Exp Brain Res
DOI 10.1007/s00221-007-1099-5

RESEARCH ARTICLE

Mental concatenation of perceptually and cognitively specified depth to represent locations in near space

Bing Wu · Roberta L. Klatzky · Damion Shelton · George Stetten

Received: 19 April 2007 / Accepted: 3 August 2007

Abstract The purpose of this study was to examine how discrete segments of contiguous space arising from perceptual or cognitive channels are mentally concatenated. We induced and measured errors in each channel separately, then summed the psychophysical functions to accurately predict pointing to a depth specified by both together. In Experiment 1, subjects drew a line to match the visible indentation of a probe into a compressible surface. Systematic perceptual errors were induced by manipulating surface stiffness. Subjects in Experiment 2 placed the probe against a rigid surface and viewed the depth of a hidden target below it from a remote image with a metric scale. This cognitively mediated depth judgment produces systematic under-estimation (Wu et al. in IEEE Trans Vis Comput Grap 11(6):684–693, 2005; confirmed here). In Experiment 3, subjects pointed to a target location detected by the indented probe and displayed remotely, requiring mental concatenation of the depth components. The model derived from the data indicated the errors in the components were passed through the integration process without additional systematic error. Experiment 4 further demonstrated that this error-free concatenation was intrinsically spatial, rather than numerical.

B. Wu (✉) · R. L. Klatzky
Department of Psychology, Carnegie Mellon University,
Pittsburgh, PA 15213, USA
e-mail: bingwu@andrew.cmu.edu

B. Wu · D. Shelton · G. Stetten
Robotics Institute, Carnegie Mellon University,
Pittsburgh, USA

R. L. Klatzky
Center for the Neural Basis of Cognition,
Carnegie Mellon University, Pittsburgh, USA

G. Stetten
Department of Bioengineering, University of Pittsburgh,
Pittsburgh, USA

Animals, including humans, possess remarkable abilities to construct a global spatial representation from pieces of local information. The problem of how such spatial integration is achieved has been studied at different levels and from a variety of perspectives. At the level of early perception, there is work on how the brain combines information from successive saccades into a common spatial map (e.g., Irwin and Andrews 1996). Other research has examined intermediate-level processes that combine successive views. For example, a large expanse of the ground surface can be accurately represented in the visual system by integrating the view of patches of the ground seen through an aperture (Wu et al. 2004). Still another level of analysis can be found in the literature on spatial cognition, which has investigated how organisms come to conjoin "cognitive maps" that are learned independently (Blaisdell and Cook 2005; Golledge et al. 1985; Holding and Holding 1988; Moar and Carleton 1982; Sturz et al. 2006). Neuroscience research has pointed to the critical role of hippocampal neurons in extracting place information from sensory inputs and forming a cognitive map representing the geometry of the environment (Battaglia et al. 2004; O'Keefe and Burgess 1996).

The construction of space from discrete samples often involves assembling information from different sources. Space may be combined across different sensory modalities, as when one reaches into a visible opening to find a hidden object by touch. Also, contiguous spatial extents may be specified not only perceptually, but symbolically.

For example, when giving instructions, people may combine portions of space that are visible ("go to *that corner*") and those that are not ("and walk *another half mile*"). To derive representations of space from such symbolic descriptions, interpretive processes are required to translate them into spatial form. We thus refer to the resulting representations as *mediated*.[1]

The present research concerns how perceptual and mediated representations of space are integrated, in particular, whether such integration specifies a level of metric precision that supports action. Our particular concern is with people's ability to concatenate abutting extents in space, when one extent is specified through direct perception and the other, being symbolic in nature, must be subjected to mediating processes.

Perhaps surprisingly, there is substantial evidence that mediation allows for representations that are intrinsically spatial, to the point that they can function to guide action in the same way as those created through perception. For example, spatial language has been shown to be functionally equivalent to vision in encoding locations and directing walking to specific destinations (Avraamides et al. 2004; Klatzky et al. 2003; Loomis et al. 2002). Maps and diagrams, which are arbitrarily rescaled representations of environments, constitute another source of cognitively mediated spatial representation. Simple graphics have been shown to provide sufficient information to support walking and estimating distance and direction in the represented world (Richardson et al. 1999). From such results, researchers have suggested that separate streams of spatial information, perceptual and cognitive, converge on a single representation of the environment that supports action (Bryant 1992; Loomis et al. 2002). Klatzky et al. (2003) suggested that posterior parietal cortex may be a brain locus for spatial representations that are essentially amodal.

Because mediated representations of space result from very different neural processing than directly perceiving space, the question of how these can be combined is an important one. To address how people concatenate discrete segments of contiguous space, when the segments are conveyed by symbolic input and direct perception, we use a method we term "error pass-through". This method establishes patterns of systematic error that arise from encoding each of the two contiguous components, one perceived and one mediated, and then independently assesses errors in the representation of the two together. It assesses whether the integration process adds additional systematic error, or alternatively, simply passes through the error observed for each component in isolation.

[1] We use this term to refer to internal processes that construct a representation, not in Gibson's (1979) sense of a mediating stimulus like a picture that depicts the external world.

In the present task, participants made pointing responses to targets within what Cutting and Vishton (1995) called *personal space*. By use of ultrasound, the total target depth was parsed into two contiguous segments. As shown in Fig. 1, the participant indented an ultrasound probe into a surface until it "bottomed out" against a barrier, then read the depth of a target below the barrier by viewing an image on a remote screen. The first depth segment—the extent of indentation—was perceptually cued by vision and touch, and the second—depth in the image—was presented symbolically, requiring cognitively mediated processing as opposed to direct perception. In order to locate the target and point accurately, the participant would have to mentally integrate the two depths, ideally by linear summation. We wish to detect if the integration adds systematic error beyond those observed in the components.

