PERGAMON

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 43 (2003) 2603–2613

Vision Research

www.elsevier.com/locate/visres

Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts

Joseph E. Atkins, Robert A. Jacobs *, David C. Knill

Department of Brain and Cognitive Sciences, The Center for Visual Science, University of Rochester, Rochester, NY 14627-0268, USA

Received 6 March 2003; received in revised form 12 June 2003

Abstract

We studied the hypothesis that observers can recalibrate their visual percepts when visual and haptic (touch) cues are discordant and the haptic information is judged to be reliable. Using a novel visuo-haptic virtual reality environment, we conducted a set of experiments in which subjects interacted with scenes consisting of two fronto-parallel surfaces. Subjects judged the distance between the two surfaces based on two perceptual cues: a visual stereo cue obtained when viewing the scene binocularly and a haptic cue obtained when subjects grasped the two surfaces between their thumb and index fingers. Visual and haptic cues regarding the scene were manipulated independently so that they could either be consistent or inconsistent. Experiment 1 explored the effect of visuo-haptic inconsistencies on depth-from-stereo estimates. Our findings suggest that when stereo and haptic cues are inconsistent, subjects recalibrate their interpretations of the visual stereo cue so that depth-from-stereo percepts are in greater agreement with depth-from-haptic percepts. In Experiment 2 the visuo-haptic discrepancy took a different form when the two surfaces were near the subject than when they were far from the subject. The results indicate that subjects recalibrated their interpretations of the stereo cue in a context-sensitive manner that depended on viewing distance, thereby making them more consistent with depth-from-haptic estimates at all viewing distances. Together these findings suggest that observers' visual and haptic percepts are tightly coupled in the sense that haptic percepts provide a standard to which visual percepts can be recalibrated when the visual percepts are deemed to be erroneous.

1. Introduction

Perceptual environments are often redundant in the sense that they provide observers with cues from multiple sensory modalities, such as visual, auditory, and haptic (touch) cues. Environments that provide cues in only a single modality can also be redundant. For example, we know of nearly a dozen visual cues to depth, including cues arising from object rotation, observer motion, binocular vision, texture and shading gradients in retinal images, and many other factors (Cutting & Vishton, 1995). One useful role for a redundant cue is to maintain calibration of another cue. This article focuses on how observers can use a reliable cue from one modality (haptics) to recalibrate depth interpretations of a cue (visual stereo disparities) from another modality.

Several theorists have speculated that observers can use information derived from consistencies and inconsistencies among cues for the purposes of visual learning. Berkeley (1709), for instance, hypothesized nearly three hundred years ago that infants acquire aspects of visual perception by correlating visual sensations with sensations arising from motor movements (a famous quote from Berkeley's book is "touch educates vision"). More recently, Wallach (1985) hypothesized that in every perceptual domain, such as depth or shape perception, there is one primary source of information, usable innately and not modifiable by experience. Other cues are acquired later, through correlation with the innate process.

A second way that cue redundancy can aid visual learning is through the experience-dependent adaptation of visual cue combination rules. Subjects in a study by Atkins, Fiser, and Jacobs (2001) viewed cylinders defined by texture and motion cues, and also grasped the cylinders. The cylinder depth indicated by one of the visual cues (e.g., texture) was consistent with the depth

* Corresponding author. Tel.: +1-716-275-0753; fax: +1-716-442-9216.

E-mail address: robbie@bcs.rochester.edu (R.A. Jacobs).

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0042-6989(03)00470-X

indicated by the haptic cue, whereas the depth indicated by the remaining visual cue (motion) was inconsistent. After prolonged exposure, subjects adapted their visual cue combination rules so as to place greater weight on their depth estimates derived from the visual cue that was consistent with the haptic cue (depth-from-texture estimates) and less weight on the estimates derived from the other visual cue (depth-from-motion estimates). Related results were reported by Ernst, Banks, and Bülthoff (2000).

A third way in which cue redundancy can aid visual learning is through cue recalibration (Epstein, 1975). Of greatest relevance for our purposes are studies showing that observers can recalibrate their interpretations of a visual stereo cue. Subjects in a study by Wallach, Moore, and Davidson (1963) viewed a rotating three-dimensional wire figure through a telestereoscopic mirror arrangement. Initially, depth percepts based solely on the kinetic depth cue were veridical, but depth relations based solely on the stereo cue were overestimated due to the fact that the telestereoscopic arrangement doubled the effective interocular distance and disparity associated with the figure. After prolonged exposure to environments containing both cues, however, subjects recalibrated their depth-from-stereo estimates so that they were in greater agreement with their depth-from-motion estimates. Subjects in a study by Epstein and Morgan (1970) wore a horizontal magnifier over one eye while they walked through the hallways and staircases of a building. Although the magnifier caused subjects to initially misperceive depth, subjects adapted their depth judgments over time in a way which the authors interpreted as a recalibration of subjects' depth-from-stereo estimates so that they were more veridical. Adams, Banks, and van Ee (2001) performed a study similar to that of Epstein and Morgan (1970), and showed that the change was indeed due to a modification of the mapping between retinal disparity and perceived depth and not to adaptation of subjects' visual cue combination rules or to monocular adaptation. The recalibration of depth-from-stereo judgments reported in the articles cited above may have been due, at least in part, to a recalibration of viewing distance estimates based on vergence angle or accommodation. Maddox (1893) showed that the resting level of vergence can be adapted by placing a prism in front of one eye so that additional vergence is necessary for fusion. Judge and Miles (1985) showed that the coupling between vergence eye movements and accommodation can be adapted due to viewing the world through periscopic glasses which increase the effective interocular separation.

