

A comparison of visuomotor cue integration strategies for object placement and prehension

HAL S. GREENWALD AND DAVID C. KNILL

Center for Visual Science, University of Rochester, Rochester, New York

(RECEIVED February 12, 2008; ACCEPTED April 16, 2008)

Abstract

Visual cue integration strategies are known to depend on cue reliability and how rapidly the visual system processes incoming information. We investigated whether these strategies also depend on differences in the information demands for different natural tasks. Using two common goal-oriented tasks, prehension and object placement, we determined whether monocular and binocular information influence estimates of three-dimensional (3D) orientation differently depending on task demands. Both tasks rely on accurate 3D orientation estimates, but 3D position is potentially more important for grasping. Subjects placed an object on or picked up a disc in a virtual environment. On some trials, the monocular cues (aspect ratio and texture compression) and binocular cues (e.g., binocular disparity) suggested slightly different 3D orientations for the disc; these conflicts either were present upon initial stimulus presentation or were introduced after movement initiation, which allowed us to quantify how information from the cues accumulated over time. We analyzed the time-varying orientations of subjects' fingers in the grasping task and those of the object in the object placement task to quantify how different visual cues influenced motor control. In the first experiment, different subjects performed each task, and those performing the grasping task relied on binocular information more when orienting their hands than those performing the object placement task. When subjects in the second experiment performed both tasks in interleaved sessions, binocular cues were still more influential during grasping than object placement, and the different cue integration strategies observed for each task in isolation were maintained. In both experiments, the temporal analyses showed that subjects processed binocular information faster than monocular information, but task demands did not affect the time course of cue processing. How one uses visual cues for motor control depends on the task being performed, although how quickly the information is processed appears to be task invariant.

Keywords: Cue integration, Reaching, 3D orientation, Task dependence, Stereopsis

Introduction

The human hand is highly versatile and allows us to perform many tasks including eating, using tools, and communicating. Goal-directed hand movements such as object placement, pointing, and prehension are normally performed under visual control, but how visual information is processed and used for guiding these movements may depend on the specific task requirements (Schrater & Kersten, 2000). The experiments presented in this paper tested whether there are dissimilarities between how humans integrate cues about three-dimensional (3D) orientation for controlling the orientation of the hand while picking up objects and placing objects.

Cue integration is an important function of the human visual system because it enables us to take advantage of the multiple types of information provided by our environment about the sizes, shapes, positions, and orientations of the objects around us. Combining the information provided by different cues enables us to

make judgments with less variance than if the estimates were based on single cues alone (Landy et al., 1995; Ghahramani et al., 1997; Knill & Saunders, 2003; Hillis et al., 2004). Experimental evidence suggests that humans integrate information from different cues according to their relative reliabilities in a nearly statistically optimal way, with more reliable cues having greater influences over the combined estimates (Jacobs, 1999; Deneve et al., 2001; Ernst & Banks, 2002; Knill & Saunders, 2003; Alais & Burr, 2004). Cue integration also depends on relative processing speeds; cues that are processed faster have greater overall influences on movements with short durations (Greenwald et al., 2005). Knill (2005) showed a dissociation between how cues were integrated for perceptual and motor tasks, but the reason for this was unclear.

The current study examined whether task demands affect how the brain integrates visual cues to control natural hand movements. Several authors have argued that binocular information is particularly important for grasping based on the observation that subjects grasping objects using binocular vision were faster and more accurate than when binocular disparity was unavailable (Servos et al., 1992; Jackson et al., 1997; Marotta et al., 1997; Watt & Bradshaw, 2000; Bradshaw et al., 2004; Glover, 2004,

Address correspondence and reprint requests to: Hal Greenwald, Center for Visual Science, University of Rochester, 274 Meliora Hall, Box 270270, Rochester, NY 14627-0270. E-mail: hgreenwald@bcs.rochester.edu

Melmoth & Grant, 2006). A grasping model by Smeets and Brenner (1999) proposed one possible reason for this. They suggested that humans grasp objects by independently controlling the trajectories of the thumb and index finger so that they arrive on appropriate points on the object that result in a stable grasp. If prehension depends on identifying these positions in depth, which change with the 3D orientation of an object even when the position of its center is constant, binocular depth cues might be expected to influence hand orientation during grasping more than during object placement because binocular cues provide direct information about the 3D positions of the grasp points. Monocular cues like texture gradient and contour compression do not; they provide direct information about the object's 3D orientation and shape. Thus, since object placement requires subjects to orient their hands to match the 3D orientation of the target surface, one might expect monocular cues to contribute relatively more to the orientation component of movements for object placement than for grasping.

We showed subjects a textured disc whose orientation in depth was specified by monocular and binocular cues, and they either grasped the disc or placed an object on it. In Experiment 1, we had different subjects perform the grasping and object placement tasks and compared how the available visual cues, which were identical between tasks, influenced their hand orientations. Since this experiment showed that subjects who grasped a disc relied more on binocular cues than those who placed an object on the disc, we performed a second experiment to determine whether subjects performing both tasks in separate interleaved sessions would maintain separate cue processing strategies for each. Since we previously found differences in the processing speeds of monocular and binocular cues and that these speeds interacted with cue reliability to determine the overall influence of the cues on subjects' motor performance (Greenwald et al., 2005), we examined the temporal dynamics of cue processing in the different tasks. This allowed us to dissociate effects on cue integration strategies that resulted from differences in task demands from those that resulted from possible task-related differences in cue processing speeds.

Experiment 1

In this experiment, we measured the separate influences of monocular and binocular cues on movements aimed at placing an object on a slanted disc or grasping this same slanted disc. We also performed a temporal analysis of the trajectories of the object in the placement task and those of the subjects' fingers in the grasping task to make inferences about how subjects processed incoming visual information over time. These experiments replicated the object placement task in Greenwald et al. (2005) using stimuli that were smaller and thicker to facilitate grasping.

Method

Subjects

Thirty-one naïve volunteers from the University of Rochester community participated in this study; 15 performed the grasping task, and 16 performed the object placement task. Each had normal or corrected-to-normal vision, was sensitive to binocular disparities of 40 s of arc or less, and was right-handed. Written informed consent was obtained from each volunteer, and subjects were paid \$10 per hour for their participation. The experiment was conducted according to the guidelines set by the University of Rochester Research Subjects Review Board, who approved the study.

