Binocular and monocular depth cues in online feedback control of 3D pointing movement

Bo Hu

Center for Visual Science, University of Rochester, Rochester, NY, USA

David C. Knill

Center for Visual Science, University of Rochester, Rochester, NY, USA

Previous work has shown that humans continuously use visual feedback of the hand to control goal-directed movements online. In most studies, visual error signals were predominantly in the image plane and, thus, were available in an observer's retinal image. We investigate how humans use visual feedback about finger depth provided by binocular and monocular depth cues to control pointing movements. When binocularly viewing a scene in which the hand movement was made in free space, subjects were about 60 ms slower in responding to perturbations in depth than in the image plane. When monocularly viewing a scene designed to maximize the available monocular cues to finger depth (motion, changing size, and cast shadows), subjects showed no response to perturbations in depth. Thus, binocular cues from the finger are critical to effective online control of hand movements in depth. An optimal feedback controller that takes into account the low peripheral stereoacuity and inherent ambiguity in cast shadows can explain the difference in response time in the binocular conditions and lack of response in monocular conditions.

Keywords: online feedback control, goal-directed movement, depth perception, binocular disparity, cast shadow

Citation: Hu, B., & Knill, D. C. (2011). Binocular and monocular depth cues in online feedback control of 3D pointing movement. *Journal of Vision, 11*(7):23, 1–13, http://www.journalofvision.org/content/11/7/23, doi:10.1167/11.7.23.

Introduction

Recent studies on goal-directed pointing movements (Brenner & Smeets, 2003, 2006; Izawa & Shadmehr, 2008; Liu & Todorov, 2007; Prablanc & Martin, 1992; Sarlegna et al., 2004; Saunders & Knill, 2003, 2004, 2005; van Mierlo, Louw, Smeets, & Brenner, 2009) converge on the conclusion that the visual system continuously uses visual information from both the target and the moving hand to guide it to a target. The main paradigm used in these studies has been to measure subjects' corrective responses to perturbations in the visual information about the position of the finger and/or the target. Comparisons of subjects' corrective responses with those predicted by ideal observers/actors show that the CNS uses information online in an optimal way—integrating visual signals from the moving hand and the target over time in a manner commensurate with the reliabilities of the signals (Izawa & Shadmehr, 2008; Saunders & Knill, 2004).

While subjects movements in many of these studies included movement in depth, the perturbations applied, except for Brenner and Smeets (2006), included components that were readily present in the retinal image plane. This makes it difficult to examine the contribution of visual depth cues to online control in goal-directed pointing tasks. In fact, the tasks in most of these studies were essentially 2D: either the target and starting points were at the same depth or the movement was confined in a plane (typically a horizontal plane). Yet many natural pointing tasks consist of free movements in three dimensions, including motion in depth. The visual information about the hand's position and movement in depth is qualitatively different from visual information about its position and movement in the image plane. The latter is given directly by the 2D projection of the hand on the retina, while depth information is carried more indirectly by a rich set of visual depth cues, both binocular and monocular. We set out to investigate how humans use visual depth cues from the hand for online feedback control of 3D movements.

Binocular depth information from a target has been shown to be important in online control in grasping (Bradshaw & Elliott, 2003; Greenwald & Knill, 2009b; Jackson, Jones, Newport, & Pritchard, 1997; Loftus, Servos, Goodale, Mendarozqueta, & Mon-Williams, 2004) and object placement tasks (Greenwald, Knill, & Saunders, 2005; Knill, 2005; Knill & Kersten, 2004; van Mierlo et al., 2009). The CNS also uses monocular depth cues about the orientation of a target object (texture and figure outline shape) to control the orientation of the hand online during grasping and object placement movements (Greenwald & Knill, 2009a; Greenwald et al., 2005; Knill, 2005; Knill & Kersten, 2004). Little is known, however, about the role played by different depth cues in providing feedback about the depth of the moving hand for online control.

A recent study by Brenner and Smeets (2006) provides information about the speed at which subjects can respond

doi: 10.1167/11.7.23 Received March 30, 2011; published June 30, 2011 ISSN 1534-7362

to abrupt changes in depth of a target or an effector controlled by an observer. In their task, subjects used a mouse to move a cursor quickly from a starting location on a computer screen to a target location. The target location appeared at two different depths relative to the screen and both the cursor and the target were shown in stereo. Early in the movement on some trials, either the cursor or the target jumped 15 cm in depth toward or away from the observer. Subjects corrected for these perturbations within approximately 200 ms of the perturbations; thus, it is clear that in principle subjects can correct for changes in depth of an effector quickly, though not quite as quickly as they can correct for perturbations in the image plane.

These data provide a useful bound on the possible speed with which the CNS can correct for movement errors in depth based on visual feedback, but it is unclear whether it reflects the behavior of an autopilot feedback system that normally controls hand movements or something else. First, the perturbations used in the study were extremely large, much larger than the errors one would normally find from noise in the sensorimotor system. Second, the large perturbations created sizable sharp temporal transients in the visual feedback from the hand that is not normally associated with naturally occurring movement errors. Finally, in part because of the size and unnaturalness of the perturbations, the perturbations must have been clearly detectable by subjects who could then have been attuned to detect and correct for them.

In the current experiments, subjects performed a pointing task in free 3D space behind a mirror through which visual feedback was given by a binocularly rendered virtual finger optically co-aligned with the subjects' real finger. The subjects' goal was to touch a target ball positioned in the workspace by a robot arm and rendered in the virtual display to be optically co-aligned with the actual ball. We measured corrective responses on trials in which the position and/or motion of the virtual finger was perturbed by a small amount in depth or in the image plane (within the variance of movement trajectories) when the finger disappeared behind a virtual occluder.

The first experiment was designed to assess basic properties of how the CNS uses visual depth information about the moving hand during online control of pointing movements. Subjects made unconstrained pointing movements in free space, so that the only visual objects in the scene were the starting position, the target ball, and the finger. We measured subjects' corrective responses to small perturbations of the virtual finger in depth and compared them with subjects' responses to equal-sized perturbations in the image plane. We further explored the relative efficacy of position and motion information by looking at corrective responses to simple step perturbations in depth and perturbations in which the virtual finger is rotated around the target, causing an initial shift in depth but a change in motion direction to keep the motion of the virtual finger relative to the target unchanged.

In the first experiment, visual cues to finger depth included both binocular disparity cues (static and dynamic) and a few monocular cues—finger size (static and dynamic) and motion parallax (using kinesthetic motion signals as a baseline for scaling velocity in the image plane to estimate depth and motion in depth). In a second experiment, we created a visual scene designed to maximize monocular cues to depth and measured the contributions of these cues to online feedback control by measuring subjects' corrective responses to depth perturbations in monocular conditions and comparing them to their corrective responses to image plane perturbations. In particular, we rendered a textured tabletop over which subjects moved their fingers and illuminated the scene with a directional light source to add cast shadows of the finger and the target. While cues like size change may be ineffective for feedback control because of the relatively small difference in the amount of size change created by natural variations in hand movements, numerous studies (Kersten, Knill, Mamassian, & Bulthoff, 1996; Mamassian, Knill, & Kersten, 1998) have shown that cast shadows can greatly affect depth perception in both static and dynamic scenes. In the "ball in box" experiment, for example, Kersten, Mamassian, and Knill (1997) demonstrated that cast shadows induce a strong sense of motion in depth even when the size of an object stays constant. It, thus, seems plausible that the visual system can utilize cast shadows to guide finger movement. On the other hand, the depth information from cast shadows is inherently ambiguous; it depends on the visual system's knowledge of the light source and the geometry of the surface that the shadows are cast on. In this experiment, we presented subjects the needed information to remove the ambiguity, making cast shadows a theoretically viable depth cue. Similar to binocular disparities, the CNS could use the shadow information to infer depth information or it could use the relative motion of the finger's and target's shadows directly to guide the hand.