This article examines observers' abilities to use haptic cues as a standard against which they can estimate errors in their depth-from-stereo estimates, and to recalibrate these estimates if necessary. Experiment 1 studied whether observers can adapt their depth-from-stereo estimates so that these estimates are in greater agreement with their depth-from-haptic estimates when placed in an environment in which stereo and haptic cues to depth are inconsistent. Subjects observed a scene consisting of two fronto-parallel surfaces, one relatively narrow and the other relatively wide. The surfaces were placed so that the narrow surface was closer to the subject, and so that it occluded the middle portion of the wide surface (as shown in Fig. 2). In a pre-test and a post-test, subjects viewed the scene and judged whether the distance in depth between the two surfaces was less than or greater than the width of the near surface. Importantly, the visual environment contained only one cue to the distance between the two surfaces, namely a stereo cue. Between pre-test and post-test, subjects both viewed the scene and grasped the two surfaces using their thumb and index fingers, thereby obtaining a haptic cue to the distance in depth between the two surfaces. The stereo and haptic cues were inconsistent in the sense that they indicated different distances between the two surfaces. As explained below, this inconsistency was a consequence of the fact that the viewing distance to the surfaces suggested by vergence angle differed from the reaching distance to the surfaces suggested by proprioception when grasping the surfaces. Comparisons of each subject's performances on the pre-test and post-test indicate that prolonged exposure to the discrepancy between visual and haptic cues caused subjects to recalibrate their interpretations of the stereo cue so that their depth-from-stereo estimates were in greater agreement with their depth-from-haptic estimates.

While consistent with the hypothesis that subjects recalibrated their depth-from-disparity judgments, the results of Experiment 1 are also compatible with a simpler model based on cue weighting. Suppose that observers make depth-from-stereo estimates by forming a linear weighted average of the depth indicated by the pattern of binocular disparities and a depth assumed by an observer on the basis of prior experience (sometimes referred to as a default depth value; a Bayesian statistician would define this as the most probable depth value based on an a priori distribution indicating the observer's beliefs about likely depth values before viewing the scene). The recalibration shown by subjects in Experiment 1 could be accounted for by an adaptation of the weight values used by subjects in their linear weighted averages. Experiment 2 was designed to evaluate this possibility.

Experiment 2 tested whether subjects could adapt to a situation in which the visuo-haptic discrepancy took one form when the two surfaces were near the subject and a different form when the surfaces were far from the subject. In particular, when the two surfaces were near the subject, the visual distance in depth between the two surfaces was less than the haptic distance, whereas the visual distance was greater than the haptic distance when the surfaces were far from the subject. Adaptation of stereo-depth to both of these conflicts cannot be accounted for by the simple linear weighting model mentioned above, but can result from appropriate recalibration of the depth-from-stereo mechanism (e.g., by biasing the estimate of viewing distance used to scale depth-from-disparity estimates to a point midway between the two distances used in the experimental stimuli). The results indicate that subjects recalibrated their interpretations of the stereo cue so that depth-from-stereo estimates were larger at the near viewing distance after prolonged exposure to the visuo-haptic discrepancy and smaller at the far viewing distance. Overall, our findings suggest that haptic percepts provide a standard to which visual percepts can be recalibrated when they are deemed to be erroneous.

2. Experiment 1

Experiment 1 studied whether observers can adapt their depth-from-stereo estimates so that these estimates are in greater agreement with their depth-from-haptic estimates when placed in an environment in which visual and haptic cues to depth are inconsistent.

2.1. Methods

2.1.1. Experimental apparatus

The experiment was carried out using a visuo-haptic virtual reality environment. The experimental apparatus consisted of virtual reality goggles (model V8 head-mounted display from Virtual Research) and two PHANToM™ 3D Touch interfaces (SensAble Technologies) that were attached by two fingerholders to the subject's thumb and index fingers (see Fig. 1). This apparatus allowed subjects to physically interact with virtual objects viewed via the goggles in a natural way using a wide range of movements (e.g., grasping, moving, or throwing objects). The 3D Touch interfaces generated force fields that created haptic sensations (e.g., weight, hardness, friction) appropriate to the motor interactions with the objects displayed in the goggles. The experimental apparatus allowed us to independently manipulate the visual and haptic cues regarding these objects.

Subjects placed their chins on a chinrest which was attached to a height-adjustable table. The position and orientation of a subject's head in the virtual reality environment was monitored using a sensor (Polhemus Fastrak) attached to the top of the head-mounted display. Paddles on the side of the chinrest constrained a subject's head position and orientation. The chinrest was positioned at one end of the workspace such that a typical subject's interocular midpoint was 110 mm above the virtual workspace floor and approximately 450 mm from its center.

2.1.2. Stimulus

The scene depicted in the experiment consisted of a wooden floor, a wooden wall at the back of the workspace, and two fronto-parallel surfaces (see Fig. 2). The rear surface was rendered in a dark blue color, had fixed dimensions, and was always larger than the front surface. The front surface was rendered in a green color, and had varying height and width dimensions and varying vertical position. The surfaces were placed directly in front of a subject such that the front surface occluded a middle portion of the rear surface, and such that the subject always viewed the scene head-on (roughly an orthogonal view; see the left image in Fig. 2). Scenes differed in the position of the surfaces relative to the subject; that is, the position in depth of the point midway between the two surfaces (referred to either as the viewing distance or the reaching distance) could vary from a near position to a far position relative to the subject. Scenes also differed in the distance in depth between the two surfaces, and in the height, width, and vertical position of the front surface. The front surface's height and vertical position were varied randomly so

Fig. 1. A subject is shown in the experimental apparatus. Visual stimuli were delivered via a head-mounted display, and haptic stimuli were delivered via the force feedback devices attached to the subject's thumb and index fingers. The subject is grasping two (virtual) fronto-parallel surfaces, with her thumb against the front surface and her index finger wrapped around behind the back surface.

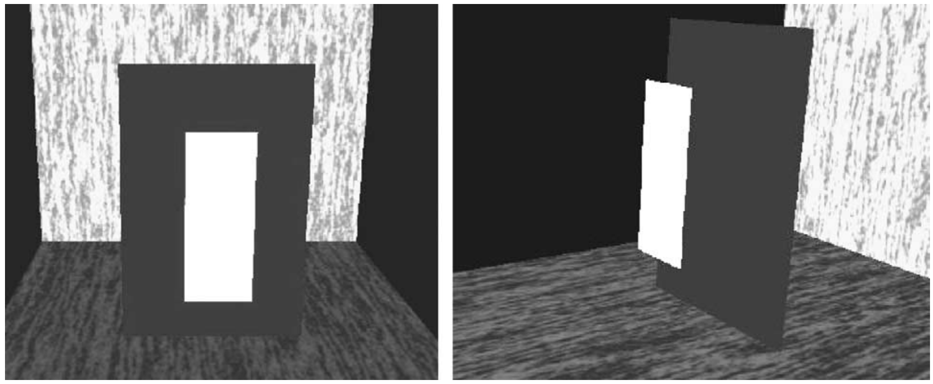

Fig. 2. (Left) One image of a stereo pair depicting the two fronto-parallel surfaces. Subjects judged whether the distance in depth between the two surfaces was greater than or less than the width of the nearer surface. (Right) The scene as it would appear from the side. This view was never shown to subjects, but is presented here to allow the reader to see the distance in depth between the two surfaces. Subjects were only able to see the scene head-on (roughly an orthogonal view) due to side paddles on the chinrest which constrained the subject's head position.