Apparatus

Participants viewed a 20-in. computer display through a horizontal half-silvered mirror so that the virtual image of the monitor appeared as a horizontal surface behind the mirror (Fig. 1). An opaque backing placed underneath the mirror during experimental trials prevented the subjects from seeing anything other than the image on the monitor, and opaque covers placed along the outer portions of the mirror masked the edges of the computer monitor and eliminated external visual references. These occluders were well outside of Panum's area (the region within which binocular fusion can occur), so they did not provide effective references for

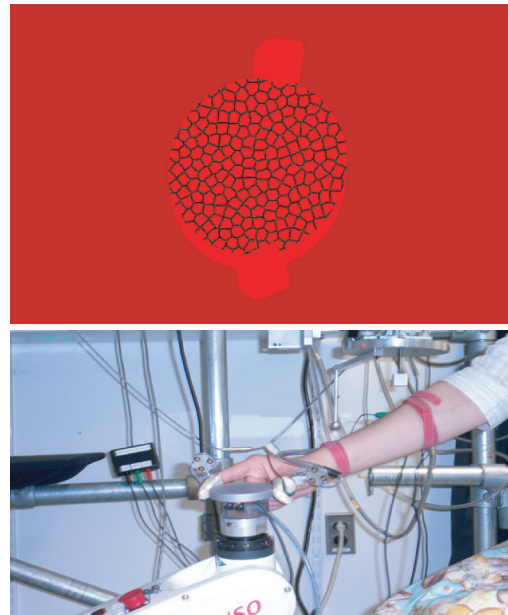
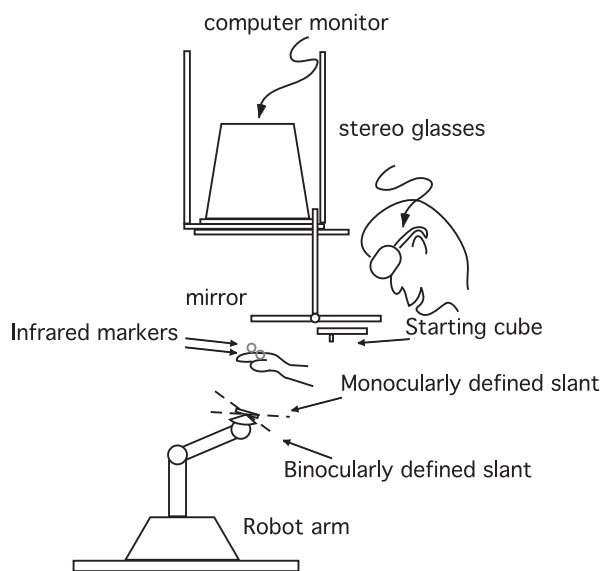


Fig. 1. (Color online) The experimental setup, the virtual disc with fingers visible at target contact, and a photograph of a subject preparing to lift the physical coin (see text for details).

judging 3D position or orientation. Images were displayed at a resolution of 1152×864 pixels at 118 Hz (59 Hz for each eye) through CrystalEyes stereo glasses (StereoGraphics Corporation, San Rafael, CA). A chin rest supported the subject's head and oriented their view approximately 30 deg downward toward the mirror.

Grasping

The main visual stimulus was a large virtual coin textured with a Voronoi pattern. A Denso VS 6-axis robot (Denso Corp., Aichi, Japan) placed an aluminum disc with the same dimensions as the virtual coin and a mass of 116 g at the same position and orientation in the workspace to provide both a clear endpoint for the grasp and a haptic feedback at the end of each trial. Having a physical object for subjects to grasp was important because people can behave differently when they are interacting with an object versus when they are pretending to interact with it (Goodale et al., 1994). The metal disc was fit on top of an optical sensor attached to the end of the robot that identified whether the disc was in place or had been removed, and a Northern Digital Optotrak Data Acquisition Unit II (Northern Digital, Inc., Waterloo, Ontario, Canada) recorded the state of the disc at 120 Hz.

Subjects wore snugly fitting rubber thimbles on their right thumb and index finger that were used to hold rigid metal tabs in place directly above their fingernails. Small, flat metal surfaces on which we mounted three infrared markers per finger to track the positions of the two fingers throughout each movement were rigidly attached to these tabs. The positions of these markers in 3D space were recorded at 120 Hz using a Northern Digital Optotrak 3020. Thimble covers made of electrically conductive fabric (nylon coated with a $2\text{-}\mu\text{m}$ layer of silver; Schlegel Electronic Materials, Rochester, NY) were placed over the rubber thimbles to help identify when subjects touched either the starting position or the physical coin while minimizing interference with tactile feedback. A metal cube used as the starting position and the metal disc were connected to 5 V sources, and wires sewn to the conductive fabric thimbles were connected to the Northern Digital Optotrak Data Acquisition Unit II, which recorded voltages across both the starting cube and the metal disc at 120 Hz to identify when either the thumb or the index finger was in contact with either object.

Object placement

Subjects held an acrylic cylinder (6.4 cm diameter, 12.7 cm height, 227 g mass) that they placed on a surface mounted on the end of the robot arm. The cylinder had four infrared markers mounted in fixed positions in a triangular pattern (three markers formed the base) on one side. The virtual surface was the same coin used in the grasping task, but the physical surface was a larger (15 cm diameter) aluminum disc that reduced the likelihood that subjects would miss the physical target. Subjects began each movement by first resting the cylinder on a square metal platform that was positioned so that the starting positions were similar in both the grasping and the object placement tasks. The corner of the platform furthest from the target had a raised lip to serve as a fixed starting point. An aluminum plate mounted on the bottom of the cylinder provided contact information like the electrically conductive thimble covers used during grasping trials, and 5 V sources were connected to the starting platform and target surface.

Calibration procedures

Spatial calibration of the virtual environment required computing the coordinate transformation from the reference frame of

the Optotrak to the reference frame of the computer monitor and the location of each subject's eyes relative to the monitor. At the beginning of each session, the backing of the half-silvered mirror was removed so that subjects could see the monitor and their hand simultaneously. Subjects positioned an Optotrak marker at a series of visually cued locations so that the marker and a symbol presented monocularly on the monitor appeared to be aligned. Thirteen positions were matched for each eye at two different depth planes, and we calculated the 3D position of each eye relative to the center of the display by minimizing the squared error between the measured position of the marker and the position we predicted from the estimated eye locations. Subjects then moved the marker around the workspace and confirmed that a symbol presented binocularly in depth appeared at the same location as the marker.

A second calibration procedure determined the coordinate transformation between the Optotrak reference frame and the reference frame of the robot arm. Three infrared markers were attached to the end of the robot arm and moved to different locations within three planes that varied in depth relative to the Optotrak. The transformations computed from this procedure and the viewer calibration procedure allowed us to position the robot arm so that the physical target appeared at the same position and orientation as the virtual disc displayed in the workspace during the experiment.

A third calibration procedure performed only for the grasping task determined the locations of the points on each subject's fingertips that they used as the contact points for grasping objects. This required a calibrated device composed of three pieces of metal joined to form a corner and a base. Three infrared markers were attached to one side. Once the rubber thimbles, tabs with infrared markers, and conductive thimble covers were attached to the subject's fingers, he or she placed each finger into the corner formed by the metal pieces, and the positions of the markers on the metal device and the subject's fingers were recorded to determine the locations and orientations of the fingertips. Graphical fingertip markers appeared at the computed positions, and subjects confirmed that the markers lined up with their actual fingertips when they moved them within the workspace. Next, a line appeared between the 90 and the 270 deg positions along the top surface of the disc held by the robot arm, and subjects grasped the disc from the side at the ends of the line using a precision grip as if they were going to pick it up. This was repeated five times, and the measured positions of their fingers relative to the line were averaged to determine the contact point for each finger. Crosses were drawn on the display so that they appeared to be centered at the computed contact points, and subjects verified that these symbols appeared at the appropriate locations on their fingertips.