Methods

General design

Subjects performed a pointing task by moving their index finger from a starting ball to a target ball in 3D space. Visual information about the finger and the balls was provided by computer graphics displayed on a CRT monitor and reflected into the workspace by a mirror (Figure 1). In over half of the trials, the virtual finger rendered in the display was displaced from the real finger position by a small amount. This perturbation happened when the finger was behind a virtual occluder, which masked the onset of the perturbations (Figures 2 and 3).

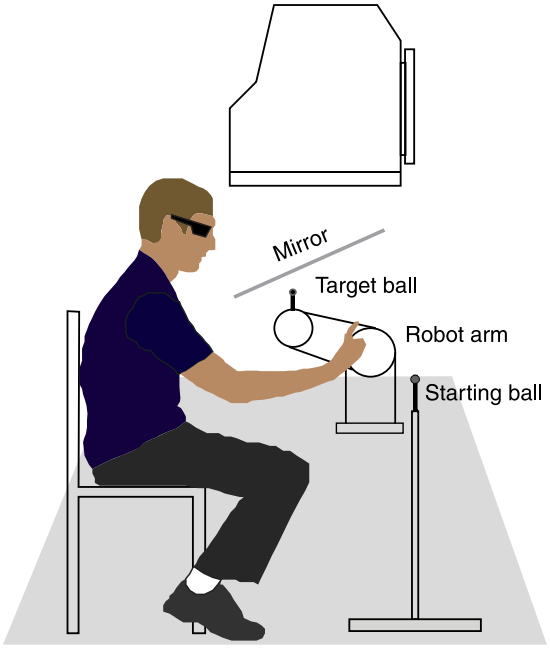

Figure 1. The schematic of the experiment setup. Subjects moved their finger to reach for the target ball mounted on a robot arm. The visual information of the finger and the balls was provided by computer graphics on the computer monitor and reflected into the workspace. Subject could not see their hand or the balls during the movement.

The perturbations were applied in a moving coordinate frame centered at the current position of the real finger. At each instance, the line of sight defined by the cyclopean eye (the point midway between the two eyes) and the finger was what we call the depth dimension and the plane perpendicular to it was the image plane (Figure 4a). The goal was to separate and compare how visual signals in depth and in the image plane were processed; we, therefore, added the same perturbations both in depth (in-depth perturbations) and in the image plane (in-image perturbations).

We applied two types of perturbation to the virtual finger in Experiment 1. The first was a small, 1-cm step perturbation that added a fixed offset between the real and virtual finger until the end of movement (Figure 4b). The second was a small rotation perturbation, in which the angle between the virtual finger and the target was increased or decreased by 2.6–3.9 degrees (so as to create an initial 1-cm shift in position when the virtual finger appeared from behind the occluder). These were also imposed in a moving coordinate frame in which the z-axis was aligned with the line of sight and included rotations in the X–Z plane (in-depth rotation perturbations) and rotations in the X–Y plane (image plane rotations). Rotation perturbations caused positions shifts that started at 1 cm when the finger emerged from the occluder and decreased over time, vanishing at the target position (Figure 4c). The rotation perturbations keep the motion of the virtual finger *relative* to the target unchanged and corrective responses actually decrease endpoint accuracy.

The visual display in Experiment 2 was richer than that in Experiment 1 to give subjects information of where the light source was and the distance and orientation of the plane that the shadows were cast on. We used a fixed directional light source in all trials and set the light source direction from above the subjects' head—in agreement with people's prior assumptions on light source direction. We put a checkerboard texture on the ground plane, which all the shadows were cast on and covered the field of view (Figure 3b). The ground plane coincided with the tabletop used during the initial calibration of subject's eye positions (see Procedures section) at the beginning of each session. The ground plane was about 50 cm from a subject's cyclopean eye. The virtual target ball was rendered on top of a pole, which was perpendicular to the ground plane and whose height changed with the position of the target ball. The pole provided extra information along with its shadow to estimate light source direction and to localize the virtual target. This scenario also created the possibility for subjects to use the relative position of the two cast shadows directly as a control signal for correcting movement errors in depth.

In view of humans' high sensitivity to dynamic monocular cues to motion in depth, for example, results showing a dramatic influence of cast shadow motion on perceived motion in depth (Kersten et al., 1997), it is clear that motions of various kinds in the image can provide strong monocular cues to motion in depth. Static monocular depth cues, however, are poor indicators of absolute depth from the viewer. We, therefore, perturbed the direction of motion of the virtual finger rather than its position when it emerged from behind the occluder. The perturbations started at 0 and increased over time so that if subjects did not correct for the perturbation the virtual finger would be 1 cm away from the target in the appropriate dimension (in-image or in-depth; Figure 4d).

Subjects

Sixteen subjects participated in the study, eight in each experiment. All subjects had corrected vision and the eight subjects in Experiment 1 had scores of eight or higher on the Randot (Precision Vision, IL) stereo test. All subjects were right-handed. All subjects were students of the University of Rochester and naive for the purpose of the experiments. All provided informed consent in accordance with the guidelines from the University of Rochester Research Subjects Review Board.

Apparatus and display

Figure 1 illustrates the physical setup for the experiments. The tasks were performed in a calibrated virtual

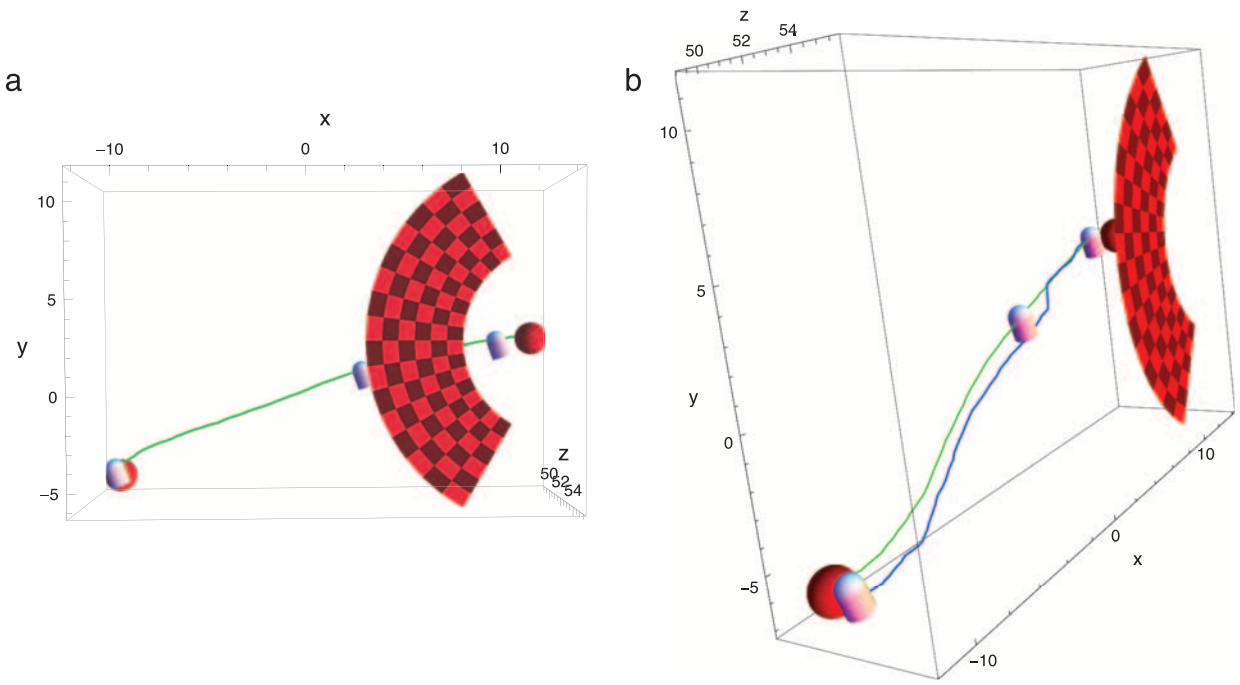

Figure 2. A trial sequence. (a) Subjects start from a fixed starting point in their right field of view and moved their finger toward a target ball 30 cm away. At the initial stage of the movement, the displayed finger (virtual finger) coincided with the unseen real finger (the green line). The point of view here is from above, not from the subject's point of view. (b) In perturbed trials, the position of the virtual finger was changed (the blue line) when it was behind the occluder. Subjects would have to compensate for the perturbation to consistently reach the target. The scene is drawn from a side view to show the otherwise occluded perturbation.