What error levels might we expect of the component processes? Let us first consider the indentation of the probe, as perceived by the modalities of vision and touch ($d_{\text{perceptual}}$ in Fig. 1). Experimental evidence suggests that error can arise from both perceptual modalities, particularly at small scale. It has been reported that visual judgments of length in near space are proportional to physical length, as indicated by power-function exponents of 1.0 (Seizova-Cajic 1998; Teghtsoonian and Teghtsoonian 1965); however, proportion relationships do not guarantee veridicality. For example, Keyson (2000) reported a linear function with a slope of 0.877 relating visually perceived length to physical length, which would introduce considerable systematic under-estimation. Haptic length perception generally yields more error than vision. The power-function exponents reported tend to vary between 0.8 and 1.2, depending on variables such as the

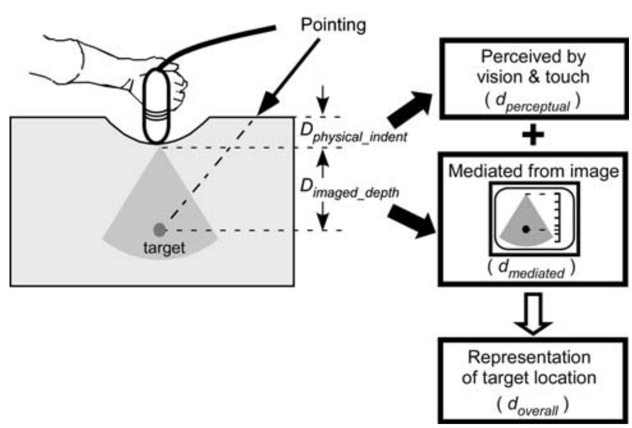

Fig. 1 Target localization requiring the integration of depth information from perception and cognitive mediation. $D_{\text{physical\_indent}}$ and $D_{\text{imaged\_depth}}$ denote the physical value of indentation and target depth relative to the probe tip. Correspondingly, the judged depths are $d_{\text{perceptual}}$ and d_{mediated}, respectively. d_{overall} denotes the judged total depth by combining two depth components

range of lengths and measuring methods used (Lanca and Bryant 1995; Seizova-Cajic 1998; Stanley 1966; Teghtsoonian and Teghtsoonian 1965, 1970). For example, whereas subjects judged length relatively accurately if the fingers of separate hands were used or the length was traced with the fingertip (Stanley 1966; Teghtsoonian and Teghtsoonian 1965), they overestimated the size of stimuli held between the thumb and forefinger (Seizova-Cajic 1998; Teghtsoonian and Teghtsoonian 1965, 1970). In addition to observing any intrinsic errors in depth perception by the individual modalities, we specifically manipulated error here by capitalizing on inter-sensory interactions. The perception of stimulus properties by touch and vision is known to be subject to cross-modal influence (eg., Lederman et al. 1986; Srinivasan et al. 1996). In particular, forces involved in exploring objects have been found to influence perception of their geometric properties (Drewing and Ernst 2006; Robles-De-La-Torre and Hayward 2001; Wydoodt et al. 2006). We demonstrate here the illusion that resisting forces arising from penetration heighten the judgment of perceived depth of indentation, and we use this to modulate error in the perceived depth segment.

Next consider what might be expected with regard to error in the mediated depth judgment (d_{mediated} in Fig. 1). The probe-relative depth specified in the image was numerically precise, but arbitrarily related to the physical scale at which it was portrayed; moreover, the display was separated in space from the action response. Our previous research on this task (Wu et al. 2005) revealed systematic errors in achieving the requisite representation. Given targets located at depths from 3.5 to 6.5 cm and localized by ultrasound, subjects pointed to locations that were systematically displaced upward relative to the true target placements. Two further findings showed that the pointing response provides a direct measure of the represented target location. First, the locations indicated by pointing coincided with those indicated by reaching responses. Second, a control experiment, in which the targets were visible and accurately perceived, showed that pointing per se introduced no systematic error.

Having constructed paradigms for assessing depth as judged through perceptual and mediational processes, with both expected to show systematic error, we can now consider how the errors are passed through a spatial integration process that concatenates them to compute a value called d_{overall}. As was mentioned above, accurate pointing to the target requires achieving a spatial representation that sums the two depth components—unindented surface to probe tip; probe tip to target. If the concatenation process introduces no further error beyond those in the component depth estimates, the error in the final pointing response should be equal to the sum of the errors in the components. More specifically, the psychophysical function relating the depth response to the actual depth value should be the sum of the psychophysical functions for the components. However, it is possible that the concatenation will introduce error, so that the equation for the sum is:[2]

$$d_{\mathrm{overall}} = d_{\mathrm{perceptual}} + d_{\mathrm{mediated}} + \mathrm{err}_{\mathrm{integration}}\,. \qquad (1)$$

A series of four experiments were conducted to assess the errors in the components and the integrator. The first three experiments implemented the error pass-through method by measuring the errors in the individual depth components and the integrated combination. Experiment 1 and 2 required subjects to localize targets at depths that were encoded through perception or mediation, respectively. From these studies, systematic errors in each process were identified and measured. Experiment 3 then examined their performance on a pointing task requiring the conjoining of the component depths and confirmed that the integrator added no additional error. Finally, Experiment 4 tested whether the integration process is intrinsically spatial, or alternatively, simply a process of numerical addition.

Experiment 1. Perception of indentation from visual and haptic cues

This experiment measured errors in the participants' perception of probe tip indentation, $d_{\mathrm{perceptual}}$. We manipulated not only the objective physical indentation, but also the surface stiffness, which introduced a resisting force. If participants interpreted greater resisting force at the stopping point of the probe as deeper penetration, perceived indentation should increase with final resisting force (i.e., stiffness) as well as physical indentation. Systematic error tendencies would then be passed forward to the process that combines perceived indentation with perceived target depth.

Method

Participants

Sixteen undergraduate students, naïve to the purpose of the experiment, were tested. All had normal or corrected-to-normal vision and stereo acuity better than 40″ of arc.

[2] There are no coefficients on $d_{\mathrm{perceptual}}$ and d_{mediated} in our model, because the concern here is with the concatenation of two distinct spatial segments. There is no reason to assume that the spatial estimates are differentially weighted at the point where they are combined. It should be noted that in this respect the present study is different from those of cue integration models, where different sources compete for a description of the same region of space. As will be shown in the final results, the data are well fit without coefficients.