Stimuli

The target object that subjects grasped was rendered in the workspace as a bright red circular disc (7 cm diameter, 0.9525 cm height) with randomly generated Voronoi textures on its visible face (Fig. 1). The disc appeared to be floating in space in front of a darker red background. Stimuli were drawn in red to take advantage of the monitor's relatively faster red phosphors and prevent interocular cross talk. The Voronoi textures displayed on the disc were broadband in spatial frequency and contained texture elements that offered a clear texture gradient when the coin was slanted in depth. The contours of the surface and the Voronoi textures provided monocular information about the target's orientation, and the disparities between features in the images presented to the eyes provided binocular cues. During the grasping task, red

thimble-shaped markers coaligned with the fingertips appeared in the workspace near the end of each trial after at least one of the fingers had made contact with the target. This real-time visual feedback about the hand was limited to the final portion of each movement to prevent subjects from comparing the disparities of their fingers to the disparity of the coin prior to grasping it. The finger symbols remained visible in the workspace while subjects lifted the coin and disappeared after the coin was replaced on the robot arm. Similarly, in the object placement task, a virtual cylinder that was coaligned with the real cylinder appeared at the end of each trial to provide visual feedback once the bottom of the cylinder had made contact with the physical target. The virtual cylinder remained visible in the workspace until the cylinder was removed from the target surface.

We describe 3D orientation using two angles, *slant* and *tilt*. Slant is the angle between the surface normal and the line of sight and specifies the rotation about the tilt axis. An object with 0 deg of slant would be frontoparallel relative to the viewer. Tilt determines the axis about which the surface is slanted, and it is the angle between the projection of the surface normal into the image plane and a fixed vector along the surface (Witkin, 1981; Stevens, 1983). Our experiments used a tilt of 90 deg, so the objects were slanted about an axis that was parallel to the vector between subjects' eyes, and the top edges of the objects were further from the subject than the bottom edges. In our setup, a slant of about 38 deg would be horizontal in the world.

The virtual coin was displayed at slants of 20–45 deg relative to the viewer in 5 deg increments (six conditions) for grasping and at slants of 25–45 deg (five conditions) for object placement, and there were 64 trials per condition. In cue-consistent trials, the disc was presented at the specified slant, but in cue-conflict trials, the monocular and binocular cues suggested different 3D orientations. Cue conflicts were introduced about a base slant of 35 deg by changing the slant suggested by one or both cue classes by either +5 or –5 deg (six mismatched combinations of 30, 35, and 40 deg). We generated cue conflicts by rendering the disc and texture at the slant specified for the binocular cues and then distorting them so that when they were projected from the binocular slant to the cyclopean view, the contour and texture indicated the slant specified for the monocular cues on that trial. We computed the appropriate distortion by projecting the vertices of the surface and texture into the virtual image plane of a cyclopean view of the disc using the slant for the monocular cues and then back-projecting the transformed vertices onto a plane with the specified binocular slant. On some of the trials (six conditions), the cue conflicts were present when the disc first appeared in the virtual environment and remained unchanged throughout each trial; we used these trials to quantify the cues' overall contributions to the movements. In an additional eight conditions, the disc was initially presented at a slant of 35 deg without cue conflicts, but conflicts appeared after movement onset by changing the slant suggested by one or both cue classes by +5 or –5 deg. When both cue classes were perturbed, they could change in either the same or the opposite directions. The responses to the cue conflicts in these trials, to which we refer as perturbation trials due to the slant manipulations, allowed us to separately analyze the time courses for processing the monocular and binocular information.

The physical target held by the robot was always centered at the same position as the virtual target in the workspace, and the robot oriented the physical target at the average of the slants specified by the monocular and binocular cues plus a random shift of up to ± 2 deg selected from a uniform distribution. The random

shift was added to prevent subjects from using haptic feedback provided by the physical target to learn a dependency on either cue (Ernst et al., 2000). The height of the coin allowed subjects to deviate by up to ± 7.75 deg from the actual alignment of the physical coin and still contact it along its edge.

Procedure

Grasping

Subjects participated in four 1-h sessions, each consisting of four 80-trial blocks, for a total of 1280 trials per subject. We administered practice trials at the start of the first session until subjects understood the task and could perform it correctly. Subjects indicated that they were ready to begin each trial by grasping a metal cube suspended in the work area to their right and slightly in front of them from the right side so that the pad of their thumb was against the side of the cube that faced them and their index finger was against the opposite face. This hand pose helped ensure that the infrared markers attached to the fingers were visible throughout each trial, and the cube served as a common starting point for the movements. After the fingers were stabilized on the cube, the coin appeared in the virtual environment. After 750 ms, a beep instructed participants to begin reaching toward the coin. The coin remained visible until they released the cube, at which point the display flickered at 29.5 Hz between black and white for 10 video frames (169 ms). This visual mask eliminated motion cues that would otherwise have revealed any stimulus perturbations. None of the subjects included in the data set reported noticing the perturbations when questioned after completing their final session. At the end of the mask, the coin reappeared in the workspace either as it was before movement onset or with the specified cue perturbations and remained visible until the end of the trial. The subjects' task in each trial was to grasp the coin along its lower and upper edges (6 o'clock and 12 o'clock) from the right side with their right hand using a precision grip and lift it a few centimeters above its initial position. Symbols indicating the current positions and orientations of the thumb and index finger appeared once one of the fingers made contact with the coin and remained visible until the disc was replaced on its base at the end of the robot arm; visual feedback about the positions of the subjects' fingers was never available until they touched the disc. Typically, the most natural way to pick up a coin would be by approaching from above with the fingers perpendicular to the coin, probably because it allows greater variability in the positions of the fingers. The specific pose subjects used in this study required them to position their fingers accurately along the edges of the coin. Although we did not analyze the lifting portion of the movements, we included it so that subjects would grasp the coin in a more stable manner than they might if only required to touch its edges. Subjects were allowed 2.5 s from the start signal to grasp and remove the coin from its base, but they had to touch the coin within 600–900 ms after releasing the starting cube. The permitted durations still allowed subjects to move at natural speeds. The maximum trial duration of 2.5 s provided subjects time to respond to the beep and to firmly grasp and lift the disc; pilot subjects reported feeling rushed when allowed a maximum time of 2 s. If subjects performed the trial correctly, the virtual disc exploded and disappeared. Otherwise, if subjects released the cube prior to the "go" signal, failed to touch the coin within the specified period of time, or did not remove the coin from its base within 2.5 s after the go signal, the computer aborted the trial, displayed an error message, and repeated the trial later in the same block.