reality setup. The stimuli and visual feedback were displayed on a CRT monitor and reflected into the workspace by a half-silvered mirror. The monitor had a resolution of 1152 × 768 pixels and a refresh rate of 120 Hz. Subjects in Experiment 1 wore LCD shutter glasses (CrystalEye, RealD, CA) and viewed the scene binocularly. Subjects in Experiment 2 viewed the scene with their left eye and covered their right eye with an eye patch. The scene was drawn in red to take advantage of the fast decay time of the red phosphor of the CRT display and minimize cross-talk between the two eyes in Experiment 1. Subjects could not see their hands during the experiment.

The starting ball was fixed to a platform in the right field of the subjects' visual space and the target ball was moved by a Denso V-series robot (Aichi, Japan). Both balls

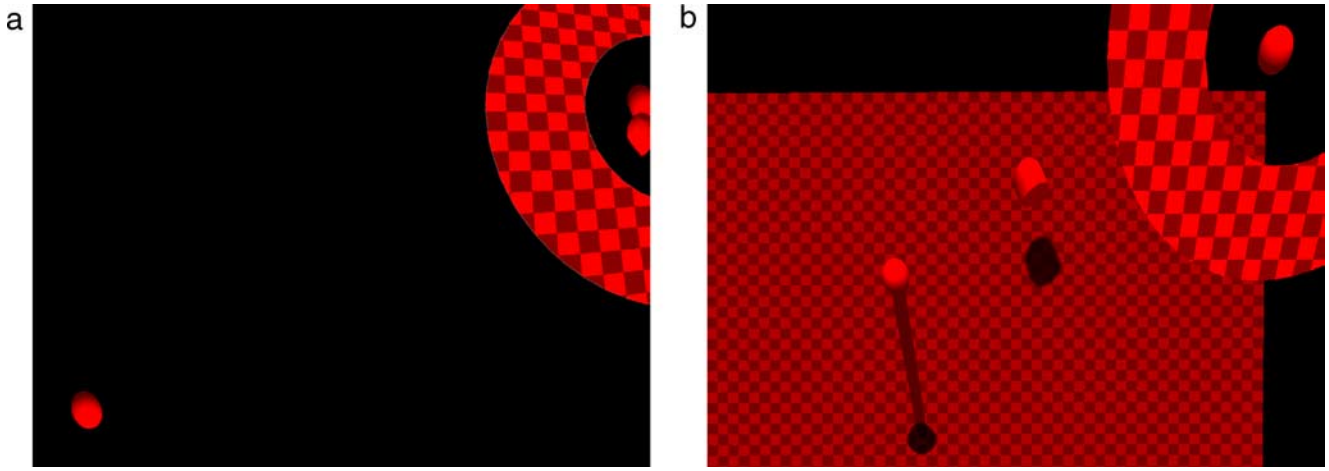

Figure 3. Visual stimuli. (a) In Experiment 1, subjects saw the virtual finger, the starting ball (on the right side), the target ball, and the occluder, rendered to both eyes. (b) In Experiment 2, the target ball was displayed on top of a pole. Subjects also saw the cast shadows of the objects. The scene was rendered only to the left eye.

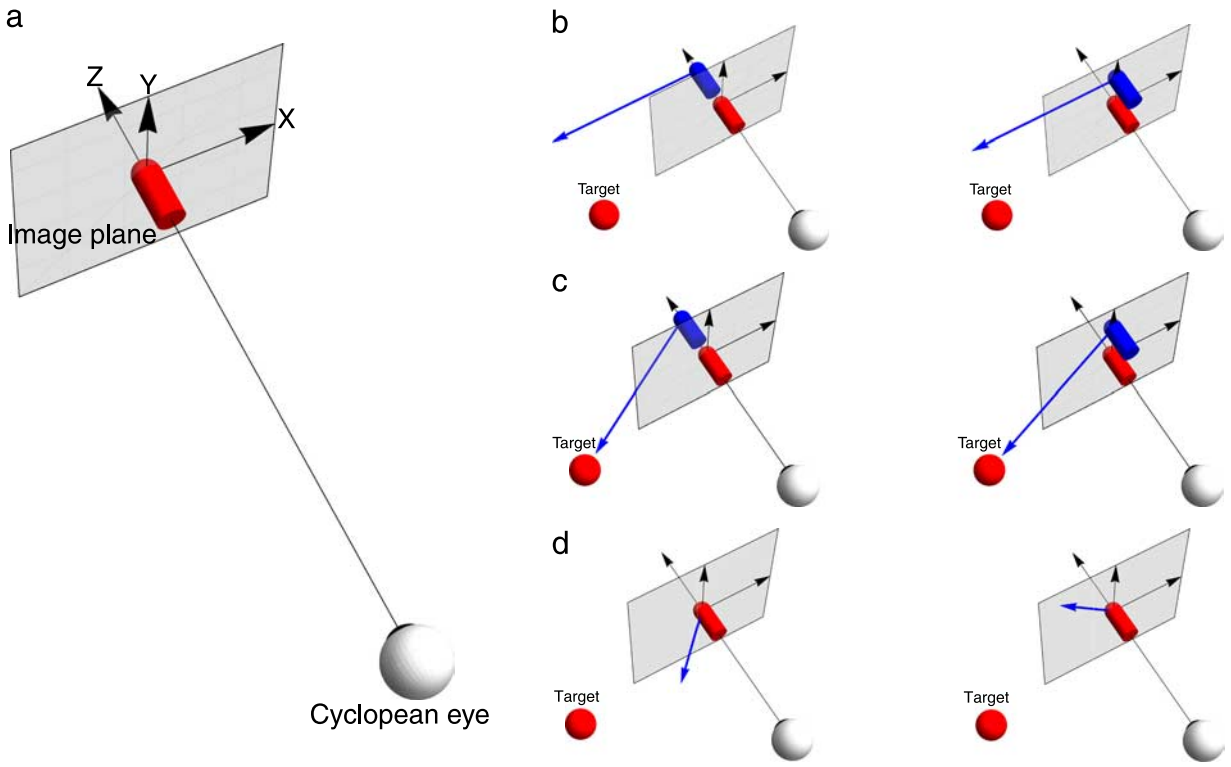

Figure 4. Experiment conditions. (a) All perturbations were applied in a local coordinate frame. The depth or the *z*-axis is defined by the cyclopean eye and the current real finger position. The *x*-axis is horizontal in the image plane and the *y*-axis is defined by the right-hand rule. In-depth perturbations were added along the *z*-axis and in-image perturbations along the *y*-axis. The blue arrows in the figure show the motion of the virtual image of the finger when the motion of the real finger is directly toward the target. (b) Step perturbations added a fixed amount (±1 cm) relative to the current real finger position (left) in depth or (right) in the image plane. (c) Rotation perturbations (left) in depth and (right) in the image plane were ±1 cm when the finger came out of the occluder and decreased so that it would be 0 when the finger reached the target. (d) Direction perturbations (left) in depth and (right) in the image plane were 0 when the finger emerged from the occluder and increased so that the virtual finger would be ±1 cm away from the target in the appropriate dimension.

were rendered in the virtual display as shaded spheres with their true physical sizes and locations (Figure 3). The center-to-center distance of the start and target balls was always 30 cm. The position of the starting ball was fixed across all trials and was designated the origin of the *frontal* plane, whose normal was the vector from the cyclopean eye to the center of the reflected screen and the horizontal axis was parallel to the vector from the left eye to the right eye. The position of the target ball was randomly chosen from a patch of the sphere centered at the starting ball with a radius of 30 cm. The patch was centered on a point in the frontal plane 28.2 cm to the left of the starting position and 10.3 cm above the midline of the display. The patch spanned ±7.5 degrees in azimuth and elevation around the center point, translating into a range in the image plane and in depth of approximately ±4 cm.