that the relative heights and sizes of the two surfaces could not be used as reliable cues to the distance between the surfaces. A stereo cue was the only reliable visual cue to the relative depth of the surfaces. A haptic cue regarding a scene was obtained when a subject grasped the two surfaces (a subject placed his thumb against the front surface and wrapped his index finger behind the rear surface; see Fig. 1). Subjects' hands were not visible during a grasp.

In some circumstances, the visual and haptic cues were inconsistent in the sense that they indicated different distances between the two surfaces. This inconsistency arose from the fact that the viewing distance to the surfaces suggested by vergence angle differed from the reaching distance to the surfaces suggested by proprioception when grasping the surfaces. Based on interviews conducted at the end of each subject's participation in the experiment, subjects were unaware of the visuo-haptic discrepancy. The discrepancy was created as follows. Let D_V denote the viewing distance—the distance along the depth axis from a point midway between the left and right eyes to the fixation point, where we assume that subjects fixated a point midway in depth between the front and rear surfaces. Let D_H denote the reaching distance—the distance along the depth axis from a point midway between the left and right eyes to the point midway in depth between a subject's thumb and index fingers when the subject grasped the two surfaces. During some stages of the experiment, we set the viewing and reaching distances to differ by 60 mm. When doing so, we wanted the binocular disparity computed using points on the front and rear surfaces, denoted η, to be consistent with both the placement of two surfaces at a viewing distance D_V and with the placement of two surfaces at a reaching distance D_H. To achieve this, the distance between the two surfaces as indicated by the haptic cue was scaled relative to the distance indicated by the visual cue. Let Δ_V denote the distance between the two surfaces as indicated by the disparity η (in radians) when the surfaces were centered at the viewing distance D_V. Let Δ_H denote the distance between the two surfaces as indicated by the same disparity η when the surfaces were centered at the reaching distance D_H (this is also the distance between the two surfaces as indicated by the haptic cue). Then the value of Δ_H was selected using the equation

$$\eta = (I * \Delta_V)/(D_V)^2 = (I * \Delta_H)/(D_H)^2, \quad (1)$$

where I is the interocular distance (see Hershenson, 1999, for discussion regarding this equation). When the reaching distance D_H was greater than the viewing distance D_V, then Δ_H was larger than Δ_V. When D_H was less than D_V, then Δ_H was smaller than Δ_V.

The experiment used multiple viewing and reaching distances. If the visual and haptic cues were consistent, then the near and far viewing and reaching distances were 435 and 525 mm, respectively. On trials when the cues were inconsistent, the near and far viewing distances were 375 and 465 mm, whereas the near and far reaching distances were 435 and 525 mm. The near and far viewing distances on test trials when only the visual cues were available were 425 and 525 mm.

The distance between the two surfaces as indicated by the stereo cue took one of seven possible values, evenly spaced from 32 to 68 mm. The width of the front surface was varied randomly. On training trials when both visual and haptic cues were available, the width was either 38, 50, or 62 mm. On test trials, when only the visual cues were available, it was either 44 or 56 mm. The height of the front surface varied randomly between 90 and 110 mm on each trial, and its vertical position (the vertical distance from the workspace floor to the center of the surface) varied randomly between 88.5 and 118.5 mm on each trial. For the rear surface, its width was 120 mm, its height was 180 mm, and its vertical position was 90.5 mm. As mentioned above, the properties of the front surface were varied so that the relative height and sizes of the surfaces were not reliable cues to the distance between the surfaces.

2.1.3. Procedure

Experimental trials can be classified as either training or test trials. Subjects received both visual and haptic cues regarding a scene on a training trial; they received only visual cues on a test trial. A large blue cube covered the experimental workspace between training trials and between test trials so that subjects could not see or touch the items in a scene as the scene was being visually and haptically rendered.

On a training trial, subjects first viewed a scene for as long as needed and made a judgment comparing the width of the front surface to the distance in depth between the front and rear surfaces. If a subject perceived the width as greater, then the subject verbally responded "wider". A subject responded "deeper" if the subject perceived the distance in depth to be greater. Otherwise the subject responded "same". Next a subject grasped the front and rear surfaces between his thumb and index finger for as long as needed by pressing his thumb against the front of the front surface and his index finger against the rear of the rear surface. The subject once again made a judgment comparing the width of the front surface to the distance in depth between the front and rear surfaces. Subjects were instructed that they should use all of the information, both visual and haptic, when making this second judgment. Subjects were not provided with feedback regarding the correctness or incorrectness of their first or second judgments. Trials using different widths of the front surface, different distances in depth between the two surfaces, and different viewing and reaching distances were randomly intermixed and counterbalanced.

On a test trial, a subject viewed a scene for 2.25 s and made a judgment comparing the width of the front surface to the distance in depth between the front and rear surfaces. In this case, subjects could respond either "wider" or "deeper" but were not permitted to respond "same". As before, subjects were not provided with feedback regarding the correctness of their responses. Trials using different widths of the front surface, different distances in depth between the two surfaces, and different viewing distances were randomly intermixed and counterbalanced. The possible values for the width of the front surface were different on training and test trials so that subjects' responses on test trials could not be based on memories of specific stimulus situations observed on training trials.