Stimuli and procedure

The stimuli were water tanks with specially-built two-tier lids as shown in Fig. 2a. The center part of each lid (8 cm in diameter) was cut away and then overlaid by a soft, textured rubber layer. Under it was placed a rigid plastic grid to support the ultrasound probe at the desired penetration depth. In addition, resistive bands were placed just beneath the upper layer to manipulate surface stiffness. There were two sub-designs, requiring the creation of nine tanks, with different combinations of stiffness and penetration: Five penetration depths (0.4, 1.0, 1.5, 2.0 and 2.6 cm) were tested at the minimum stiffness level 0.93 ± 0.12 N/cm. Penetration depths of 1.0 cm and 2.0 cm were tested at two additional stiffness levels, 2.53 ± 0.35 and 6.51 ± 0.88 N/cm. (Stiffness values, determined by static load tests, are averaged over tanks having the same nominal stiffness level. Some variability arose, as shown by the SD, because each tank was hand-fabricated; however, the variability was small relative to the difference in levels of stiffness.)

Subjects performed the experiment binocularly. In each trial, the ultrasound probe was held by the subject and placed on the indentation so that it "bottomed out" on the rigid under-surface. The subject was asked to estimate the extent of the indentation and then to draw a vertical line that matched it. Each indentation was tested twice at random. No feedback was given to the subjects regarding their performance.

In addition, in order to assess possible systematic errors due to the measurement method per se, another group of subjects ($n = 16$) was tested in a line-duplication task using the same procedure, where the stimuli were five lines of length equal to the depths tested at the minimum stiffness level. The subjects performed the draw-a-matching-line task twice for each stimulus without feedback.

Results

The open circles in Fig. 2b represent the average length of lines drawn in the line-duplication task. Virtually no systematic error was shown in the drawing response; the data were fit by a line with slope of 0.982 ($r^2 = 0.998$, $P < 0.001$). In contrast, the lengths of the lines drawn to match the perceived indentation (solid circles) were significantly shorter than the physical depths. Clearly, the indentation was under-estimated at this lowest level of stiffness. The data were best fit by $d_{\text{perceptual}} = 0.695 \times D_{\text{physical\_indent}}$ ($r^2 = 0.993$, $P < 0.001$).

Figure 2c plots the mean judged indentation of the two depth values that were observed across variations in stiffness. It clearly shows, as expected, that $d_{\text{perceptual}}$ increased as the surface stiffness increased ($F(2,30) = 65.611$, $P < 0.001$). The interaction between physical depth and stiffness was also significant ($F(2,30) = 10.038$, $P = 0.001$). Additionally, another group of subjects ($n = 12$) in Experiment 3b was tested in the same task with

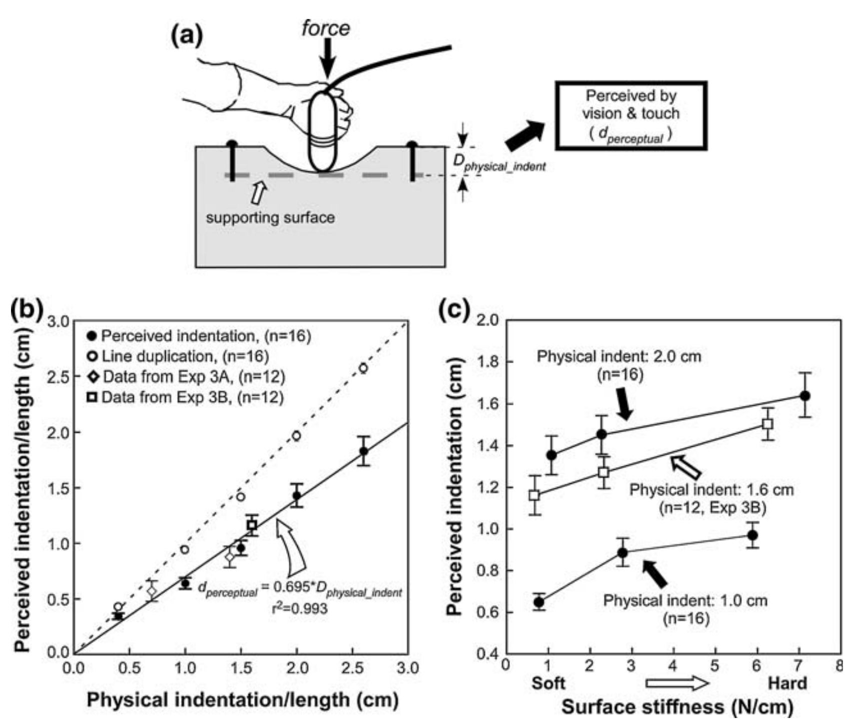

Fig. 2 **a** The experimental setup for the measurement of perceived indentation. **b** *Solid circles* show the mean judgments at the lowest stiffness for each physical indentation, with a linear regression fit to the data. *Open circles* show the average length drawn in the line-duplication task in which the stimuli were lines of length equal to the indentations tested. Also shown (*open square* and *diamonds*) are data obtained from Experiment 3, using different groups of subjects in the same task. **c** Mean perceived indentation as a function of surface stiffness (*filled circles*). Also shown (*open squares*) are data from Experiment 3 with a physical indentation of 1.6 cm using a different group of subjects. Error bars represent the standard error of the mean

a penetration depth of 1.6 cm. Their results (open squares in Fig. 2c) showed the same pattern ($F(2,22) = 26.828$, $P < 0.001$).