Three infrared markers were attached to a metal splint worn on the subjects' index finger and the position of the markers was tracked by an Optotrak 3020 system (NDI, Ontario, Canada) at 120 Hz. The position and orientation (or pose) of the finger was calculated from the tracking data in real time. The virtual finger was rendered as a half-sphere with a radius of 0.85 cm on top of a cylinder with a length of 1 cm, about the same size of the metal splint worn by subjects on their right index finger. It was rendered at the measured 3D position and orientation of a subjects' finger, except when it was perturbed from this position on perturbation trials. To compensate for the delay caused by the tracking system and the computer rendering loop, the pose of the displayed finger was determined by linearly extrapolating for 17 ms (2 frames) from the latest 25 ms (3 frames) of marker data.

An annular-shaped virtual occluder was displayed 5 cm above the frontal plane. At this distance, the finger was behind the occluder in a natural movement. The virtual occluder was centered at the starting ball with an inner radius of 5 cm and a width of 5 cm. With this configuration, the finger disappeared behind the occluder about 1/6 of the total distance to the target and emerged from behind the occluder about 1/3 of the total distance.

The starting and target balls were made of aluminum, with a radius of 0.94 cm. The balls and the splint were

connected in a contact detection circuit. When the finger touched either ball, it closed one of the two circuit loops and the state of the circuit was sampled at the same rate (120 Hz) as the infrared markers by the ODAU module of the Optotrak system. This provided a direct measure of the beginning and end of a movement.

Procedures

Each subject completed five 1-h sessions on separate days, each of which consisted five to eight blocks, depending on how fast the subject could perform the task. Each block had thirty baseline trials and twenty perturbed trials of all 4 types of perturbations. The trials were randomly intermixed under the constraint that no two consecutive trials were of the same perturbation type.

At the beginning of each session, subjects performed a geometric calibration procedure by repeatedly moving an Optotrak marker on a flat tabletop and aligning the marker to reference points displayed on the monitor. Geometric transformations among the coordinate frames associated with the Optotrak, the computer monitor, and the subjects' eyes were computed from the procedure. The information was used to render the scene with correct perspective and binocular disparity to both eyes.

Subjects initiated a trial by touching and resting on the starting ball with their index finger. Once the contact detection circuit signaled that the finger was on the ball, the robot moved the target ball to a randomly chosen position and the virtual target was displayed on the screen. Subjects started the movement on an auditory cue. Data collection started when the finger left the starting ball. Subjects were instructed to move to reach for the target at their natural pace. A successful trial ended when the finger touched the target ball within 500–1200 ms. If the finger did not touch the target within the time limit but was close (within 3 cm) to the target at any time during the movement, the trial was recorded as a miss. Otherwise, the trial was treated as invalid and recorded data were discarded. A text message would inform the subject to speed up or slow down accordingly. On a successful trial, subjects saw the virtual finger touching the virtual target and also felt the physical contact between the splint and the target ball. The virtual target disappeared upon contact and the program waited for the subject to initiate the next trial.

Data processing

We removed various irregularities before analyzing the data. We used linear interpolation to fill in missing frames in the recorded marker data if necessary. However, if there were 3 or more consecutive missing frames, the trial was discarded. We removed the slowest and fastest 10% of the trials. We then removed all of the remaining trials that had a maximal acceleration of 57 m/s^2 or higher. These were typically trials in which subjects rushed to the target and searched around the target before the trial time expired.

We measured the timing and strength of subjects' responses to the perturbation using an autoregressive (AR) model. Let $t = 0$ be the time (in Optotrak frames) that the virtual finger emerged out of the occluder and the finger position at time t be x_t; thus, we have

$$x_t = \sum_{i=1}^{n} a_i x_{t-i} + u_t + \varepsilon_t, \quad (1)$$

where a_i ($i = 1 \ldots n$) are the coefficients of the AR model and u_t is a constant term. The residual ε_t has zero mean by construction. We fitted the AR model from baseline trials and computed the raw perturbation influence function w_t of each perturbation type from

$$x_t = \sum_{i=1}^{n} a_i x_{t-i} + u_t + w_t p_t, \quad (2)$$

where p_t is the amount of perturbation (set to +1 or −1). The raw influence function and the residual noise were smoothed by a causal exponential filter, $f(t) = e^{\lambda t} / \lambda$, $t < 0$, with the time constant $\lambda = 37$ ms.

We used the following procedure to estimate each subject's reaction time to begin correcting for perturbations. For each subject, we estimated the standard error of the smoothed influence functions under the null hypothesis of no correction ($w_t = 0$) using a bootstrap procedure in which we fit Equation 2 to resampled data from the no perturbation trials with the perturbation (p_t) set randomly to ±1 on each trial. The standard deviation of the resulting bootstrapped estimates of the smoothed influence functions provides a measure of the standard deviation of weights expected under the null hypothesis that subjects did not correct—technically, it is the standard deviation of the values of w_t one would expect if subjects did not begin correcting at a time less than or equal to t. We used as our measure of reaction time the first time at which a smoothed perturbation function deviated from 0 by one standard error and remained more than one standard error away from 0 for more than 15 frames (125 ms). Reaction time greater than 40 frames (333 ms) indicated a late correction and we considered the subject as not responding to the perturbation.

The response at time t, $c(t)$, was given by the time integral of w_t, $c(t) = \sum_{i=1}^{t} w_i$. The result gives a normed measure of subjects' responses, where a response of 1 corresponds to a 1-cm deviation in movements with perturbations from movements on unperturbed trials. We, therefore, express subjects' responses in centimeters. The magnitude of the response at the end of the movement is computed by summing the weights up to the point that subjects first touched the target.

We used a 10th order (n = 10) AR model in the analysis, though any model above 4th order gave similar results.

Results

Experiment 1: Binocular conditions

Subjects performed well in the binocular task and the average miss rate of all eight subjects was 6.8%. The ratio of the missed trials of each perturbation type was similar to the overall ratio, so the perturbations themselves were not the cause of the misses.

All subjects responded to the perturbations in the image plane, which was consistent with the findings in Saunders and Knill (2004). The mean reaction time to the step perturbation was 173 ms (*SEM* 11 ms) and that to the rotation perturbation was 187 ms (*SEM* 12 ms). All subjects responded to the step perturbation in depth with an average reaction time of 230 ms (*SEM* 12 ms). Only 2 of the 8 subjects showed responses to the rotation perturbation in depth and their average reaction time was 175 ms (*SEM* 13 ms).

Figure 5 shows the average influence functions of all subjects. The apparent divergence from 0 of the influence function for the in-depth rotation perturbation was totally driven by the 2 subjects that responded to the perturbations (Figure 6); however, even those subjects responded more strongly to the rotation perturbations in the image plane. Within the in-image and in-depth conditions, subjects showed stronger responses to the step perturbations than the rotation perturbations, which agreed with the previous findings (Saunders & Knill, 2004). Figures 5 and 6 also show the transitory effect to the rotation perturbations, both in the in-image and in-depth conditions.