Subjects participated in experimental sessions on four days. On Day 1, subjects performed 210 consistent-cue training trials meaning that visual and haptic cues regarding the scenes were consistent. Subjects also performed 56 test trials. The session on Day 1 is regarded as a practice session during which subjects were in the process of learning about the experimental environment and task. Consequently, subjects' responses on test trials during this session were not analyzed. On Day 2, subjects performed 126 consistent-cue training trials followed by 224 test trials. This set of test trials is referred to as the pre-test. Subjects performed 210 inconsistent-cue training trials and 56 test trials on Day 3. Because subjects were in the process of learning about the inconsistent-cue environment during this session, their responses on the test trials were not analyzed. Subjects performed 126 inconsistent-cue training trials followed by 224 test trials on Day 4. This set of test trials is referred to as the post-test. [1]

Our prediction is that subjects' judgments on the post-test trials will differ from their judgments on the pre-test trials due to their experiences during the inconsistent-cue training trials. Because the haptic cue consistently indicated a larger relative distance between the front and rear surfaces than the stereo cue on these trials, we expect that subjects will recalibrate their depth-from-stereo estimates by making them larger. If so, subjects would be adapting their depth-from-stereo estimates so that these estimates are in greater agreement with their depth-from-haptic estimates.

2.1.4. Subjects

Subjects were five undergraduate and graduate students at the University of Rochester. They had normal or corrected-to-normal vision and were naive to the purposes of the study.

2.2. Results

The graph on the left of Fig. 3 shows a typical subject's data (subject TL) on the pre-test trials at the near viewing distance. The horizontal axis gives the difference between the relative depth between the surfaces and the width of the front surface (measured in millimeters); the vertical axis gives the probability that the subject judged the relative depth as greater than the width. A logistic psychometric function was fit to the nine data points, using a maximum likelihood procedure. The estimated value of the threshold parameter was the difference between the relative depth and width at which the logistic function yields a probability of 0.5, the point of

[1] Unlike other subjects, subject JJ received inconsistent-cue training on Days 1 and 2 of the experiment and consistent-cue training on Days 3 and 4. In principle, the order of these two types of training should not matter so long as this order is taken into account when comparing the subject's responses on pre-test trials versus post-test trials. A comparison of this subject's data with those of other subjects confirms this hypothesis. We have therefore included this subject's data in our analyses.

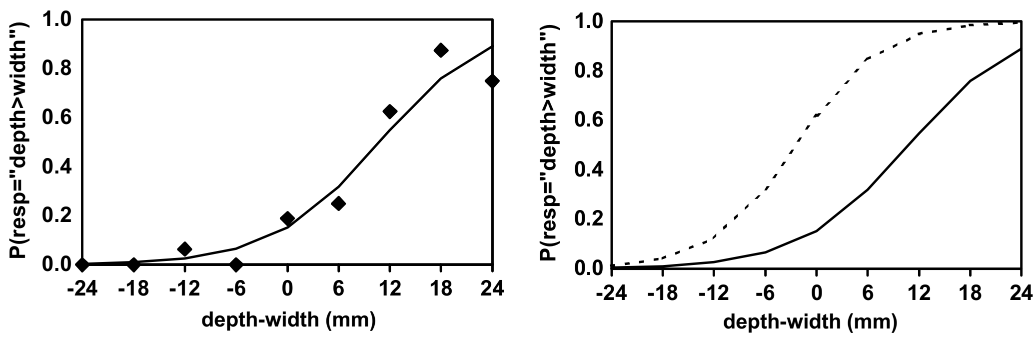

Fig. 3. (Left) Subject TLs pre-test data (solid diamonds) and a logistic function which has been fit to these data points (solid line) at the near viewing distance. (Right) Subject TLs pre-test (solid line) and post-test (dotted line) logistic fits at the near viewing distance. The leftward shift from pre-test to post-test denotes an overall increase in depth-from-stereo estimates.

subjective equality (PSE), between the perceived relative depth and width of the surface.

The graph on the right of Fig. 3 shows the logistic fits to this subject's data at the near viewing distance on the pre-test (solid line) and post-test (dotted line) trials. Clearly, this subject's judgments were different on the pre-test versus the post-test; the post-test data is generally shifted to the left of the pre-test data. The subject's PSE on the pre-test trials was 10.8, and his PSE on the post-test trials was −2.4. We hypothesize that this shift is due to the fact that the subject adapted his depth-from-stereo judgments, based on his experiences during the inconsistent-cue training trials, so that these judgments were generally larger, thereby making them more consistent with his depth-from-haptic judgments. The data for all five subjects is shown in the graphs on the left (near viewing distance) and right (far viewing distance) of Fig. 4. The horizontal axis indicates the subject; the vertical axis indicates a subject's PSE on the pre-test (black bars) and post-test (gray bars) trials. All subjects had larger PSEs on the pre-test than on the post-test.

Fig. 5 shows the average shift in subjects' PSEs (post-test PSE minus pre-test PSE) for the near and far viewing distances (the error bars give the standard errors of the means). The bars toward the left labeled 'exp-near' and 'exp-far' refer to the subjects that we have been discussing so far. At the near viewing distance, the average shift was −17.9 mm (two-tailed t-test; $T(4) = 4.21$; $p = 0.014$). At the far viewing distance, the average shift was −14.3 mm ($T(4) = 3.96$; $p = 0.017$).

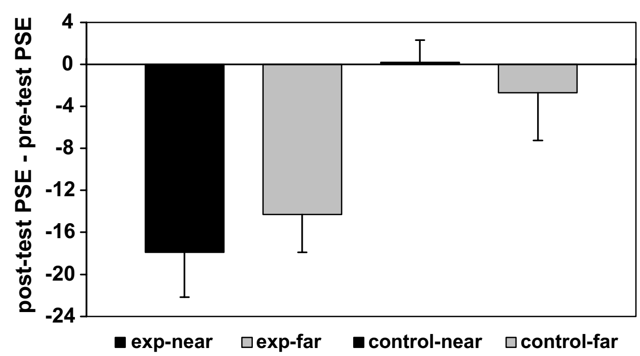

Fig. 5. Post-test PSEs minus pre-test PSEs for the experimental (left two bars) and control (right two bars) subjects at near (black bars) and far (gray bars) viewing distances.