It is evident from these results that $d_{\mathrm{perceptual}}$ is affected by visual cues to surface deformation, here including binocular depth and textural distortions (Devisme et al. 2005), and also by stiffness, which would affect the resisting forces that arose when the probe compressed the surface. This is consistent with previous findings that the perceptual judgments of 3D geometric properties such as surface slant (Ernst et al. 2000; Ernst and Banks 2002) and curvature (Drewing and Ernst 2006) are jointly specified by visual and haptic cues. We further tested whether the cue combination could be described by a weighted linear combination. A model with two variables ($d_{\mathrm{perceptual}} = 0.61 \times D_{\mathrm{physical\_indent}} + 0.031 \times \mathrm{Force}$) was found to describe the data well ($r^2 = 0.981$, $P < 0.001$). Here force is defined as the product of stiffness and the indentation at the bottom-out position; the coefficient of Force is a distortion factor in (perceived cm)/N. For a constant physical indentation, higher force leads to greater perceived indentation. This constitutes an illusory error that would be passed on to the integration process that combines perceived indentation with the contiguous target depth encoded by mediational processes. Experiment 2 examined the latter component.

Experiment 2. Mediated perception of 3D locations

The experiment was to assess errors in the mediated representation of target depth, as defined relative to the probe tip. Rather than being visible to the subjects, target depths were indicated graphically. The representation of target depth was determined using a triangulation-by-pointing paradigm. Given our previous results (Wu et al. 2005), systematic under-estimation was expected.

Method

Participants

Twenty-four subjects were tested. All had normal or corrected-to-normal vision and stereo acuity better than 40″ of arc.

Stimuli and procedure

The stimuli included four water tanks, which were the same as those in the previous experiment except for having no indentation of the lid. Under the cutaway section of the lid were four beads (1.0 cm in diameter) mounted at different depths (3.5, 5.0, and 6.5 cm for the target beads and a random depth for an additional dummy). The *x*–*y* locations of these beads were marked on the tank cover and labeled with the numbers 1–4 in a random pattern. Their relative locations were fixed within a tank but varied between tanks so as to create an unpredictable stimulus environment.

Targets were observed as ultrasound images, displayed on a 5″ LCD that also provided a numerical metric for depth. In addition, to quantify the measurement error in the experimental procedure, an augmented-reality display called the Sonic Flashlight, which enables observers to directly see the target by displaying its ultrasound data in 3D space as a virtual image at its actual location (Stetten and Chib 2001), was used as a control. We have shown (Wu et al. 2005) that the Sonic Flashlight is equivalent to direct viewing of the target in terms of the pointing task.

Subjects performed the experiments binocularly. On each trial, the subject held the ultrasound probe upright and placed it over one of the four beads. After obtaining a clear image of the target, the subject estimated its location. Numerical depth in cm could be read from the image, and the subject had reference to a standard 1 × 1 cm grid printed on a sheet of paper placed on the table. To demonstrate the judged location, the subject pointed a stylus mounted with a tracker (miniBIRD) at the judged location of the target. The subject pointed the stylus in turn from four positions on the lid to complete a trial. After all targets in a tank had been tested, the next tank was introduced. The test order of devices and tanks was counterbalanced across subjects. No feedback was provided.

The judged location of each target was computed off-line by combining the subject's four pointing responses at that target using an algorithm described in Wu et al. (2005). Ideally, all lines along which the subject pointed at a given target would converge to where the subject judged the target to be, and hence the perceived location could be computed using any pair of pointing lines. In reality, there exists no exact intersection of pointing lines due to variability in human performance and measurement noise. To assess the represented location, we first estimated it from each pair of pointing lines. Midpoints were found between each pair of pointing lines, where the lines passed closest to each other. With many pairs of pointing lines, the intersection points would form a volume around the represented location of the target. The centroid of those points was then used as an estimation of the location encoded by the subject.

Results

Figure 3b shows the mean judged depth, along the gravitational axis. Consider the remote image first, where target depth was displayed as an image accompanied by a metric. The $d_{mediated}$ was significantly underestimated and best fit by $d_{mediated} = 0.91 \times D_{imaged\_depth}$ ($r^2 = 0.989$, $P < 0.001$). In contrast, subjects performed the pointing task accurately when using the control 3D display, showing no systematic error due to the pointing responses per se. With the control device the judged target depths (solid crosses) had an average slope of 1.00 relative to the physical values, and none of the mean judged values differed significantly from the physical target depth ($t(11) = 1.174$, $P = 0.265$; $t(11) = 1.352$, $P = 0.204$; $t(11) = 1.024$, $P = 0.328$ for the target depth of 3.5, 5.0, and 6.5 cm, respectively). This indicates that observed errors in $d_{mediated}$ for the remote display were not due simply to motor control or response bias.

Experiment 3. Target localization by uniting two depth components

Experiment 3 asked whether the conjoining of $d_{perceptual}$ and $d_{mediated}$ added any additional systematic error over and above that observed by simply summing these two components. The same pointing task as used in Experiment 2 was implemented to measure the perceived overall depth ($d_{overall}$), but now with an indentation to be considered ($d_{perceptual}$) as well as probe-relative depth ($d_{mediated}$). As discussed in the Introduction section, $d_{overall}$ is theoretically given by

$$d_{overall} = d_{perceptual} + d_{mediated} + err_{integration} \quad (1)$$

Substituting the parameters of the psychophysical functions from Experiments 1 and 2,

$$\begin{aligned} d_{overall} = {} & 0.61 \times D_{physical\_indent} + 0.031 \times \text{Force} \\ & + 0.91 \times D_{imaged\_depth} + err_{integration} \end{aligned} \quad (2)$$

The unknown is $err_{integration}$, the error (constant or proportional) that might be brought about by the process of integration itself. To evaluate it, our approach here was to assess $d_{overall}$ and to compare it with the sum of $d_{perceptual}$ and $d_{mediated}$.

Method

Participants

Twenty-four subjects with normal or corrected-to-normal vision and stereo acuity were tested.

Stimuli and procedure

Similarly to Experiment 1, two sub-designs were conducted, with 12 subjects each, for different combinations of stiffness and penetration. The experimental settings were: (a) three penetration depths (0.0, 0.7, and 1.4 cm) at the minimum stiffness level (0.74 ± 0.06 N/cm) and (b) three stiffness levels (0.81 ± 0.16, 2.41 ± 0.25, and 6.38 ± 0.54 N/cm) at the penetration depth of 1.6 cm. These will be referred to as Experiment 3a and 3b, respectively (see previous figures). Each combination was tested twice, requiring a total of 12 tanks. The structure of these tanks and the placement of targets inside were the same as described before. The overall target depths were 3.5, 5.0, and 6.5 cm relative to the lid of the tank.