Figure 7a plots the average response to perturbations as a function of the distance to the target. For step perturbations, negative responses reflect appropriate corrections to bring the finger closer to the target at the end of the movement. For rotation perturbations, responses (positive or negative) result in larger endpoint errors. The mean endpoint correction of the in-image step perturbation was 0.53 cm, with 95% confidence interval of [0.49, 0.57], corresponding to a 53% correction, while it was 0.3 cm (95% confidence interval of [0.24, 0.36]) for the in-depth perturbations (Figure 7b).

No subject noticed the perturbations as reported in the exit debriefing, even after we explained what the perturbations were. Fixation during the task was not

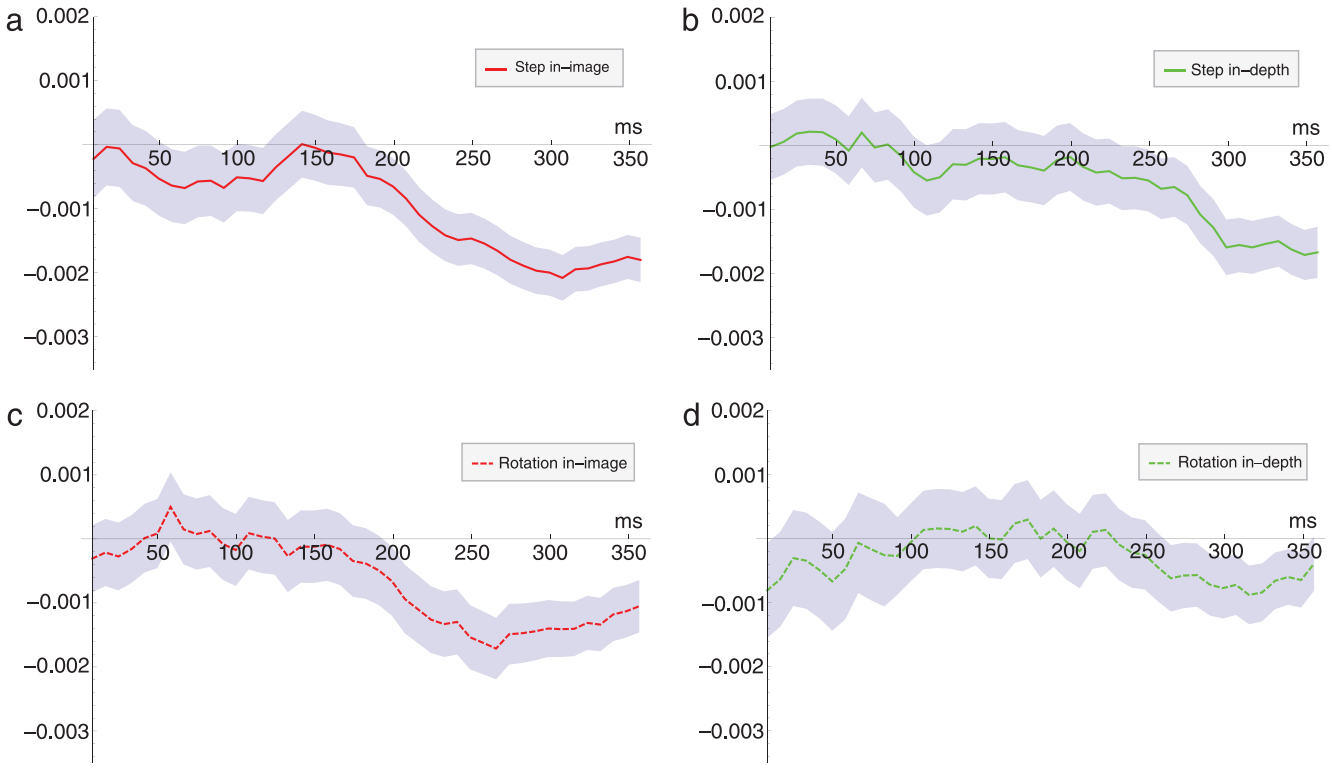

Figure 5. Average influence functions of all the subjects in Experiment 1. Error bands are ±1 *SE*. The reaction time was defined when the influence function deviated from 0 by 1 *SE*. Time 0 was when the virtual finger emerged from the occluder. The four panels (a–d) correspond to the four perturbation conditions. The response to the in-depth perturbation was driven by the data from two subjects (see Figure 6).

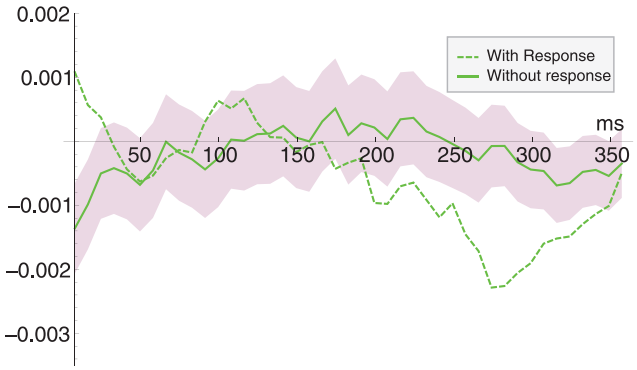

Figure 6. Average influence functions of the 2 subjects who responded to the in-depth rotation perturbation (dotted line) and the 6 subjects who did not (solid line). Note that the reaction time was defined as the first time that the influence function was different from 0 by 1 standard error and remained different for 125 ms. Under this definition, the 6 subjects did not show response to the rotation perturbation in depth.

enforced by an eye tracker, but all subjects reported to have fixated at the target.

Discussion

Assuming subjects indeed fixated on the target, the average eccentricity when the finger came out of the occluder was 17 degrees and the average pedestal disparity was 114 arcmin (crossed), corresponding to a position approximately 2.8 cm nearer to the observer than the geometric horopter, though this varied from trial to trial with a standard deviation of 126 arcmin. Though stereoacuity values are highly dependent on the measurement method and stimuli, and there are no direct data available at said eccentricity and pedestal disparity, extrapolating one measurement (Howard & Rogers, 1995), the threshold disparity difference for depth discrimination would be expected to be above 400 arcmin. The stereo acuity is, thus, very low compared to the acuity of position judgments on the retina—approximately 51 arcmin at that eccentricity—computed using an estimated Weber fraction for position estimates of 0.05 (Burbeck, 1987; Burbeck & Yap, 1990; Whitaker & Latham, 1997). Because of the markedly low acuity of binocular disparities in the periphery, it would take longer for an optimal controller to integrate the disparity signals into its estimate of finger position, explaining why subjects responded slower and more weakly to the in-depth step perturbation than the in-image one. A similar difference in delay was observed by Brenner and Smeets (2006) as well. An alternative explanation of the apparent difference in the delay for corrective responses is that the visual system takes longer to process information about depth than position signals in the retinal plane. Measurements of the time constant for integrating inputs from the two eyes for stereopsis (Ludwig, Pieper, & Lachnit, 2007; Ogle, 1963) put the latency of binocular depth processing in perceptual tasks at 50 ms to 100 ms, which could explain the observed differences in delay. However, a recent study found that the latency can be much lower, about 14 ms, in visuomotor tasks (Wilson, Pearson, Matheson, & Marotta, 2008). Since the two factors lead to similar effects on performance, we cannot sort out from the current data the relative contribution of each one to the difference in apparent delay.