We ran three subjects in a control condition to insure that the change in relative depth judgments was not simply a result of practice effects or perceptual drift over time. Control subjects received the same training and testing as experimental subjects with the exception that during training, they never experienced a discrepancy

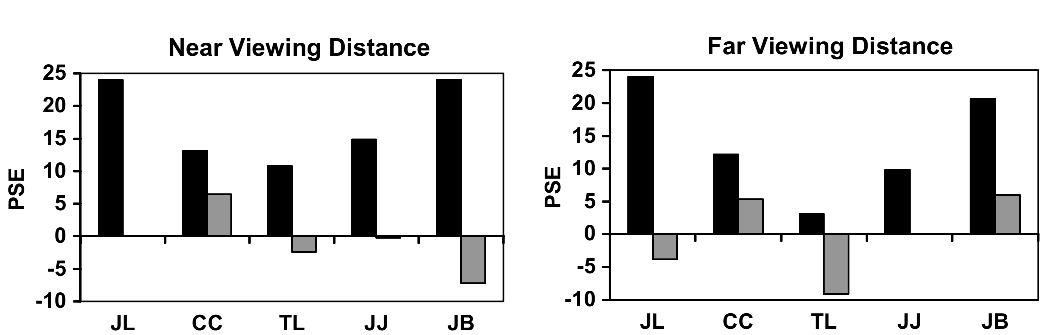

Fig. 4. Points of subjective equalities (PSEs) for the five experimental group subjects at the near (left) and far (right) viewing distances. Pre-test PSEs are given by black bars and post-test PSEs are given by gray bars.

between stereo and haptic cues to the distance between front and rear surfaces. Whereas experimental subjects performed inconsistent-cue training trials during sessions on Days 3 and 4, control subjects performed consistent-cue training trials. For control subjects, sessions on Days 3 and 4 were identical to those on Days 1 and 2. The results for the control subjects are given by the two rightmost bars in Fig. 5. At both near and far viewing distances, the average shifts in their PSEs were not significantly different than zero (near viewing distance: $T(2) = 0.076$; $p = 0.947$; far viewing distance: $T(2) = 0.600$; $p = 0.609$). We conclude that relative depth-from-stereo estimates did not change due to practice alone.

2.3. Discussion

The first striking feature of the data is the large PSEs found in pre-test. When the relative depth of the front and back surfaces was equal to the width of the front surface (as specified by the viewing geometry), subjects perceived the width of a surface to be significantly larger than its depth. In theory, this effect could result from subjects underestimating the relative depth of the surfaces (e.g., because of flatness cues in the display), from overestimating the width of the front surface, a bias in the comparison process required to make the judgments, or some combination of the above. The hypothesis that subjects underestimated relative depth is argued against by the combined experimental and control data. The post-test data from the inconsistent-cue condition show that haptic feedback can change visual judgments of relative depth, yet no changes appeared in post-test for the control condition in which consistent haptic feedback was provided. If subjects were initially underestimating relative depth from vision, however, the control condition would have effectively been a cue conflict condition (unless haptic depth just happened to be underestimated by the same factor as visual depth) and would have led to a change in visually perceived relative depth. This suggests that the perceptual bias reflected in the pre-test PSEs derives from biases in the apparent width of the front surface or in the process by which visual and haptic estiates are compared. Further, visually perceived relative depth was, on average, equivalent to haptically perceived relative depth when visual and haptic cues were consistent.

Since subjects were not given haptic information about the width of the stimulus surfaces, the shifts in subjects' PSEs for width versus relative depth judgements can best be explained by changes in their visual estimates of relative depth, induced by the depth percepts obtained from haptic information. Two broadly different types of perceptual mechanisms could underlie the learning effects: recalibration of relative depth-from-stereo estimates or re-weighting of relative depth estimates from stereo disparity and from other cues and prior biases.

2.3.1. Recalibration

Eq. (1) shows that in order to estimate relative depth-from-stereo disparity, the visual system must scale the disparity measurements by a gain factor dependent on the absolute viewing distance and the interocular distance:

$$\Delta_{\mathrm{V}} = (D_{\mathrm{V}})^2 * \eta / I. \tag{2}$$

Similarly, in order to estimate the real width of the front surface, Δ_{w}, from its projected width, ω, the visual system must scale the projected width by the viewing distance:

$$\Delta_{\mathrm{w}} = (D_{\mathrm{V}}) * \omega. \tag{3}$$

Thus, the perceived ratio between the width of the front surface and its depth relative to the rear surface is given by

$$R = (1/D_{\mathrm{V}}) * (\omega * I/\eta). \tag{4}$$

These considerations suggest that a logical locus of recalibration is in the estimate of viewing distance.

The virtual display used in the current experiments contained two principal cues to viewing distance—accommodation and the vergence angle of the eyes. Because accommodation was uninformative in the experiment (it was fixed at a constant 0.914 m in the virtual reality goggles), only vergence provided useful information about viewing distance in the test conditions. This cue alone is not perfectly reliable and, thus, one might expect the gain factor on disparity to be malleable and subject to recalibration based on feedback from other sensory modalities such as haptics (see Judge & Miles, 1985; Maddox, 1893, discussed above).

We must also consider two possible sites for recalibration. The first is a recalibration of a single estimate of viewing distance that is used to determine the gain factor for both width and depth percepts. The second is a recalibration that is localized in the relative depth-from-disparity calculation. The former model predicts a post-adaptation change in PSE proportional to the difference between the reaching distance suggested by the haptic cue and the viewing distance suggested by vergence. At the near viewing distance, this model predicts that the perceived ratio of width to depth in the post-test stimuli should shrink by 14% relative to pre-test, leading to a 14% decrease in the PSE. At the far viewing distance, it predicts an 11% change. The width of the front surface on post-test was either 44 or 56 mm, so the model predicts that on average, PSEs should change by an average of 6.9 mm for the near test surfaces and an average of 5.8 mm for the far test surfaces. The measured changes in subjects' judgments due to adaptation were slightly greater than twice these values.

The latter model, a local recalibration of the relative depth-from-disparity computation, predicts a much larger shift in PSE, since it is not balanced by an increase in the perceived width of the front surface. Eq. (2) predicts a decrease in PSE of 13.0 mm for the near viewing distance and 10.9 mm for the far viewing distance. These values are within the standard error of our estimates; thus, if recalibration was the cause of the post-adaptation change in PSEs, the data suggest that this recalibration is local to the relative depth-from-disparity computation. Our data does not allow us to determine the specific computation underlying the recalibration. It could be a simple recalibration of the viewing distance estimate used to scale disparity measurements. It could involve a re-weighting of the cues to viewing distance to bias estimates of this distance toward the value suggested by the accommodative state of the eye or to a default value greater than that used in the displays. Finally, it might not, strictly speaking, involve a recalibration of viewing distance at all, but rather a straightforward change in the gain on the output of the relative depth-from-disparity calculation.