The same ultrasound machine was used to localize targets. Subjects were tested binocularly. On each trial, the subject was told to place the ultrasound probe over a target and if necessary, to press the probe so that it "bottomed out" on the rigid under-surface. He or she was clearly instructed that both depth components had to be considered in order to locate the target. A standard 1 × 1 cm grid was provided for subjects' reference to the remote display for the probe-relative depth component. Their judgments of target locations were measured using the same pointing paradigm as in Experiment 2. Before the pointing trials (to

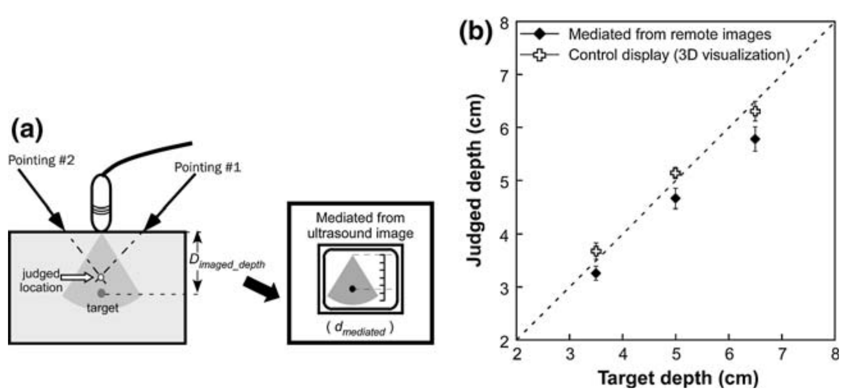

Fig. 3 **a** The experimental task for $d_{mediated}$ judgments. **b** Judged depth in Experiment 2 as a function of the physical target depth using the remote image and control display (*filled diamonds* and *open crosses*, respectively)

avoid bias from the experimental judgments), their perception of indentation, $d_{perceptual}$, was assessed using the same line-drawing paradigm as in Experiment 1. (Results were shown in Fig. 2 to fall on the same regression line fit to that experiment.) The test order of tanks and targets was counterbalanced across subjects. Subjects received no error feedback.

Results

As shown in Figs. 4b and c, significant effects of physical indentation and surface stiffness were found on judgments of $d_{overall}$. On the whole, $d_{overall}$ was systematically underestimated. Underestimation increased with physical indentation ($F(2,22) = 6.150$, $P < 0.01$) and was counteracted by surface stiffness ($F(2,22) = 22.051$, $P < 0.001$). These trends had been found for the estimate of $d_{perceptual}$ in Experiment 1, and were found here with the line-drawing task, which was a limited replication of that experiment. That is, $d_{perceptual}$ (insets, Fig. 4) was under-perceived, more so with greater indentation ($F(1,11) = 17.343$, $P < 0.01$) and less so with greater stiffness ($F(2,22) = 26.828$, $P < 0.001$). However, it should be noted that $d_{overall}$ was underestimated by more than the error in $d_{perceptual}$, suggesting further errors from judgment of the imaged depth ($d_{mediated}$) and possibly other sources (i.e., $err_{integration}$).

The next analysis was conducted to test the model of Eq. 2 with the assumption that the error term, $err_{integration}$, is equal to zero. The judged depth from Experiment 3 ($d_{overall}$) was compared to the sum of the values from Experiments 1 ($d_{perceptual}$) and 2 ($d_{mediated}$), with the discrepancy being an estimate of $err_{integration}$. As shown in Fig. 5a, the estimated $err_{integration}$ was close to zero across the range of depths under study. Statistically, there was no evidence of a systematic $err_{integration}$ [$t(8) = 0.516$, $P = 0.620$ and $t(11) = 0.691$, $P = 0.504$ for the comparison of Experiments 3a and 3b, respectively, to the sum of Experiments 1 and 2]. Overall, then, the depth-concatenation process can be modeled as in Fig. 5b, where it is a simple summation of the two component depths.

Experiment 4. Judgments of numerically-specified locations

Experiments 1–3 demonstrate a process that integrates contiguous segments by concatenating them, without introducing systematic error. We hypothesize that the integration process is intrinsically spatial, because its output is sufficient to guide an action response, pointing.

Alternatively, however, it might be argued that the integration process is a purely numerical calculation. That is, $d_{overall}$ might conceivably be calculated by converting two judged depths into arbitrary unit values and then adding them together. Assuming unbiased (if not fully accurate) addition, as seems highly reasonable, this process would add no further systematic error. In order to guide action, its output would then have to be translated into a