Six of the eight subjects did not respond to the rotation perturbations. Given that they responded to the step perturbations, with the same initial sizes (1 cm), it seems plausible that those subjects relied mainly on the motion of the finger in depth relative to the target and/or

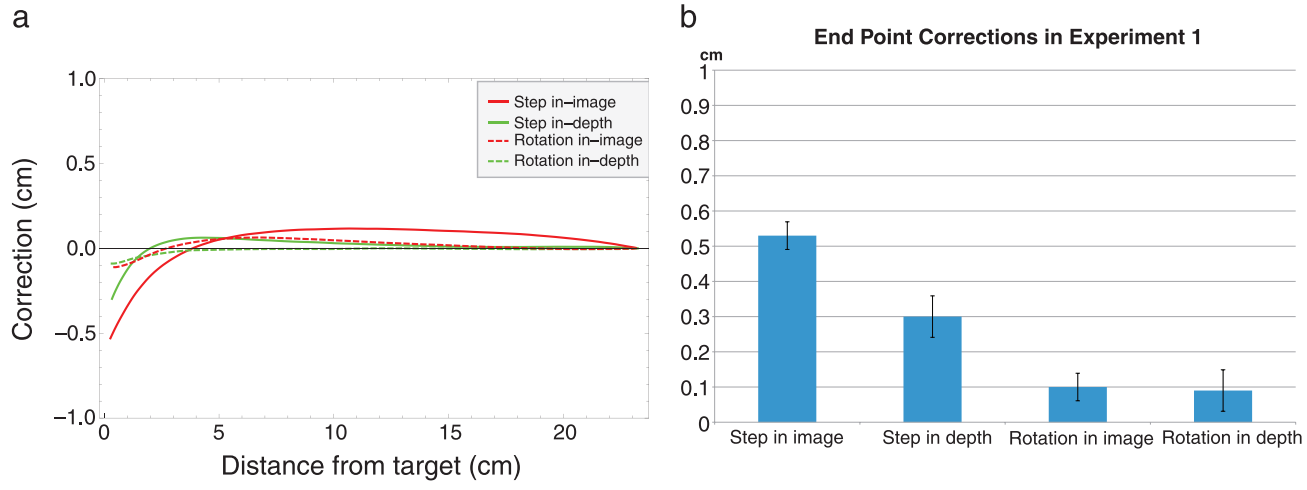

Figure 7. Responses to perturbations in Experiment 1. Negative responses reflect corrections that bring the finger closer to the target for step perturbations. For rotation perturbations, responses presumably move the finger away from the target at the end. (a) The average response to perturbations during the movement. (b) Plots of the endpoint corrections for all the conditions. Error bars are 95% confidence intervals.

the change in relative disparity to reach for the target (a homing strategy) in contradiction with the findings of Brenner and Sweets (2006). Two considerations argue against this. First, those same subjects showed a response to the rotation perturbations in the image plane. One would, therefore, have to posit that the CNS uses a different strategy for controlling hand movement in depth and for controlling hand movement in the image plane. On the face of it, this seems unlikely, as the distinction between depth and image plane dimensions does not map naturally to the motor system being controlled.

Second, Saunders and Knill (2004) have shown that for the image plane perturbations, the subjects' behavior is consistent with an optimal controller that simultaneously estimates the position and velocity of the finger and uses these estimates as input to a controller that generates online motor commands. While such a system necessarily displays corrective responses to simple position shifts of the visual feedback, its responses to rotational perturbations depends on the relative contributions of the sensory signals for position and velocity to their internal estimates. As the relative uncertainties of those signals change, the response to the rotation perturbations changes. As was found here, Saunders and Knill found that a simple rotation of visual feedback in the image plane led to fast initial responses in a direction opposite to the position shift in the visual finger position created by the rotation. The response was weaker than the response to a position shift perturbation, consistent with a model in which sensory information about the motion of the finger contributes to internal estimates of finger motion, mitigating the effects of the position shift signaled by a rotation perturbation. They found that when the visual feedback was perturbed by a combination of a rotation and a shift so as to initially create a feedback signal shifted 1 cm in one direction away from the finger but with movement toward a point shifted 1 cm away from the target in the *opposite* direction (for which subjects should correct in the same direction as the positional shift), responses to the perturbation were delayed by over 100 ms relative to the responses to simple rotation perturbations. Simulation results showed this pattern to be consistent with the known sensory noise parameters on position and velocity signals; that is, early in the movement, the sensory position signal contributes more strongly to internal estimates of finger position than do velocity signals, so that the motion feedback had to be doubled relative to the position feedback to effectively cancel out the responses.

Given the consistency of subjects' behavior in response to image plane perturbations with an optimal feedback controller constrained by sensory noise on position and motion signals, it is natural to interpret the current data to indicate that the sensory noise on position and motion signals in depth for 6 of the eight subjects are such that the two signals effectively are given equal weight when integrated into internal estimates of finger position and motion, while for two of the subjects, the sensory noise on motion in depth is effectively higher than the sensory noise on position in depth (relative to the noise levels in the other subjects). Since we do not have good data on the relative uncertainties of those signals in depth as we do for signals in the image plane, we are unable to parameterize an optimal feedback control model to test these predictions. The fact that six subjects showed early corrective responses for step perturbations but not for rotational perturbations provides strong evidence that the CNS uses not only sensory information about depth but also about motion in depth for online control.

Experiment 2: Monocular conditions

Subjects performed worse in the monocular conditions than in the binocular conditions; the average miss rate of all subjects was 17.7%, significantly higher than that in the binocular conditions in Experiment 1 (6.8%). Subjects were also 23% more likely to miss in the trials with in-depth perturbations than the in-image ones (22.2% for in-depth perturbations and 18.1% for image plane perturbations).

All subjects responded to the in-image perturbations. The reaction time for the step perturbation was 173 ms (*SEM* 12 ms) and that of the direction perturbation was 221 ms (*SEM* 19 ms; see Figure 8). The average amount of correction to the in-image step perturbations was 0.73 cm, with 95% confidence interval of [0.65, 0.81] and that to the in-image direction perturbations was 0.51 cm, with 95% confidence interval of [0.45, 0.57]. Subjects' corrections to in-image step perturbations were larger than we found in Experiment 1, but this can be explained by the increased movement duration in Experiment 2—777 ms compared to 659 ms in Experiment 1. Subjects showed no significant response to the perturbations in depth (see Figures 8 and 9). Average corrections to the step and rotation perturbations in depth were 0.01 cm, with 95% confidence interval of [−0.068, 0.088], and 0.03 cm, with 95% confidence interval of [−0.048, 0.108], respectively. The 95% confidence bounds on these estimates put the maximal corrections at 0.088 and 0.108; thus, while we cannot conclude definitively that subjects did not correct for the perturbations in depth, we can conclude that any corrections they made were very small in proportion to the size of the perturbations.

No subject noticed the perturbations and no subject reported to have used a simple homing strategy by aligning the shadow of the finger to the shadow of the target ball.