Object placement

Subjects completed four 1-h sessions of an object placement task, each consisting of four 76-trial blocks for a total of 1216 trials per subject. Instead of picking up the target surface as in the previous experiment, subjects placed a cylinder on it. The conditions were the same as for grasping but without the 20 deg cue-consistent trials, which were extraneous. In addition, at the end of every session, we gave subjects feedback about their performance based on the standard deviation (adjusted for the range of orientations by dividing by the slope of the best-fitting line relating the true target orientation to cylinder orientation) of the cylinder orientation on cue-consistent trials. This measured how accurately they performed the object placement task but was not directly related to any of our hypotheses. Subjects who produced standard deviations significantly larger than 5 deg on unperturbed, cue-consistent trials when adjusted for their orientation range were dismissed early from the study. Otherwise, the procedures for the two tasks were identical.

Results

For the grasping condition, we excluded two subjects from the data set who reported noticing the stimulus perturbations and one subject who paused to wait for the visual mask to end on each trial before completing her movements despite being instructed otherwise because this prevented us from analyzing her responses to the perturbations over time. We excluded an additional two subjects because they performed at chance levels on the unperturbed, cue-consistent trials, and two others were removed as outliers because their adjusted standard deviations (around 9 deg) on unperturbed, cue-consistent trials were six standard deviations above the mean standard deviation of the remaining eight subjects (mean \pm s.d. = 3.98 ± 0.81 deg). These remaining subjects all had standard deviations at or below 5.5 deg, which we used as a threshold for performance in the object placement condition and Experiment 2. For the object placement condition, we excluded five subjects from the data set for noticing the perturbations and dismissed three subjects early for poor performance on unperturbed, cue-consistent trials. Eight subjects per task were included in the data set.

We quantified the effects of the visual cues on task performance in the object placement task by evaluating the relationship between the 3D orientations suggested by the monocular and binocular cues and the 3D orientation of the cylinder. We used the positions of the infrared markers, which were rigidly attached to the side of the cylinder, to locate the major axis of the cylinder. Across trials, the only dimension in which the orientations of the target surfaces differed was in their slant about the horizontal axis, so this was the dimension we used for analyzing the 3D orientation of the cylinder. Since subjects were instructed to place the cylinder flush upon the target surface, the major axis of the cylinder represented their estimate of the surface normal. This vector was invariant to the rotation of the cylinder about its major axis, which depended on where subjects grasped it. We redefined the vector in viewer-centered coordinates, projected the vector into the sagittal plane to isolate the slant component, and computed the angle between this vector and the line of sight, which would be perpendicular to a frontoparallel surface, to determine the cylinder's slant. Similarly, we measured the effects of the visual cues in the grasping task using "grasp orientation," which characterized the orientation of the fingers relative to the orientation of the coin (Fig. 2). The problem was similar to finding the slant of the cylinder except that the fingers did not provide a rigid coordinate

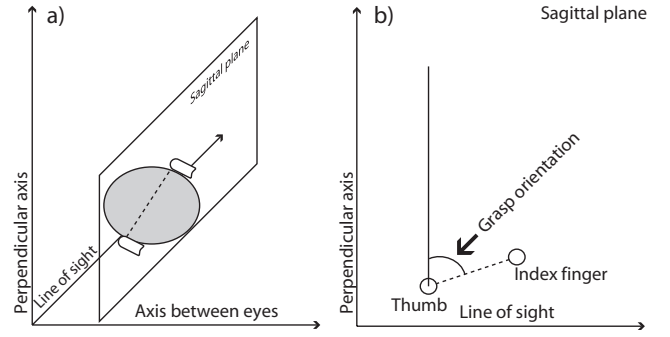


Fig. 2. Calculation of grasp orientation. (A) The thumbs grasping the coin and the vector between the contact points, which is projected into the sagittal plane. (B) A view of the sagittal plane from the subject's right and the grasp orientation angle specified by the projected vector.

system for the grasp. Instead, we identified fixed grasp points on the thumb and index finger relative to the rigidly attached infrared markers and computed the vector between these contact points from each Optotrak data frame. As we did for the major axis of the cylinder, we projected the grasp vector into the sagittal plane to isolate the slant component of the grasp. This projection ensured that our measure of slant for any stable grasp was invariant to rotation of the fingers about the circumference of the coin. Fig. 3 shows the mean grasp and cylinder orientation trajectories for one subject from each task in cue-consistent conditions. The figure shows that subjects used different orientation trajectories to grasp the coin or orient the cylinder when the surface was at different slants and that their performance was reasonably accurate. Also, the trajectories for different conditions diverged early in the movements, and the perturbation trials in the figure show that subjects clearly responded online to stimulus orientation changes.

To determine how the monocular and binocular cues influenced how subjects performed each task, we calculated the cue weights based on unperturbed trials (30–40 deg cue-consistent trials and all cue-conflict trials) according to the following model:

$$\sigma_{\text{effector}} = (k_1 + k_2 t)(w_{\text{mono}} \sigma_{\text{mono}} + w_{\text{bin}} \sigma_{\text{bin}}) + b_1 + b_2 t, \quad (1)$$

where σ_{effector} is the grasp orientation or the cylinder orientation at target contact, σ_{mono} and σ_{bin} are the slants suggested by the monocular and binocular cues throughout each trial, and w_{mono} and w_{bin} are the relative weights that quantified the respective contributions of the monocular and binocular cues. The cue weights were constrained to sum to 1; a weight of 0 would indicate that the particular cue did not affect the task, and a weight of 1 would show that it was entirely responsible for determining how subjects oriented their fingers or the cylinder. The terms k_1 and b_1 , respectively, accounted for multiplicative and additive biases. We also found that some subjects showed small biases that grew over the course of the experiment, so we included additive and multiplicative bias terms (k_2 and b_2) to account for effects that were correlated with trial number (t). The results are shown in Fig. 4. The error bars in this and subsequent figures indicate standard deviations for individual subject data and the standard error of the mean computed across subjects for group data. Binocular cues were 75% responsible for determining the endpoint orientation of the fingers in the grasping task, which was higher than the 64% contribution of binocular cues to the object placement task. This difference was significant according to an unpaired t -test, $t(14) = 2.72$, $P < 0.05$. Given our findings from Greenwald et al. (2005)

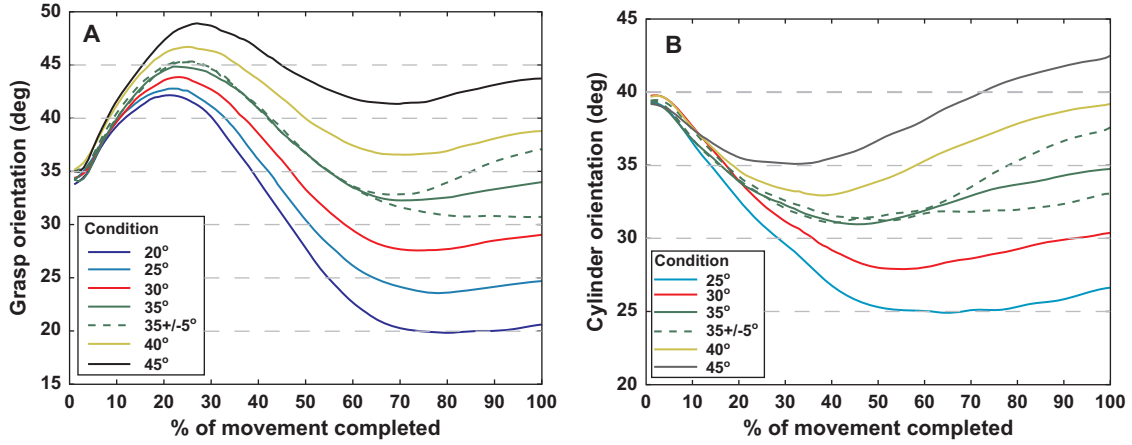


Fig. 3. (Color online) (A) Grasp orientation trajectories for one subject. (B) Cylinder orientation trajectories for one subject. Only trials from cue-consistent trials are shown, and the two cue-consistent perturbation conditions are included (dashed lines). Individual subjects differed in the ranges of effector orientations they used.