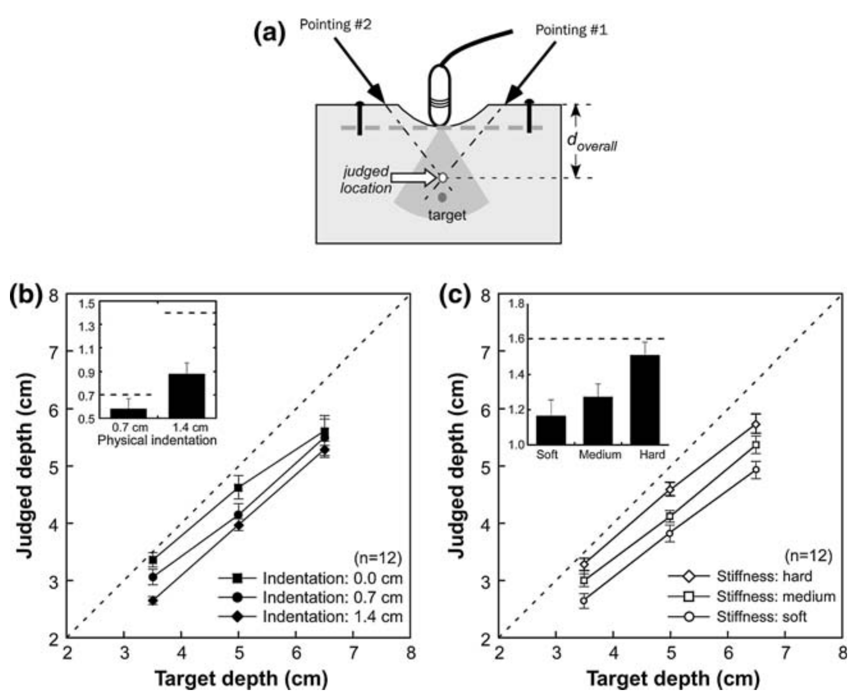

Fig. 4 **a** The stimulus and task used for $d_{overall}$ judgments; **b** judged $d_{overall}$ at the lowest stiffness for each physical indentation; **c** judged $d_{overall}$ as a function of surface stiffness; the *inset* figures show the mean judged $d_{perceptual}$. Error bars represent the standard error of the mean

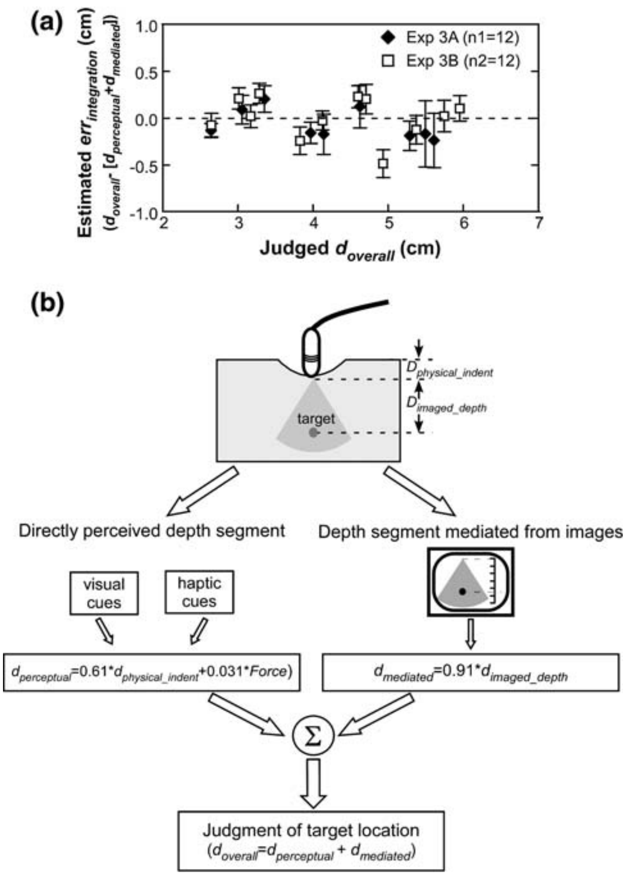

Fig. 5 **a** Estimated $\mathrm{err}_{\mathrm{integration}}$ as a function of the observed d_{overall}. Error bars represent the standard error of the mean. **b** A schematic of the model for judging d_{overall}

spatial representation. Note, then, that this numerical-addition version of integration must specify two processes that follow the encoding of the individual depth segments. One is numerical addition; the second is conversion of the sum to a spatial representation of depth. Because results of Experiment 3 indicate that no further error is added, once each depth segment is determined, up to the point of the response, the hypothesis requires that both of these processes must be error-free. It seems reasonable to presume that addition of two numerically represented depths could be performed without error, but it is less clear that the conversion from numerical magnitudes to action-guiding spatial representations could also be error-free.

To assess this, the experiment measured subjects' accuracy in pointing to locations that were specified numerically, by a verbal metric. If no errors were made in this task, it would be consistent with the idea that the process of concatenating perceived and mediated depth is numerical. However, if errors were made, it would indicate that the numerical output is not sufficient to guide action and hence the underlying integration process must provide a spatial representation as its output.

Method

Participants

Thirteen subjects were tested. All had normal or corrected-to-normal vision and stereo acuity better than 40″ of arc.

Stimuli and Procedure

Subjects' judgments were measured at five depths (2.5, 3.5, 5.0, 6.5 and 7.5 cm). A plastic tank served as the stimulus, on the lid of which eight locations were marked with letters A to H. On each trial, a location was announced by the experimenter, using a description like "3.5 cm below Position A". A standard 1 × 1 cm grid was provided for reference. The subjects were tested binocularly using the same procedure as in Experiment 3. Each target depth was measured with three repetitions at random locations. Additionally, a baseline measure was performed using two other tanks that had five beads (1.0 cm in diameter) mounted at the experimental depths and a cut-away lid that enabled subjects to view the beads directly. This was conducted last to avoid bias. No feedback was provided during the entire experiment.

Results

The judged target locations were computed from subjects' pointing responses using the aforementioned algorithm. Figure 6 shows the mean judged depths along the gravitational axis along with the regression lines. Consider first the baseline condition, where the targets were viewable. The subjects' performance (solid square and the dashed lines) was relatively accurate with a slope relating actual and estimated depths near unity (slope = 1.026, r^2 = 0.963, $P < 0.001$). Performance was similar to that obtained with the control device (open crosses) in Experiment 2. One exception was the judgment of the shallowest depth, 2.5 cm, which was significantly overestimated ($t(12)$ = 4.223, $P < 0.01$). This may be attributed to a bias in the pointing response such that people tend to overshoot small angles (the vertical angle required to aim at the 2.5 cm-deep target from different positions ranged from 13.5° to 24.5° in this experiment).