Discussion

The reaction time to the in-image step perturbation was the same as that in Experiment 1. The reaction time to the in-image direction perturbations was slightly longer,

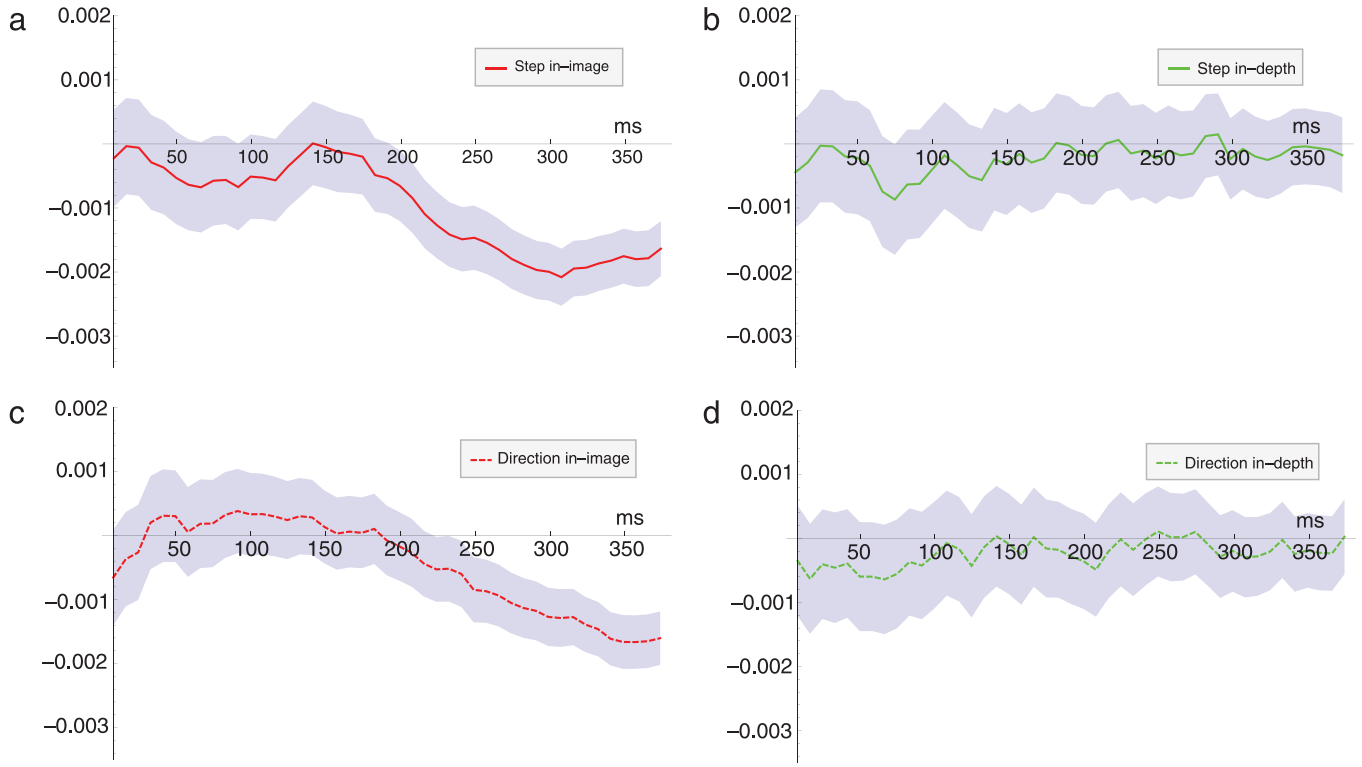

Figure 8. Average influence functions of all the subjects in Experiment 2. Error bands are ±1 *SE*. The four panels (a–d) correspond to the four perturbation conditions. That the influence functions of the in-depth perturbations were not significantly different from 0 showed that subjects did not respond to the in-depth perturbations.

which is consistent with the results in Saunders and Knill (2004). Subjects showed no significant response to the perturbations in depth. This suggests that none of the monocular cues present in the display, including cast shadows, are effective online cues for online control of fast pointing movements.

Several factors may hinder the use of shadows as a reliable online depth cue. Although in theory, given that the light source is directional, its direction can be solved from just two frames of finger–shadow correspondences, in practice, the solution involves the intersection of the vectors connecting those correspondences and the estimate can be very noisy. Strong prior assumptions about light source directions can also negatively affect how people estimate depth from cast shadows. Though we put the virtual light source above the subjects' heads, Mamassian

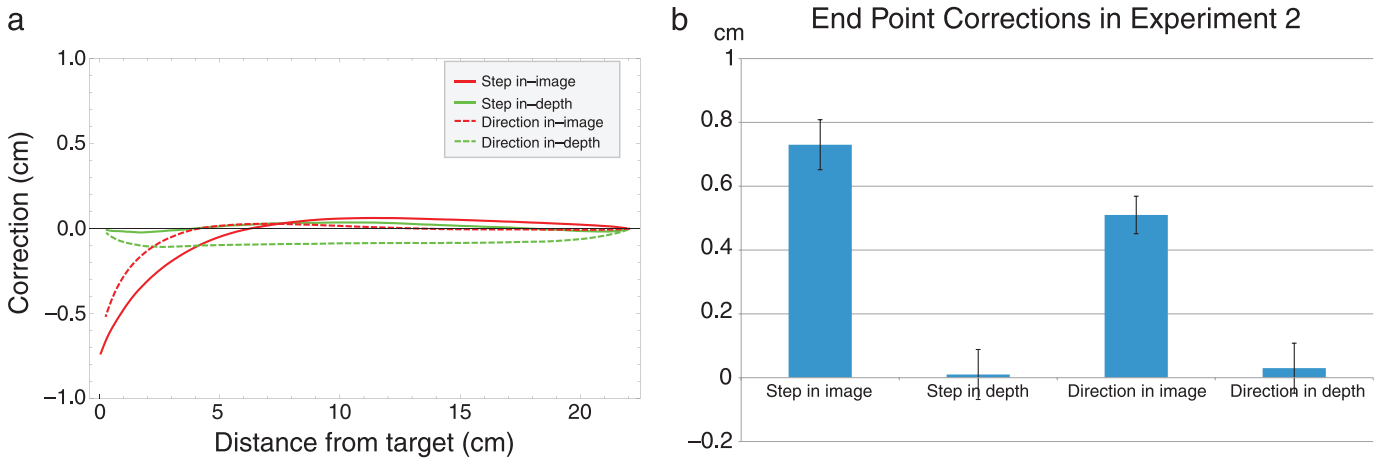

Figure 9. Responses to perturbations in Experiment 2. As before, negative responses reflect corrections that improve pointing accuracy. (a) The average amount of correction during the movement. (b) Subjects made little corrections to the in-depth perturbations and the endpoint corrections were not significantly different from 0 (error bars are 95% confidence intervals). Subjects corrected more to the in-image perturbations than to the same perturbations in Experiment 1.

et al. reported that the prior on light source position can be biased toward the left by as much as 26 degrees (Mamassian & Goutcher, 2001). Given a strong and likely biased prior and noisy visual input, it could take an optimal estimator very long to infer the 3D position of the finger from its shadows, rendering cast shadows ineffective as an online depth cue. That subjects did not use shadows as an online visual cue does not mean they cannot use them to reach for the target at the very end. Studies on how shadows help in motor tasks are sparse, but everyday examples abound.

It is always a concern that the scene was not rendered realistically enough for subjects to trust the shadow information. For example, since we used a directional light source, no penumbrae were rendered, while in real world penumbrae are always present. This, however, does not change the underlying geometry; furthermore, the effectiveness of shadows in perceptual tasks is remarkably robust to the realism of the rendering (Mamassian et al., 1998).

General discussion

All of the reaction times to the in-image perturbations are somewhat longer than those of corresponding perturbations in Saunders and Knill (2004; 143 ms for rotation and 146 ms for step perturbations) and all of the reaction times to the in-depth perturbations are longer than that in Brenner and Smeets (2006). This can be attributed to at least two factors. First, the size of the perturbations is only half of those used in Saunders and Knill (1 cm vs. 2 cm) and much smaller than those in Brenner and Smeets. Given the conservative statistical techniques for detecting the beginning of corrective responses, smaller perturbations (hence, smaller corrective responses) necessarily result in later estimates of response times. Second and perhaps more importantly, the current task was different from previous ones. Subjects had to bring the speed of the finger to near zero when approaching the target ball for an accurate touch, whereas in Saunders and Knill subjects could hit the tabletop at any speed, knowing they would be stopped. Subjects in Brenner and Smeets were required to hit virtual targets with a virtual cursor and had no endpoint speed constraint either. Liu and Todorov (2007) observed that when subjects were required to stop at the target, the movement duration was longer than simply asked to hit the target at arbitrary speed. This trade-off between endpoint stability and accuracy is nicely captured by an optimal feedback control model. The model also predicts that the amount of correction is smaller in the stop-at-target case, which explains why the corrections to the step perturbations in both experiments (53% and 73%, respectively) are smaller than that (80%) in Saunders and Knill.