that the relative influence of binocular information depended on the amount of time allowed for monocular information to accumulate, one might ask whether the higher influence of binocular information in the grasping task was because subjects were initiating their movements sooner and therefore had more time to view the stimuli during the planning phase in the object placement task. There was a significant difference in movement initiation times, $t(14) = 2.59$, $P < 0.05$ but in the opposite direction; subjects began their movements an average of 61 ms later in the grasping task. However, since initially binocularly fusing the stimuli takes time, the higher binocular influence on the grasping task could have been due to subjects having had additional time for binocular fusion to occur. If this were true, one would expect a larger binocular influence on movements with longer movement initiation times. We divided the data from each subject into thirds based on movement initiation times and compared the cue influences computed from trials with the shortest and longest movement initiation times, but there was no significant difference in the relative binocular cue weights for object placement, $t(7) = 1.12$,

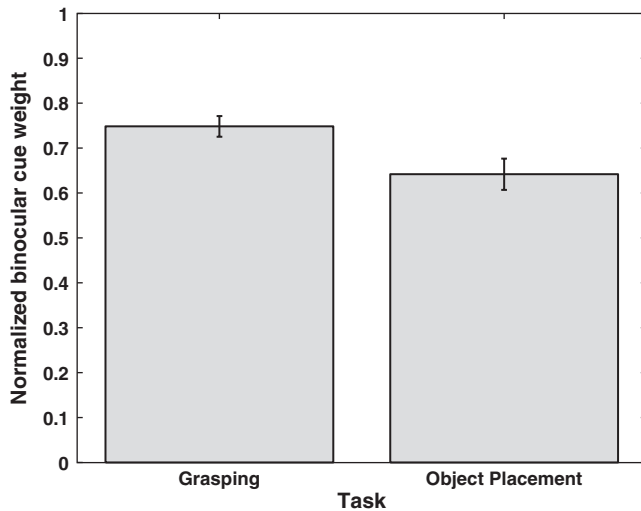


Fig. 4. Cue influences on the grasping and object placement tasks performed separately. These were computed from unperturbed trials at target contact.

$P = 0.30$, or grasping, $t(7) = 0.31$, $P = 0.76$. Experiment 2 provides further evidence that the 61-ms difference in movement initiation times could not have been responsible for the effect.

To separate the effects of task demands on cue integration from the effects due to task-related differences in cue processing speeds, we measured the temporal dynamics of cue processing based on subjects' responses to the stimulus perturbations over time by fitting an autoregressive function that predicted the orientation of the fingers from a linear combination of the effector orientations recorded in the previous 120 Hz data frames. Correlating the residual error of the predictions over time with the perturbations in the monocular and binocular cues during perturbed trials revealed how each cue separately affected the subjects' movements. Equation (2) provides the form of the model:

$$s_t = w_1(t) \times s_{t-1} + w_2(t) \times s_{t-2} + \dots + w_n(t) \times s_{t-n} + k_{\text{mono}}(t) \times \Delta_{\text{mono}} + k_{\text{bin}}(t) \times \Delta_{\text{bin}}, \quad (2)$$

where s_t is the orientation of the fingers or the cylinder at time t , $w_{t-1} \dots w_{t-n}$ are the predictive weights that describe the temporal correlations between the current and the previous effector orientations, and Δ_{mono} and Δ_{bin} are, respectively, the values of the monocular and binocular perturbations, which were always 0 or ± 5 deg. The values of k_{mono} and k_{bin} over time capture the residual variances in finger orientation that reflect the subjects' responses to the perturbations; these are the *perturbation influence functions*. Before subjects can respond to the perturbations, these functions should equal 0, and they change based on the extent to which the perturbations influence subject's movements. We fit the model parameters with a sliding window that used the Optotrak data frame at time t and the seven previous 120 Hz frames. The number of previous frames included in the regression only affected the scaling of the weights but was not crucial to the model's performance because it did not affect their relative magnitudes.

Fig. 5 shows the perturbation influence functions computed using the combined data from all subjects in each task beginning when the stimulus perturbations were introduced after the end of the visual mask (169 ms after movement initiation) and ending when more than half of the trials had ended. The perturbation influence functions provide a sensitive measure of when subjects first responded to the stimulus changes, and, for both grasping and

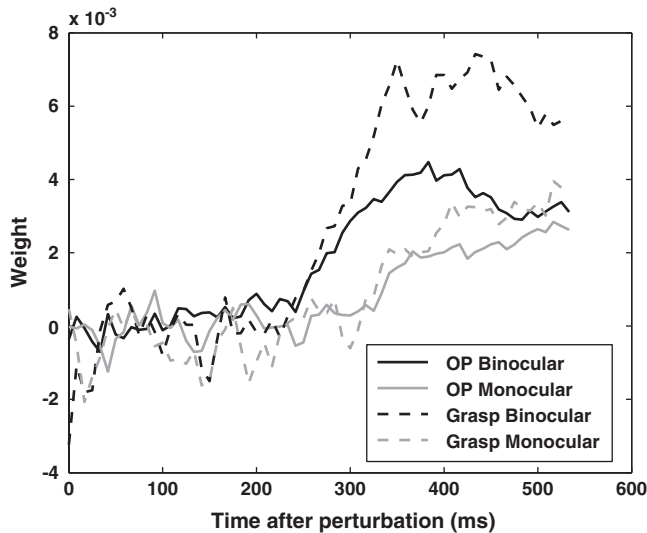


Fig. 5. Perturbation influence functions for both the grasping task and the object placement task (OP) fit by combining data from all subjects in each task. Perturbations occurred 169 ms after movement initiation.

object placement, this occurred approximately 250–300 ms after the perturbations appeared. The results from both tasks supported our prediction based on Greenwald et al. (2005) that binocular information would be processed faster than monocular information. The corresponding perturbation influence functions began increasing simultaneously, which indicated the similarity between the temporal dynamics of cue processing in each task. These functions reflect a combination of visual processing and motor output, so the accumulated influence of the information from each cue at a specific point in time, the overall influence of the cues, and the response latency of the motor system all could have affected the measured weights. An obvious feature in Fig. 5 is that the binocular influence function in the grasping task jumped to a higher level than the other influence functions; this most likely resulted from the greater overall influence of binocular information in this task but could also have resulted from subjects being able to reorient their fingers faster than the cylinder when responding to the perturbations. To account for this difference, we normalized the influence functions for each task. These relative binocular perturbation influence functions (Fig. 6) decreased over time, which indicates that subjects initially responded more to the binocular perturbations and that their influence relative to the monocular perturbations progressively declined. The rate of decline was similar for both grasping and object placement once subjects had an opportunity to begin adjusting to the perturbations.