Compared to their baseline performance, the subjects significantly underestimated depths in the numerical-description condition ($F(1,12)$ = 60.572, $P < 0.001$). The data points (open circles) were best fit by a line with slope of 0.882 (r^2 = 0.975, $P < 0.001$). For comparison, Fig. 6 (filled diamonds) also shows the data from Experiment 2,

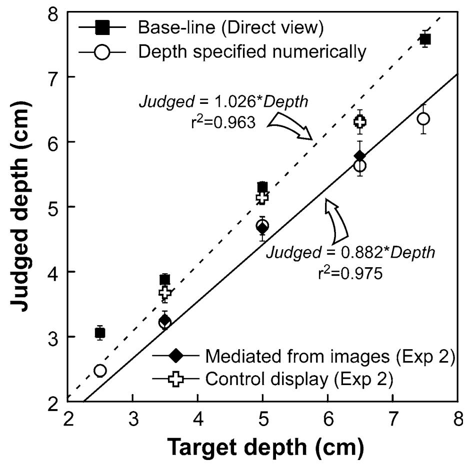

Fig. 6 The *open circles* represent the mean judgments of depths specified numerically in spatial language, while the *filled squares* denote the judgments with direct vision. The *solid* and *dashed lines* respectively represent the linear regressions of the data sets on physical depths. For comparison, data from Experiment 2 were also plotted where the targets were localized using ultrasound images (*filled diamond* for the mediated task and *open crosses* for the control display). Error bars represent the standard error of the mean

where depth information was graphically portrayed. No significant effect of depiction type (graphical vs. numerical) was found ($F(2,46) = 0.384$, $P > 0.5$) across the three common depths tested (3.5, 5.0, 6.5 cm). This suggests a common mechanism for processing the depth information conveyed via different inputs: using the metric scale on the side of the ultrasound image (Experiment 2) and using spoken numbers (this experiment).

Importantly, the above results negate the hypothesis that the error-free concatenation observed in Experiment 3 corresponds to adding two numbers. Experiment 4 was intended to bypass the integration process and give participants a numerical specification directly. If the numerical specification of depth actually resembled the output of integration in prior studies, then based on those studies' showing that all systematic errors arise prior to integration, participants in Experiment 4 should have proceeded to respond without error. However, this prediction was disconfirmed by the data, which show that some process following the numerical specification of depth in Experiment 4 must have induced error. Presumably, that process is the translation of the number into a spatial representation, which is essential for action. Therefore, we reject the argument that integration is a process of mental addition producing a numerical output. Our model assumes that in Experiments 1 –3, perceptual and mediational processes produce spatial representations prior to integration. It was those processes that induced error; the integration was itself error-free and led directly to action.

General discussion

Systematic error and noise in the process of depth concatenation

Our experiments clearly show that two contiguous depth segments, one perceptually encoded and the other cognitively mediated, can be mentally integrated. Further, Experiment 4 indicates that the process of integration operates on the basis of intrinsically spatial representations, which are achieved from both perceptual and cognitive pathways. We take these results to be consistent with the hypothesis that spatial information from different modalities can converge on a unitary, and possibly amodal, representation (Bryant 1992; Loomis et al. 2002).

One finding, using the error pass-through approach, was that the integrator in our studies added no systematic error to the process of arriving at a depth representation. Each channel was found to be subject to error; however, as shown in Fig. 5, the magnitude of underestimation error observed in the final judgments matched with the sum of errors arising from each channel in isolation. Hence, mean judgments of overall depth were predicted well by the sum of psychophysical functions for depth components that were estimated for each information source.

Additionally, one can use the error pass-through method to ask whether the process of concatenation added random variability (noise) to the final output. In the present study, the target depth was parsed into two segments, which were encoded respectively through perception or mediation and processed separately by distinct neural mechanisms. Given process independence, variability that arises at judgments of each depth components would propagate through the integrator and add in the final estimations. Assuming additional variance would be introduced while assembling two segments, we would have

$$\mathrm{var}(d_{\mathrm{overall}}) = \mathrm{var}(d_{\mathrm{perceptual}}) + \mathrm{var}(d_{\mathrm{mediated}}) + \mathrm{var}_{\mathrm{integration}}. \quad (3)$$

To test this, we took the difference between each individual's judgments in the two replications as an estimate of the within-subject variability. The values of $\mathrm{var}(d_{\mathrm{perceptual}})$, $\mathrm{var}(d_{\mathrm{mediated}})$ and $\mathrm{var}(d_{\mathrm{overall}})$ were estimated from the data of Experiments 1, 2 and 3, respectively. In order to evaluate $\mathrm{var}_{\mathrm{integration}}$, $\mathrm{var}(d_{\mathrm{overall}})$ was compared with the sum of $\mathrm{var}(d_{\mathrm{perceptual}})$ and $\mathrm{var}(d_{\mathrm{mediated}})$. No statistically significant difference in variability was found ($t(14) = 1.305$, $P > 0.2$), consistent with the pattern observed for the mean judgments. This result offers no evidence that the concatenation process adds variability; however, this conclusion is made tentatively as the design was not directed toward addressing this issue (cf. Ernst and Banks 2002).

A question arises as to how general error-free integration might be. Our experiments examined the assembling of two depths where all depth intervals were in near (reachable) space. There is evidence for amodal spatial representations at further distances reached through locomotion (e.g., 20 m or so), as cited above. However, we do not know if there is error-free integration at those distances. With greater depth values or more complex tasks involving multiple depths and variable orientations, new sources of error might come into play. These would presumably be passed on to the process of integration. Whether the process itself might come to add error under these circumstances remains to be investigated.

Haptic modulation of visually perceived depth

The results of Experiment 1 offer a novel illusion resulting from visual-haptic interaction. Specifically, judgments of depth based on visible compression of a surface are inflated by resisting forces. When the compression bottoms out, the perceived depth of penetration can be predicted by a weighted sum of the physical indentation and a distortion based on the terminal resistance. This presumably arises because many compliant surfaces in everyday life behave much like springs, where the relation between resisting force and indentation is a linear one. Following a model of Proffitt and colleagues (Proffitt 2006; Proffitt et al. 2003), the influence of resistance on perceived indentation might be mediated by a sense of energy expenditure.