2.3.2. Cue re-weighting

An alternative account of the learning effects is that haptic feedback leads to a change in the weights assigned to stereo estimates of relative depth versus estimates of relative depth derived from other cues and prior biases. Consider the following simple and popular model in the vision sciences literature, which we refer to as the linear-weighted-average (LWA) model. Let $d_s(s)$ denote a subject's estimate of the relative depth between front and back surfaces based solely on the pattern of binocular disparities, denoted s, present in a pair of left and right images. Let d_0 denote a relative depth value derived from other cues (e.g., accommodation) and prior biases assumed by a subject. According to the LWA model, a subject's estimate of the relative depth between the two surfaces, denoted d^*, is a weighted average of $d_s(s)$ and d_0 with weights w_s and w_0:

$$d^* = w_s d_s(s) + w_0 d_0,$$

where the weights are assumed to be non-negative and to sum to one.

When faced with a conflict between visual and haptic estimates of depth, the visual system might adjust the weights of the visual cues so that the cue(s) most consistent with the haptic information are given greater weight. This model can account for the results of Experiment 1 in two ways. First, suppose that non-stereo estimates of d_0 are relatively large; in fact, are larger than $d_s(s)$ for all pairs of left and right images used in the experiment. If these were more consistent with the haptic feedback, the visual system might adjust by giving more weight to the non-stereo cues, leading to greater relative depth estimates on the post-test trials than on the pre-test trials. Alternatively, suppose that non-stereo estimates of d_0 are relatively small. In this case, the overall shift in subjects' depth estimates could have been produced by increasing the weight given to stereo information. Experiment 2 was designed to test this cue re-weighting account.

3. Experiment 2

Experiment 2 tested whether subjects could adapt to a situation in which the visuo-haptic discrepancy took one form when the two fronto-parallel surfaces were near the subject and a different form when the surfaces were far from the subject. Specifically, when the two surfaces were near the subject, the haptic distance in depth between the two surfaces was greater than the visual distance (435 and 375 mm, respectively), whereas the haptic distance was greater than the visual distance when the surfaces were far from the subject (525 and 585 mm, respectively). Since simple re-weighting of stereo and non-stereo cues in the LWA model discussed above can only lead to proportionally similar effects in the two conditions, a finding that subjects performed differently in the near and far conditions would argue against the cue re-weighting hypothesis.

3.1. Methods

3.1.1. Stimuli

Stimuli were generated in a manner similar to Experiment 1. Stimuli for training trials were presented at two different viewing distances (375 and 585 mm). On consistent-cue training trials, the reaching distance used to create the haptic cue (including the relative depth between the two surfaces) was the same as the viewing distance used to create the visual images. On cue-inconsistent training trials, the reaching distance used to create the haptic cue was set to 435 mm for the near viewing distance and to 525 mm at the far viewing distance. At the near viewing distance, the haptically rendered relative depth between the two surfaces was approximately 35% greater than the visually rendered relative depth. At the far viewing distance, it was 20% less than the visually rendered depth. Other features of the stimuli (e.g., surface width and height) were the same as in Experiment 1.

3.1.2. Procedure

The procedure for Experiment 2 was identical to Experiment 1 with the following exceptions. Subjects participated in experimental sessions on six days. On Day 1, subjects performed 210 consistent-cue training trials followed by 168 test trials. Test trials were conducted at near (395 mm), medium (475 mm), and far (555 mm) viewing distances. The viewing distances for the near and far test stimuli were chosen to be between the corresponding viewing and reaching distances used in the training trials. The viewing distance for the medium condition was intermediate between these two. Because the session on Day 1 is regarded as a practice session during which subjects were in the process of learning about the experimental environment and task, their responses on the test trials during this session were not analyzed. On Day 2, subjects performed 126 consistent-cue training trials followed by 336 test trials. This set of test trials is referred to as the pre-test. On Days 3–6, subjects performed a set of inconsistent-cue training trials (Day 3: 210 trials; Days 4–5: 168 trials; Day 6: 126 trials) followed by a set of test trials (Days 3–5: 168 trials; Day 6: 336 trials). Subjects were in the process of learning about the inconsistent-cue environment during Days 3–5, and so their responses on the test trials on these days were not analyzed. The data from the set of test trials on Day 6 was analyzed. This set of trials is referred to as the post-test. [2]

3.1.3. Subjects

Subjects were seven undergraduate and graduate students at the University of Rochester. They had normal or corrected-to-normal vision and were naive to the purposes of the study.

3.2. Results

Results for one subject are shown in Fig. 6 (subject DA). The top, middle, and bottom graphs correspond to the near, middle, and far viewing distances, respectively. The horizontal axis of each graph plots the difference between the relative depth between the two fronto-parallel surfaces and the width of the nearer surface; the vertical axis plots the probability that the subject judged the distance in depth as greater. The solid line shows a