Discussion

Experiment 1 revealed that task differences between prehension and object placement led to different cue integration strategies for the two tasks. Specifically, binocular information played a significantly greater role in grasping objects than in placing them. A possible explanation based on Smeets and Brenner (1999) is that the grasping task required subjects to accurately and independently control the positions of their thumb and index fingers so that they arrived at particular points on the edges of the slanted disc that covaried with slant, and subjects may have relied more

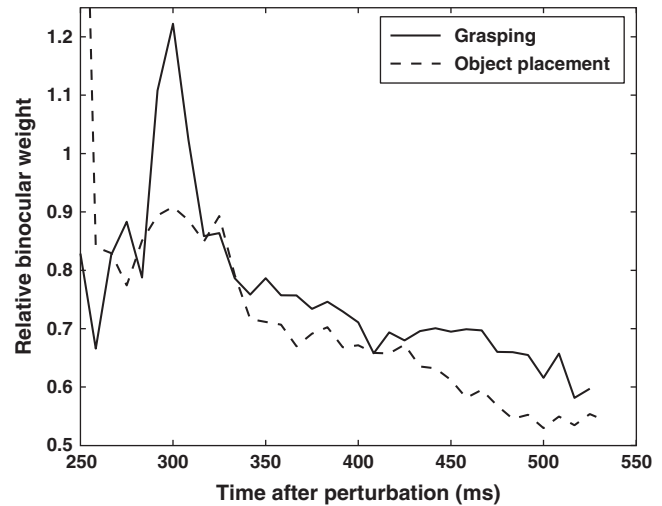


Fig. 6. Relative binocular influence function for each task over time computed by normalizing the perturbation influence functions.

on the binocular information because it provided better information about 3D position. In contrast, our object placement task only required subjects to accurately judge the orientation of the target surface since we allowed them to place the object anywhere on the target surface and the position of the center of the surface did not vary.

The temporal dynamics of cue integration were similar between tasks and matched our findings from Greenwald et al. (2005). Although task demands influenced how cues were integrated, they did not change how quickly the information provided by these cues was processed. Our data were consistent with the findings from Mamassian (1997), who showed that the 3D orientation of target objects significantly affected subjects' hand trajectories in a grasping task and that their responses to the orientations occurred early in the movements. Reaction times to the stimulus perturbations were 250–300 ms, and the responses to the perturbations indicated that binocular information accumulated faster than monocular information. The reaction times were considerably slower than the 100–150 ms reaction times observed for corrections to two-dimensional target displacements (Paulignan et al., 1991a,b; Prablanc & Martin, 1992). In Greenwald et al. (2005), we suggested that the higher response latencies were due to more complex processing required to estimate 3D orientation or due to the visual mask used to hide the stimulus perturbations. More recent evidence (D.C. Knill, unpublished) has shown that the visual mask adds a processing delay of 75–150 ms, which is consistent with evidence indicating that changing the background of a stimulus introduces processing latencies (Huang & Paradiso, 2005; Huang et al., 2005). Taking into account the delays introduced by the mask, the reaction times to the target slant perturbations appear to be similar to the 100–150 ms range found for perturbations in target position. Although the visual mask introduced a delay, it allowed subjects to maintain the target under central fixation throughout the entire trial while preventing awareness of the stimulus perturbations.

Experiment 2

In the first experiment, subjects performing the grasping task used binocular information more than subjects performing the object

placement task, and the results suggested that differences between the requirements for placing an object on a surface and those for picking up an object produced differences in how subjects integrated information from the available cues. The requirement to accurately control the positions of the fingers in depth for the grasping task could have produced these differences since stereopsis is more useful than monocular cues for estimating 3D position. Therefore, it appears that visual cue integration is task dependent and reflects how information is used for a particular task, which suggests that humans' cue integration strategies depend not only on the cues' relative reliabilities and how quickly they are processed but also on how they are used. Such a result could help generalize the findings from Knill (2005) in which subjects used different cue integration strategies for visual and visuomotor tasks; the differences may have arisen from task demands rather than from differences in cue processing for perception and action. To test this, Experiment 2 required subjects to perform the grasping task during some sessions and the object placement task during others and allowed us to perform a direct within-subjects comparison of both tasks. We predicted that the cue integration strategies observed for each task in isolation would be maintained. Alternatively, subjects might have combined their cue integration strategies into a single unified approach that they applied to both tasks, which would not have been unreasonable since picking up objects and placing them on surfaces are rarely performed in isolation outside the laboratory.

Method

Subjects

The seven subjects who completed this experiment met the same criteria as in Experiment 1. An additional subject was excluded for poor performance and another for exhibiting tremors induced by psychoactive medication that could have also affected processing speeds. No subject participated in multiple studies.

Apparatus

The experimental apparatus had different configurations for the grasping and object placement tasks that matched the descriptions provided for Experiment 1. The locations of the starting positions and oriented targets were similar in both tasks.

Calibration procedures

The robot and viewer calibration procedures remained the same as in the previous experiment. Finger calibrations were performed only for the grasping task; the cylinder used in the object placement task did not require a separate calibration for each subject.

Stimuli

The stimuli used in this experiment were the same Voronoi-textured coins used in the previous experiment.

Procedure

Subjects participated in six 1-h sessions with the two tasks interleaved across days (task order : ABBABA), and the order counterbalanced across subjects. Each day consisted of four 76-trial blocks for a total of 912 trials per task. We used the same conditions as in the object placement task from Experiment 1 and reduced the number of trials per condition from 64 to 48 so that subjects only completed six sessions rather than eight.

Additionally, we gave subjects feedback about their performance following each session as in the previous object placement task. Other than the specific requirements for the same subjects to pick up or place an object in different sessions, the procedures for both tasks in this experiment were identical to each other and to the associated tasks in Experiment 1.

Results

We computed the overall cue weights for this experiment in the same way as for the previous experiment using data from the unperturbed trials at target contact, and the results are shown in Fig. 7. The standard errors for the group means were corrected for intersubject differences (Loftus & Masson, 1994). A paired *t*-test showed that, as in the previous experiment, binocular cues influenced movements in the grasping task significantly more than those in the object placement task, $t(6) = 3.54$, $P < 0.05$.