A converse effect has previously been found, whereby a visual depiction of compression affects the perceived stiffness of springs (Srinivasan et al. 1996). These observations support the notion that the judgment of 3D geometric properties is determined by the integration of visual and haptic cues, perhaps in a statistically optimal way (Drewing and Ernst 2006; Ernst and Banks 2002; Ernst et al. 2000). Linear regression indicated that visual cues contributed much more than force in perceiving surface indentation, when the coefficients were standardized by variance. This might be related to the relative reliability of different cues: the surface deformation was reliably cued by rich visual cues such as binocular disparity (Devisme et al. 2005); in contrast, the relation between deformation and exerted force can only be used as a weak heuristic.

It is worth noting that modulation of visual perception by haptically perceived force is of interest beyond adding to the literature on illusions. This interaction has potential impact on the use of ultrasound in medical application. To the extent that errors in cognitively mediated depth and distortions from resisting forces lead to errors in the localization of ultrasound targets, they would have consequences for invasive procedures guided by ultrasound images. Previously (Wu et al. 2005), we showed that systematic errors in perception of target locations translate directly into errors in guiding action. A greater understanding of how such errors arise would help in designing methods to compensate for them.

Acknowledgments This work is supported by grants from NIH (R01-EB00860) and NSF (0308096). Parts of this work were presented at the 2005 Annual Meeting of the Vision Sciences Society.

References

Avraamides M, Loomis J, Klatzky RL, Golledge RG (2004) Functional equivalence of spatial representations derived from vision and language: evidence from allocentric judgments. J Exp Psychol: Hum Learn Mem Cogn 30:801–814

Battaglia FP, Sutherland GR, McNaughton BL (2004) Local sensory cues and place cell directionality: additional evidence of prospective coding in the hippocampus. J Neurosci 24:4541–4550

Blaisdell AP, Cook RG (2005) Integration of spatial maps in pigeons. Anim Cogn 8:7–16

Bryant DJ (1992) A spatial representation system in humans. Psychology 3(16):1

Cutting JE, Vishton PM (1995) Perceiving layout and knowing distance: the integration, relative potency, and contextual use of different information about depth. In: Epstein W, Rogers S (eds) Handbook of perception and cognition: perception of space and motion vol. 5. Academic, San Diego, pp 69–117

Devisme C, Monot A, Drobe B, Droulez J (2005) Influence of visual context on surface deformation perception based on binocular disparity. Perception 34(Suppl):114

Drewing K, Ernst MO (2006) Integration of force and position cues for shape perception through active touch. Brain Res 1078(1):92–100

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433

Ernst MO, Banks MS, Bülthoff HH (2000) Touch can change visual slant perception. Nature Neurosci 3:69–73

Gibson JJ (1979) The ecological approach to visual perception. Houghton Mifflin, Boston

Golledge RG, Smith TR, Pellegrino JW, Doherty S, Marshall SP (1985) A conceptual model and empirical analysis of children's acquisition of spatial knowledge. J Environ Psychol 5:125–152

Holding CS, Holding DH (1988) Acquisition of route network knowledge by males and females. J Gen Psychol 116:29–41

Irwin DE, Andrews RV (1996) Information integration in perception and communication. In: Inui T, McClelland J (eds) Attention and performance (16). MIT Press, Cambridge, pp 125–155

Keyson DV (2000) Estimation of virtually perceived length. Presence 9(4):394–398

Klatzky RL, Lippa Y, Loomis JM, Golledge RG (2003) Encoding, learning and spatial updating of multiple object locations specified by 3-D sound, spatial language, and vision. Exp Brain Res 149:48–61

Lanca M, Bryant D (1995) Effect of orientation in haptic reproduction of line length. Percep Mot Skills 80:1291–1298

Lederman SJ, Thorne G, Jones B (1986) Perception of texture by vision and touch: multidimensionality and intersensory integration. J Exp Psychol Hum Percep Perform 12(2):169–180

Loomis JM, Lippa Y, Klatzky RL, Golledge RG (2002) Spatial updating of locations specified by 3-D sound and spatial language. J Exp Psychol Hum Learn Mem Cogn 28:335–345

Moar I, Carleton LR (1982) Memory for routes. Q J Exp Psychol A 34:381–394
O'Keefe J, Burgess N (1996) Geometric determinants of the place cells of hippocampal neurons. Nature 381:425–428
Proffitt DR, Stefanucci JK, Banton T, Epstein W (2003) The role of effort in distance. Perception. Psychol Sci 14:106–113
Proffitt DR (2006) Embodied perception and the economy of action. Persp Psychol Sci 1:110–122
Richardson AE, Montello D, Hegarty M (1999) Spatial knowledge acquisition from maps, and from navigation in real and virtual environments. Mem Cogn 27:741–750
Robles-De-La-Torre G, Hayward V (2001) Force can overcome object geometry in the perception of shape through active touch. Nature 412:445–448
Seizova-Cajic T (1998) Size perception by vision and kinesthesia. Percep Psychophys 60(4):705–718
Srinivasan MA, Beauregard GL, Brock DO (1996) The impact of visual information on the haptic perception of stiffness in virtual environments. Proceedings of the ASME dynamic systems and control division (DSC-vol 58), pp 555–559
Stanley G (1966) Haptic and kinesthetic estimates of length. Psychon Sci 5:377–378
Stetten G, Chib V (2001) Overlaying Ultrasound Images on Direct Vision. J Ultrasound Med 20(3):235–240
Sturz BR, Bodily KD, Katz JS (2006) Evidence against integration of spatial maps in humans. Anim Cogn 9:207–217
Teghtsoonian M, Teghtsoonian R (1965) Seen and felt length. Psychon Sci 3:465–466
Teghtsoonian M, Teghtsoonian R (1970) Two varieties of perceived length. Percep Psychophys 8:389–392
Wu B, Klatzky RL, Shelton D, Stetten G (2005) Psychophysical evaluation of in-situ ultrasound visualization. IEEE Trans Vis Comput Graph 11(6):684–693
Wu B, Ooi TL, He ZJ (2004) Perceiving distance accurately by a directional process of integrating ground information. Nature 428:73–77
Wydoodt P, Gentaz E, Streri A (2006) Role of force cues in the haptic estimations of a virtual length. Exp Brain Res 171:481–489