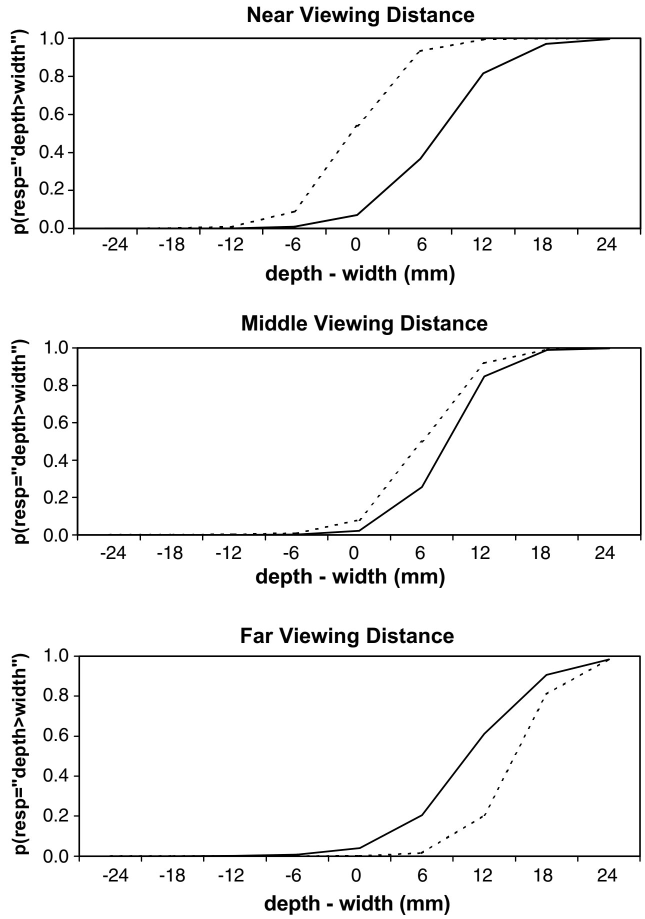

Fig. 6. Pre-test (solid line) and post-test (dotted line) logistic fits for subject DA at the near, middle, and far viewing distances. This subject showed context-sensitive recalibration of his depth-from-stereo estimates: from pre-test to post-test, his data show a shift to the left at the near viewing distance (an increase in his depth-from-stereo estimates), nearly no shift at the middle viewing distance, and a shift to the right at the far viewing distance (a decrease in his depth-from-stereo estimates).

logistic function which has been fit to the subject's pre-test data, and the dotted line shows a function which has been fit to his post-test data. At the near viewing distance (top graph), the subject's data shifts to the left from pre-test to post-test, indicating that the subject increased his depth-from-stereo estimates at the near viewing distance. There is only a small difference between the subject's pre-test and post-test data at the middle viewing distance (middle graph). At the far viewing distance (bottom graph), the data shifts to the right from pre-test to post-test, indicating that the subject decreased his depth-from-stereo estimates at this viewing distance.

The pattern in this subject's data was often found in the data of the remaining subjects, as illustrated in the graph in Fig. 7. The horizontal axis of this graph plots the viewing distance, and the vertical axis plots the subjects' average PSE value at that viewing distance. The black line is for the pre-test data; the gray line is for

[2] Two caveats should be noted. First, subject MC found it uncomfortable to grasp the front and rear surfaces when they were at a far reaching distance. To accommodate this subject, the far reaching distance was moved 20 mm closer to the subject. Second, the experiment was modified in a small way on inconsistent-cue training trials when the two fronto-parallel surfaces were at the far viewing and reaching distance. In this case only, the smallest haptic distance between the surfaces was too small to be faithfully represented by the virtual reality experimental apparatus. Consequently, the smallest depth was not used at this viewing and reaching distance; instead it was replaced by a new largest depth. This change inadvertently led to an imbalance in the number of trials in which the haptic distance between the two surfaces was greater than, as opposed to less than, the visual width of the nearer surface; there were now twice as many 'deeper' trials than 'wider' trials at the far viewing and reaching distance on inconsistent-cue training trials. One would reasonably expect that this bias would tend to lead subjects toward increasing their depth-from-stereo estimates at the far viewing and reaching distance. In fact, this did not occur, and so we conclude that this imbalance was not an important factor underlying our experimental results.

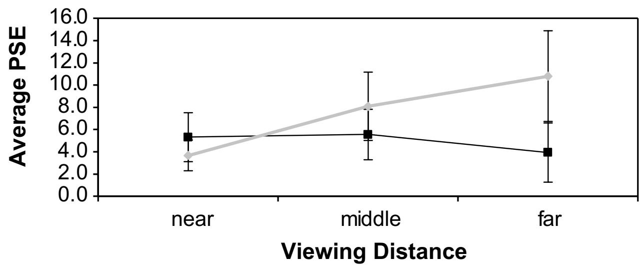

Fig. 7. The subjects' average PSEs at the near, middle, and far viewing distances. The black line is for the pre-test data, and the gray line is for the post-test data.

the post-test data (error bars give the standard errors of the means). The pre-test PSEs are relatively similar across viewing distances, but the post-test PSEs increase with viewing distance.

Fig. 7 compares average PSEs on pre-test trials versus post-test trials, but it does not compare pre-test and post-test PSEs on a subject-by-subject basis. An analysis of individual subject's data was performed as follows. For a given set of trials (either pre-test or post-test), logistic functions were fit to a subject's responses at the near, middle, and far viewing distances, and the subject's PSE at each viewing distance was computed. Next, a line was fit to these three PSE values, and the slope of this line (referred to as the PSE slope) was recorded. Fig. 8 shows the pre-test (black bars) and post-test (gray bars) PSE slopes for each of the seven subjects. All subjects have larger post-test PSE slopes than pre-test PSE slopes (the average difference between a subject's post-test and pre-test PSE slopes is significant greater than zero: $T(6) = 3.133$; $p = 0.0203$).

3.3. Discussion

The experimental data indicate that subjects' judgments on the post-test trials differed from their pre-test judgements due to their experiences during the inconsistent-cue training trials. Specifically, they adapted their depth-from-stereo estimates so that these estimates were larger at a near viewing distance and smaller at a far viewing distance. Due to these adaptations, subjects' depth-from-stereo estimates were modified so that they were in greater agreement with their depth-from-haptic estimates at all viewing distances. We now consider these results in light of the models discussed earlier.

3.3.1. Recalibration

The simplest explanation of the results of Experiment 2 is that subjects recalibrated their absolute depth-from-vergence estimates to accord with the haptic cue. To appropriately accommodate the error signal from the haptic feedback, however, subjects could not simply have adjusted a simple gain factor on their absolute depth estimates. Rather, they would have had to learn to bias their absolute depth estimates toward a position midway between the extreme depths used in the experiment, a form of preferred distance bias. This would have resulted in a positive bias for near viewing distance stimuli and a decrease in PSE (increase in perceived relative depth versus width ratio), as observed, and a negative bias for far viewing distance stimuli and consequent increase in PSE, as observed. It should have led to a minimal change in PSE at the intermediate viewing distance.