We compared the unpaired data from Experiment 1 to the paired data from Experiment 2 using a mixed linear model to determine whether having subjects perform both tasks affected their cue integration strategies for each. The results showed only a significant effect of task on the overall influence of binocular cues, $F(1, 26) = 6.70$, $P < 0.05$. There was neither a significant difference in relative cue contributions between the experiments, $F(1, 26) = 1.69$, $P = 0.21$, nor a significant interaction between experiment and task, $F(1, 26) = 1.03$, $P = 0.32$. Earlier, we expressed a concern that differences in movement initiation times may have accounted for the differences in overall cue influences between the tasks. In Experiment 2, there was no significant difference in movement initiation times, $t(6) = 0.12$, $P = 0.91$, and the relative cue influences did not differ between Experiments 1 and 2. Therefore, we concluded that the observed effects were task related and not due to differences in movement initiation times.

We performed the same temporal analysis as for the previous experiment to determine how subjects processed incoming visual information when picking up and placing objects. The perturbation influence functions are shown in Fig. 8. As we found in Experiment 1, subjects responded to the stimulus perturbations with reaction times of approximately 250 ms, and binocular information was processed faster than monocular information.

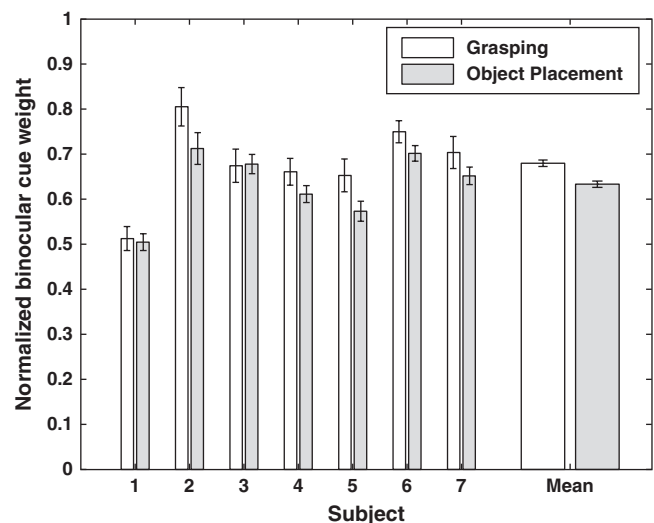


Fig. 7. Within-subjects binocular cue influences on the grasping and object placement tasks.

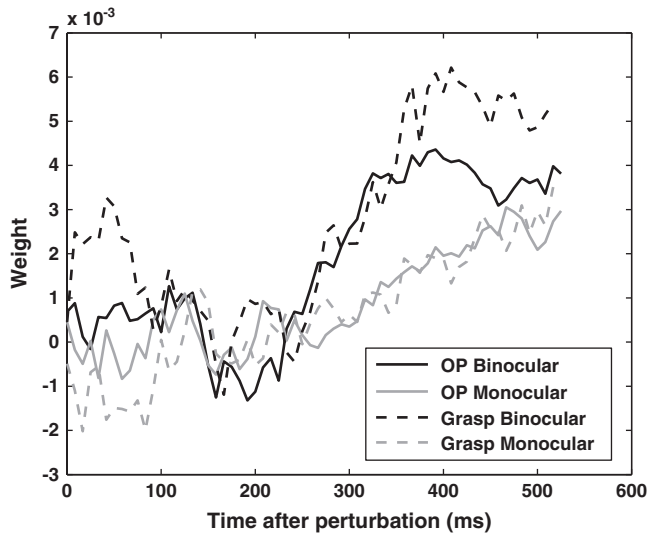


Fig. 8. Perturbation influence functions for the object placement (OP) and grasping tasks in Experiment 2.

The corresponding functions for the two tasks show a high degree of overlap beginning when subjects began responding to the perturbations, which indicates that the time course of online cue processing was similar between the two tasks. As in the previous experiment, the binocular influence function for the grasping task reached a higher value than the other influence functions.

Discussion

The results showed that subjects maintained separate cue integration strategies for the two tasks rather than merging their cue integration schemes into a unified strategy. When subjects in Experiment 2 performed both of the tasks that were performed separately in Experiment 1, binocular information influenced grasping more than object placement. This was consistent with the pattern of results from Experiment 1, and further analysis showed that the differences in treatments did not produce significant differences in task performance, which suggested that the cue integration strategies used for each of the tasks performed in isolation were maintained when subjects performed both. The perturbation influence functions we obtained in this experiment showed considerable overlap between corresponding conditions, which supported our finding that the temporal dynamics of cue processing generalized across tasks and were not affected by task demands.

General discussion

Our experiments showed that visual cue integration for motor control is determined in part by task demands in addition to relative reliabilities and processing speeds. The main visual targets in our object placement and prehension tasks were identical across tasks, but we found that subjects relied more on binocular information when grasping objects than when placing objects. We hypothesized that this was because grasping required subjects to accurately position their fingers in depth since the positions of the grasp points along the edges of the coin covaried with its 3D orientation. For object placement, the center of the target

determined the appropriate position for the object, and this point remained constant across different slants. Estimating the slant of the target was equally important in both tasks, but 3D position was more relevant for the grasping task; this led to the differences in cue integration strategies between the two tasks. When we had subjects perform both tasks in Experiment 2, the differences in cue integration strategies that we observed between the tasks in Experiment 1 were maintained.

How subjects came to use the cues differently for grasping and object placement remains an open question. One possibility is that the visual system uses the available information differently for each task. This could mean dynamically reweighting the cues according to their relative reliabilities for performing a given task or processing them in separate ways using different computational approaches. Alternatively, it could be estimating different variables; for instance, subjects could have depended on binocular disparity gradients across the surfaces for object placement but on the binocular disparities of the grasp points during grasping as Smeets and Brenner (1999) suggested. The trajectories of the cylinder and subjects' fingers do not permit us to identify the underlying cause of the effects we found, which is often a problem for psychophysical studies, and we hope that future experiments will provide an explanation.

In contrast to the task-dependent effects we found on cue integration, our investigation into the temporal dynamics of cue processing revealed that task differences do not lead to variations in how quickly the visual system extracts information from cues. The temporal analyses based on the responses to the stimulus perturbations as a function of time showed that the rates at which visual cues were processed were the same across tasks in both Experiments 1 and 2. Regardless of which task they performed, subjects in all parts of our study responded to the binocular perturbations with approximately 250-ms latencies, and binocular information accumulated faster than monocular information. This result replicates the major finding from Greenwald et al. (2005).

Our experiments have consistently shown that binocular information is processed faster than monocular information. This is not inconsistent with the common wisdom that stereopsis is slow (McKee et al., 1990); it takes time for binocular fusion to occur, but once the eyes are accommodated and verged to the correct depth and binocular fusion has occurred, the binocular information can be processed rapidly. Further evidence that there are differences in the rate at which monocular and binocular information accrue comes from the relative influence functions from Experiment 1 shown in Fig. 6. These functions indicate that binocular information initially dominated subjects' responses to the stimulus perturbations but that the relative binocular influence declined asymptotically over time. If the differences in cue processing were only due to different levels of uncertainty about the binocular and monocular information, these relative influence functions would be expected to remain constant.

