p$ is constant for all points on the plane.

Similarly, we define the monkey’s actual point of fixation by the vector $G$,

$$G = (xG, yG, zG),$$

and we define the plane of the display screen as the locus of points with $z = -s$. To simulate the point $P$ as viewed from the simulated fixation point $F$, we thus need to draw two dots on the display screen: one to be viewed by the left eye (drawn at screen location $SL$) and one to be viewed by the right eye (drawn at screen location $SR$), where

$$SL = (xSL, ySL, -s)$$

and

$$SR = (xSR, ySR, -s).$$
Computing the eye-centered coordinates
To compute the eye-centered coordinates of a point P, we start by rotating the two eyes from pointing straight ahead (down the z-axis) to pointing at the simulated fixation point F. Assuming no torsion, this can be done with a pair of rotation matrices: one to rotate the eye around the y-axis and a second to rotate the eye around the x-axis. Thus, to fixate the right eye on point F, we have the following matrices:

\[
\begin{align*}
RyFR &= \begin{pmatrix}
\cos(\beta FR) & 0 & -\sin(\beta FR) \\
0 & 1 & 0 \\
\sin(\beta FR) & 0 & \cos(\beta FR)
\end{pmatrix} \\
RxFR &= \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos(\alpha FR) & \sin(\alpha FR) \\
0 & -\sin(\alpha FR) & \cos(\alpha FR)
\end{pmatrix}
\end{align*}
\]

where \(\beta FR\) and \(\alpha FR\) are the corresponding azimuth and elevation angles, respectively. Similar matrices \(RxFL\) and \(RyFL\) can be defined to rotate the left eye to point F. Additionally, matrices \(RyGR, RyGR, RxGL,\) and \(RyGL\) are defined to rotate the two eyes to the actual point of fixation, G.

With these definitions, we can now write expressions for the location of a point \(P\) in eye-centered coordinates. Let \(PRI\) denote the location of point \(P\) in the right eye while the eyes are fixating at \(F\). The expression for \(PRI\) is thus given by the following:

\[
PRI = RxFR RyFR (P - R).
\]

A similar expression can be written for the location of point P, as viewed from F, in left-eye-centered coordinates, \(PLi\). Solutions for the rotation angles of the eyes can be derived in straightforward manner. For the right eye,

\[
\beta FR = \arctan\left(\frac{(F - R) \cdot \hat{x}}{(F - R) \cdot \hat{z}}\right)
\]

\[
\alpha FR = \arctan\left(\frac{(F - R) \cdot \hat{y}}{-\sin(\beta FR)(F - R) \cdot \hat{x} - \cos(\beta FR)(F - R) \cdot \hat{z}}\right)
\]

where \(\hat{x}, \hat{y},\) and \(\hat{z}\) denote unit vectors in the head-centered coordinate frame. Similar expressions define the rotation angles for the left eye fixating at F and for each eye fixating at G.

Computing the screen coordinates
We ultimately want the set of screen coordinates, \(SR\) and \(SL\), which produce eye-centered headings (azimuth and elevation angles) that match those given by \(PRI\) and \(PLi\). In other words, the eye-centered values of the dots on the screen, which we denote by \(SRi\) and \(SLi\), must be related to \(PRI\) and \(PLi\) by a constant that simply scales along the eye-centered heading. Thus, we can write

\[
\begin{align*}
SRi &= kR \cdot PRI \\
SLi &= kL \cdot PLi,
\end{align*}
\]

where \(kR\) and \(kL\) are the scale factors. For the right eye, we can expand this expression to

\[
RxGR RyGR (SR - R) = kR \{RyFR RyFR (P - R)\},
\]

and we can rearrange it to

\[
SR - R = kR \{RyGR' RyGR' RyFR RyFR (P - R)\}.
\]

We know that the z coordinates of \(SR\) and \(SL\) are given by \(-s\), because they are constrained to lie on the display screen. Thus, examining only the z component of the above equation allows us to solve for \(kR\):

\[
kR = \frac{-s}{\{RyGR' RyGR' RyFR RyFR (P - R)\} \cdot \hat{z}}.
\]

Similarly, \(kL\) has a corresponding solution.

Using the above solutions for \(kR\) and \(kL\), we now have complete expressions for the locations of dots on the screen that are needed to simulate viewing of point P from a fixation point F:

\[
\begin{align*}
SR &= kR \{RyGR' RyGR' RyFR RyFR (P - R)\} + R \\
SL &= kL \{RyGL' RxGL' RxFL RyFL (P - L)\} + L.
\end{align*}
\]

These equations were used to generate the random-dot stereograms used in this study. To test the validity of our derivations, we computed the vertical disparities of points on one of our simulated planes, and we compared these disparities with those computed from Mayhew and Longuet-Higgins (1982), their Equation 2. Their equation for vertical disparity assumes that the fixation distance (\(zF\)) is large relative to the interocular distance (\(IO\)), whereas our formulation makes no such assumptions. As expected, we found perfect agreement between the two sets of calculations for \(zF \gg IO\).

Experimental conditions
The formulation that we used allows us to flexibly specify the 3D structure of the display, as well as the real and simulated fixation distances. For the experimental conditions described in Materials and Methods involving frontoparallel textured planes, the stimulus conditions were derived using the following subsets of parameters. Note that the real and simulated fixation points were always located directly in front of the animal, such that \(xF = yF = xG = yG = 0\).

Disparity Only or Disparity + Texture conditions:

\[
\begin{align*}
&zG = 32 cm \text{ (real fixation on the display screen)} \\
&zF = zP = 12, 18, 32, 52, 102 cm \text{ (simulated fixation on the plane of dots at various distances).}
\end{align*}
\]

Disparity + Vergence or Disparity + Vergence + Texture conditions:

\[
\begin{align*}
&zG = zF = zP = 12, 18, 32, 52, 102 cm.
\end{align*}
\]

Vergence Only condition:

\[
\begin{align*}
&zG = 12, 18, 32, 52, 102 cm \text{ (real fixation position varies)} \\
&zF = zP = 32 cm \text{ (simulated fixation and dot locations in the plane of the displayscreen, such that the retinal image does not vary with zG).}
\end{align*}
\]

For the conditions involving texture variations, dot size was scaled inversely with the z coordinate of each point.

References