Alternatively, subjects may have adapted their estimates of relative depth-from-stereo in a context-specific, local manner, effectively learning a positive gain on perceived relative depth-from-stereo at near viewing distances and a negative gain on perceived depth-from-stereo at far viewing distances. Performance on test trials at the intermediate viewing distance is not clearly predicted by the local adaptation hypothesis, but one might expect it to result from interpolation of the local adaptation rules learned at near and far distances. Evidence for this type of context-dependent learning and interpolation has been found in the domain of prism adaptation (Ghahramani & Wolpert, 1997). Our data cannot differentiate between these recalibration models.

The results show a small overall increase in post-test PSEs over all test trials relative to pre-test PSEs. This is not predicted by either of the recalibration hypotheses.

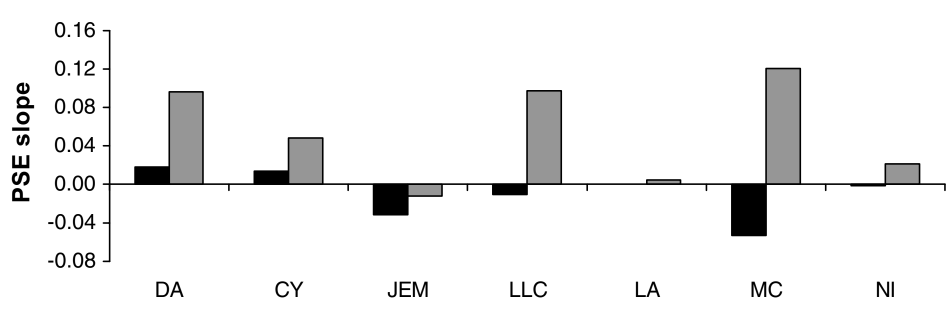

Fig. 8. A line was fit to each subject's PSEs at the near, middle, and far viewing distances. This graph plots the slope of this line on the pre-test (black bars) and post-test (gray bars) trials. To calculate the slope, we replaced the x-coordinate in Fig. 7 with the geometrically specified viewing distances used in rendering the test stimuli.

However, it could be easily reconciled with these hypotheses by positing either a learned bias in perceived viewing distance that is closer to the far distance used in the experiment (for the first recalibration model) or a stronger local adaptation effect at the far viewing distance (for the second recalibration model).

3.3.2. Cue re-weighting

Simply learning to re-weight disparity cues relative to other depth cues (accommodation, blur, prior flatness biases, etc.) cannot account for the differential test effects found at different viewing distances. This would seem to eliminate cue re-weighting in the LWA model as an account for the learning effects found here. In general, however, cue re-weighting in a linear model can fit just about any learning effects so long as the model is given the ability to learn and apply new weights in a context-specific manner (e.g., learn to down-weight disparity information at one viewing distance and up-weight it at another).

Regardless of the underlying mechanism, the learning effects were considerably smaller in the second experiment than in the first. In the first experiment, subjects PSEs indicated a full adaptation of the visually perceived relative depth to the haptically specified depth. In the current experiment, complete global recalibration of viewing distance to match the corresponding reaching distance would have predicted an average difference in PSE between near and far viewing distances of approximately 12 mm assuming that recalibration of viewing distance affects both perceived width and perceived relative depth, while complete recalibration for relative depth alone would have predicted a difference of approximately 23 mm in PSE between the two viewing distances. These values are approximate, because testing was done at slightly different depths than training. The average difference in post-test PSEs between near and far viewing distances in Experiment 2 was only 7.2 mm, slightly more than half of that predicted by the more conservative model. The decreased adaptation effect found in Experiment 2 likely reflects the increased difficulty of learning opposite signs of relative depth change at different viewing distances.

4. Concluding remarks

Perceptual environments are highly redundant, meaning that they provide observers with cues from multiple sensory modalities. As noted by many investigators, observers can take advantage of cue redundancy during visual learning. In this article, we studied the hypothesis that observers can recalibrate their percepts of depth-from-stereo when visual and haptic cues are discordant. Our findings indicate that when stereo and haptic cues to depth are inconsistent, observers recalibrate their interpretations of the stereo cue so that depth-from-stereo percepts are in greater agreement with depth-from-haptic percepts. Overall, the results suggest that observers' visual and haptic percepts are tightly coupled in the sense that haptic percepts provide a standard to which visual percepts can be recalibrated when they are deemed to be erroneous.

Acknowledgements

We thank two anonymous reviewers for their comments on an earlier version of this manuscript, D. Ballard and M. Hayhoe for allowing us to use their lab facilities, and M. von der Heyde for creating the experimental environment. This work was supported by NIH grants RO1-EY13149, RO1-EY13319, and P41-RR09283, and by NSF Graduate Fellowship DGE-9616170.

References

Adams, W. J., Banks, M. S., & van Ee, R. (2001). Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience, 4*, 1063–1064.

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research, 41*, 449–461.

Berkeley, G. (1709). An essay towards a new theory of vision. In C. M. Turbayne (Ed.), *Works on vision*. Indianapolis: Bobbs-Merrill.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. In W. Epstein, & S. Rogers (Eds.), *Perception of space and motion*. San Diego: Academic Press.

Epstein, W. (1975). Recalibration by paring: a process of perceptual learning. *Perception, 4*, 59–72.

Epstein, W., & Morgan, C. L. (1970). Adaptation to uniocular image magnification: modification of the disparity–depth relationship. *American Journal of Psychology, 83*, 322–329.

Ernst, M. O., Banks, M. S., & Bülthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience, 3*, 69–73.

Ghahramani, Z., & Wolpert, D. M. (1997). Modular decomposition in visuomotor learning. *Nature, 386*, 392–395.

Hershenson, M. (1999). *Visual space perception: a primer*. Cambridge, MA: MIT Press.

Judge, S. J., & Miles, F. A. (1985). Changes in the coupling between accommodation and vergence eye movements induced in human subjects by altering the effective interocular separation. *Perception, 14*, 617–629.

Maddox, E. E. (1893). *The clinical use of prisms*. Bristol: John Wright and Sons.

Wallach, H. (1985). Learned stimulation in space and motion perception. *American Psychologist, 40*, 399–404.

Wallach, H., Moore, M. E., & Davidson, L. (1963). Modification of stereoscopic depth perception. *American Journal of Psychology, 76*, 191–204.