The perturbation influence functions provide information about the temporal dynamics of subjects' responses to the perturbations; however, they are insufficient to show how the perceptual weights for the different cues changed over time because they do not directly measure these weights but rather indicate perceptual estimates that have been filtered through the dynamics of the motor control system and hand. In Greenwald et al. (2005), we confirmed that relative processing speeds determine how visual cues influence movements over time by simulating an optimal estimator using a Kalman filter that integrated noisy slant estimates from monocular and binocular cues and provided its

estimate to a minimum jerk controller of hand orientation. When the only differences in the sensory inputs to the filter were in the cue uncertainties, the predicted relative influence of binocular cues based on the perturbation influence functions derived from the simulated hand kinematics remained constant over time. To account for the decrease in the relative influence of binocular information on hand orientation over time, it was necessary to lowpass filter the slant estimates derived from the monocular cues, which was equivalent to modeling a system that accrues monocular information about 3D orientation more slowly than binocular information.

Our finding that binocular information accumulates more quickly suggests that 3D displays should be particularly useful for applications in which operators need to process incoming information about 3D spatial relationships as rapidly as possible. Two recent studies (Falk et al., 2001; Blavier et al., 2006) comparing performance for various tasks using a da Vinci robotic surgical system under monocular and binocular viewing conditions found that subjects performed equally well in both conditions but required significantly less time to perform the tasks when using binocular vision. Both studies were qualitative and simply compared performance under monocular and binocular viewing conditions. Since one would expect speeded performance when the reliability of the incoming information increases (Roitman & Shadlen, 2002), this result may simply reflect the added information provided in the binocular viewing conditions. Our results about the temporal dynamics of cue processing indicate that processing speed may also contribute to improved temporal performance under binocular viewing. They therefore could have important implications for displays used in certain time-critical applications that rely on 3D spatial representations. The information processing advantages of stereopsis could make stereoscopic displays worthwhile for laparoscopic surgery, air traffic control, or heads-up displays used in aircraft or automobiles.

Acknowledgments

The authors thank Leslie Richardson for assistance with data collection, Brian McCann for technical assistance, Teresa Williams for helping with the design and construction of the conductive finger thimbles, and Schlegel Electronic Materials for providing a large sample of conductive fabric. This research was supported by the National Institutes of Health grant EY013389.

References

- ALAIS, D. & BURR, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology* **14**, 257–262.
- BLAVIER, A., GAUDISSERT, Q., CADIÈRE, G.-B. & NYSEN, A.-S. (2006). Impact of 2D and 3D vision on performance of novice subjects using da Vinci robotic system. *Acta Chirurgica Belgica* **106**, 662–664.
- BRADSHAW, M.F., ELLIOTT, K.M., WATT, S.J., HIBBARD, P.B., DAVIES, I.R.L. & SIMPSON, P.J. (2004). Binocular cues and the control of prehension. *Spatial Vision* **17**, 95–110.
- DENEVE, S., LATHAM, P.E. & POUGET, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience* **4**, 826–831.
- ERNST, M.O. & BANKS, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433.
- ERNST, M.O., BANKS, M.S. & BÜLTHOFF, H.H. (2000). Touch can change visual slant perception. *Nature Neuroscience* **3**, 69–73.
- FALK, V., MINTZ, D., GRÜNENFELDER, J., FANN, J.I. & BURDON, T.A. (2001). Influence of three-dimensional vision on surgical telemanipulator performance. *Surgical Endoscopy* **15**, 1282–1288.
- GHAHRAMANI, Z., WOLPERT, D.M. & JORDAN, M.I. (1997). Computational models of sensorimotor integration. In *Self-Organization, Computational Maps and Motor Control*, ed. MORASSO, P.G. & SANGUINETI, V., pp. 117–147. Amsterdam, Elsevier Press.
- GLOVER, S. (2004). Separate visual representations in the planning and control of action. *Behavioral and Brain Sciences* **27**, 3–78.
- GOODALE, M.A., JAKOBSON, L.S. & KELLOR, J.M. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia* **32**, 1159–1178.
- 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.
- HILLIS, J.M., WATT, S.J., LANDY, M.S. & BANKS, M.S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision* **4**, 967–992.
- HUANG, X., BLAU, S. & PARADISO, M.A. (2005). Background changes delay the perceptual availability of form information. *Journal of Neurophysiology* **94**, 4331–4343.
- HUANG, X. & PARADISO, M.A. (2005). Background changes delay information represented in macaque V1 neurons. *Journal of Neurophysiology* **94**, 4314–4330.
- 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* **42**, 113–142.
- JACOBS, R.A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research* **39**, 3621–3629.
- KNILL, D.C. (2005). Reaching for visual cues to depth: The brain combines cues differently for motor control and perception. *Journal of Vision* **5**, 103–115.
- KNILL, D.C. & SAUNDERS, J.A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research* **43**, 2539–2558.
- LANDY, M.S., MALONEY, L.T., JOHNSTON, E.B. & YOUNG, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research* **35**, 389–412.
- LOFTUS, G.R. & MASSON, M.E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review* **1**, 476–490.
- MAMASSIAN, P. (1997). Prehension of objects oriented in three-dimensional space. *Experimental Brain Research* **114**, 235–245.
- MAROTTA, J.J., BEHRMANN, M. & GOODALE, M.A. (1997). The removal of binocular cues disrupts the calibration of grasping in patients with visual form agnosia. *Experimental Brain Research* **116**, 113–121.
- McKEE, S.P., LEVI, D.M. & BOWNE, S.F. (1990). The imprecision of stereopsis. *Vision Research* **30**, 1763–1779.
- MELMOTH, D.R. & GRANT, S. (2006). Advantages of binocular vision for the control of reaching and grasping. *Experimental Brain Research* **171**, 317–388.
- PAULIGNAN, Y., JEANNEROD, M., MACKENZIE, C. & MARTENIUK, R. (1991a). Selective perturbation of visual input during prehension movements. 2. The effects of changing object size. *Experimental Brain Research* **87**, 407–420.
- PAULIGNAN, Y., MACKENZIE, C., MARTENIUK, R. & JEANNEROD, M. (1991b). Selective perturbation of visual input during prehension movements. 1. The effects of changing object position. *Experimental Brain Research* **83**, 502–512.
- PRABLANC, C. & MARTIN, O. (1992). Automatic control during hand reaching at undetected 2-dimensional target displacements. *Journal of Neurophysiology* **57**, 455–469.
- ROITMAN, J.D. & SHADLEN, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience* **22**, 9475–9489.
- SCHRATER, P.R. & KERSTEN, D. (2000). How optimal depth cue integration depends on the task. *International Journal of Computer Vision* **40**, 73–91.
- SERVOS, P., GOODALE, M.A. & JAKOBSON, L.S. (1992). The role of binocular vision in prehension: A kinematic analysis. *Vision Research* **32**, 1513–1521.
- SMEETS, J.B. & BRENNER, E. (1999). A new view on grasping. *Motor Control* **3**, 237–271.
- STEVENS, K.A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics* **46**, 183–195.
- WATT, S.J. & BRADSHAW, M.F. (2000). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia* **38**, 1473–1481.
- WITKIN, A.P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence* **17**, 17–45.