While we have focused our discussion on cast shadows, the apparent size of the finger (and how it changes over time) is another potentially salient monocular cue to depth and motion in depth. With the average distance between the fingertip and the cyclopean eye at 50 cm in a natural pointing task and size of the finger at 1.7 cm, a 1-cm perturbation in depth corresponds to about 2 arcmin of size change. Considering that the Weber fraction of angular size judgment threshold is around 6% (Mckee & Welch, 1992), or 7 arcmin under our experiment condition, the uncertainty in the size cue renders it effectively useless for online control.

The results suggest that binocular disparities are indispensable for fast accurate pointing movement, not only because they are important for estimating the depth of a target object but also because they are used by the CNS for online feedback control of the moving hand. While relevant studies do not dissociate the contribution of stereopsis to estimates of target depth and online control, it is clear that stereoscopic function significantly impacts performance on close-range motor tasks (O'Connor, Birch, Anderson, Draper, & FSOS Research Group, 2010) and people with poor stereoscopic vision such as amblyopes perform worse on tasks like grasping than normals (Grant, Melmoth, Morgan, & Finlay, 2007). In general, stereoacuity has direct impact on how well we perform close-range accurate motor tasks (O'Connor et al., 2010).

Conclusions

The experiments show that the CNS uses stereoscopic depth information about the moving hand for online control of pointing movements. Monocular depth cues, however, appear ineffective for online control, suggesting that stereoscopic disparities dominate online control of hand movements in depth. That online depth information has less of an influence in online control than information about the position of the hand in the image plane is consistent with what one would expect from an optimal controller that weighs the sensory input for estimating hand state in proportion to the reliability of the signals.

Acknowledgments

This work was supported by NIH Grant R01 EY017939 to the second author.

Commercial relationships: none.
Corresponding author: Bo Hu.
Email: bhu@cvs.rochester.edu.
Address: Center for Visual Science, University of Rochester, P.O. Box 270270, Rochester, NY 14627, USA.

References

Bradshaw, M. F., & Elliott, K. M. (2003). The role of binocular information in the on-line control of prehension. *Spatial Vision, 16,* 295–309.

Brenner, E., & Smeets, J. B. J. (2003). Fast corrections of movements with a computer mouse. *Spatial Vision, 16,* 365–376.

Brenner, E., & Smeets, J. B. J. (2006). Two eyes in action. *Experimental Brain Research, 170,* 302–311.

Burbeck, C. A. (1987). Position and spatial frequency in large-scale localization judgments. *Vision Research, 27,* 417–427.

Burbeck, C. A., & Yap, Y. L. (1990). Two mechanisms for localization? Evidence for separation-dependent and separation-independent processing of position information. *Vision Research, 30,* 739–750.

Grant, S., Melmoth, D. R., Morgan, M. J., & Finlay, A. L. (2007). Prehension deficits in amblyopia. *Investigative Ophthalmology & Visual Science, 48,* 1139–1148.

Greenwald, H. S., & Knill, D. C. (2009a). A comparison of visuomotor cue integration strategies for object placement and prehension. *Visual Neuroscience, 26,* 63–72.

Greenwald, H. S., & Knill, D. C. (2009b). Cue integration outside central fixation: A study of grasping in depth. *Journal of Vision, 9*(2):11, 1–16, http://www.journalofvision.org/content/9/2/11, doi:10.1167/9.2.11. [PubMed] [Article]

Greenwald, H. S., Knill, D. C., & Saunders, J. A. (2005). Integrating visual cues for motor control: A matter of time. *Vision Research, 45,* 1975–1989.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis.* New York: Oxford University Press.

Izawa, J., & Shadmehr, R. (2008). On-line processing of uncertain information in visuomotor control. *Journal of Neuroscience, 28,* 11360–11368.

Jackson, S. R., Jones, C. A., Newport, R., & Pritchard, C. (1997). A kinematic analysis of goal-directed prehension movements executed under binocular, monocular, and memory-guided viewing conditions. *Visual Cognition, 4,* 113–142.

Kersten, D., Knill, D. C., Mamassian, P., & Bulthoff, I. (1996). Illusory motion from shadows. *Nature, 379,* 31.

Kersten, D., Mamassian, P., & Knill, D. C. (1997). Moving cast shadows induce apparent motion in depth. *Perception, 26,* 171–192.

Knill, D. C. (2005). Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception. *Journal of Vision, 5*(2):2, 103–115, http://www.journalofvision.org/content/5/2/2, doi:10:1167/5.2.2. [PubMed] [Article]

Knill, D. C., & Kersten, D. (2004). Visuomotor sensitivity to visual information about surface orientation. *Journal of Neurophysiology, 91,* 1350.

Liu, D., & Todorov, E. (2007). Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *Journal of Neuroscience, 27,* 9354–9368.

Loftus, A., Servos, P., Goodale, M. A., Mendarozqueta, N., & Mon-Williams, M. (2004). When two eyes are better than one in prehension: Monocular viewing and endpoint variance. *Experimental Brain Research, 158,* 317–327.

Ludwig, I., Pieper, W., & Lachnit, H. (2007). Temporal integration of monocular images separated in time: Stereopsis, stereoacuity, and binocular luster. *Perception & Psychophysics, 69,* 92–102.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition, 81,* B1–B9.

Mamassian, P., Knill, D. C., & Kersten, D. (1998). The perception of cast shadows. *Trends in Cognitive Sciences, 2,* 288–295.

Mckee, S. P., & Welch, L. (1992). The precision of size constancy. *Vision Research, 32,* 1447–1460.

O'Connor, A. R., Birch, E. E., Anderson, S., Draper, H., & FSOS Research Group. (2010). The functional significance of stereopsis. *Investigative Ophthalmology & Visual Science, 51,* 2019–2023.

Ogle, K. N. (1963). Stereoscopic depth perception and exposure delay between images to the two eyes. *Journal of the Optical Society of America, 53,* 1296–1304.

Prablanc, C., & Martin, O. (1992). Automatic control during hand reaching at undetected two-dimensional target displacements. *Journal of Neurophysiology, 67,* 455.

Sarlegna, F., Blouin, J., Vercher, J.-L., Bresciani, J.-P., Bourdin, C., & Gauthier, G. M. (2004). Online control of the direction of rapid reaching movements. *Experimental Brain Research, 157,* 468–471.

Saunders, J. A., & Knill, D. C. (2003). Humans use continuous visual feedback from the hand to control fast reaching movements. *Experimental Brain Research, 152,* 341–352.

Saunders, J. A., & Knill, D. C. (2004). Visual feedback control of hand movements. *Journal of Neuroscience, 24,* 3223–3234.

Saunders, J. A., & Knill, D. C. (2005). Humans use continuous visual feedback from the hand to control both the direction and distance of pointing movements. *Experimental Brain Research, 162,* 458–473.

van Mierlo, C. M, Louw, S., Smeets, J. B. J., & Brenner, E. (2009). Slant cues are processed with different latencies for the online control of movement. *Journal of Vision, 9*(3):25, 1–8, http://www.journalofvision.org/content/9/3/25, doi:10.1167/9.3.25. [PubMed] [Article]

Whitaker, D., & Latham, K. (1997). Disentangling the role of spatial scale, separation and eccentricity in Weber's law for position. *Vision Research, 37,* 515–524.

Wilson, K. R., Pearson, P. M., Matheson, H. E., & Marotta, J. J. (2008). Temporal integration limits of stereovision in reaching and grasping. *Experimental Brain Research, 189,* 91–98.