

Optimal integration of shape information from vision and touch

Hannah B. Helbig · Marc O. Ernst

Received: 29 September 2006 / Accepted: 21 November 2006 / Published online: 16 January 2007
© Springer-Verlag 2007

Abstract Many tasks can be carried out by using several sources of information. For example, an object's size and shape can be judged based on visual as well as haptic cues. It has been shown recently that human observers integrate visual and haptic size information in a statistically optimal fashion, in the sense that the integrated estimate is most reliable (Ernst and Banks in *Nature* 415:429–433, 2002). In the present study, we tested whether this holds also for visual and haptic shape information. In previous studies virtual stimuli were used to test for optimality in integration. Virtual displays may, however, contain additional inappropriate cues that provide conflicting information and thus affect cue integration. Therefore, we studied optimal integration using real objects. Furthermore, we presented visual information via mirrors to create a spatial separation between visual and haptic cues while observers saw their hand touching the object and thus, knew that they were seeing and feeling the same object. Does this knowledge promote integration even though signals are spatially discrepant which has been shown to lead to a breakdown of integration (Gepshtein et al. in *J Vis* 5:1013–1023, 2005)? Consistent with the model predictions, observers weighted visual and haptic cues to shape according to their reliability: progressively more weight was given to haptics when visual information became less reliable. Moreover, the integrated visual–haptic estimate was more reliable than either unimodal estimate. These findings suggest that observers

integrate visual and haptic shape information of real 3D objects. Thereby, knowledge that multisensory signals arise from the same object seems to promote integration.

Keywords Multisensory integration · Shape information · Optimal perception · Vision · Touch

Introduction

We perceive our environment through a variety of senses. For instance, vision and touch can both provide information about an object's shape and size. The brain integrates this redundant information to come up with the most reliable (unbiased) estimate. That is, the nervous system integrates noisy sensory information from multiple sensory modalities such that the variance of the final multimodal estimate is maximally reduced. Given that the noise sources associated with the individual estimates are independent and Gaussian, this optimal integration strategy is the maximum likelihood estimate (MLE). The MLE is a linear combination of the individual unimodal estimates—here visual and haptic cues to shape—that are weighted according to their reliability: more reliable cues are assigned a larger weight [for an introduction on optimal linear cue combination see Blake et al. (1993), Ernst and Bühlhoff (2004), Jacobs (1999), Jacobs (2002), Knill and Saunders (2003), Landy et al. (1995), Yuille and Bühlhoff (1996)]. Several researchers have addressed the question whether human observers do indeed use such an optimal cue integration strategy and found that the optimal cue combination rule (MLE rule) predicts

H. B. Helbig (✉) · M. O. Ernst
Max Planck Institute for Biological Cybernetics,
Spemannstr. 38, 72076 Tübingen, Germany
e-mail: helbig@tuebingen.mpg.de

observers' behavior for a variety of sensory modalities and perceptual tasks (e.g., Alais and Burr 2004; Bresciani et al. 2005; Ernst and Banks 2002; Gepshtein and Banks 2003; Hillis et al. 2004; Knill and Saunders 2003; Landy and Kojima 2001). For example, Ernst and Banks (2002) investigated visual and haptic discrimination of object size and found that the performance in the bimodal task was well predicted by the MLE model.

In the present study, we investigated whether this holds also for the integration of visual and haptic shape information. To this end, we presented observers with real objects that they could see and feel. The objects were planar elliptical shapes. The elliptical ridges were raised from front and backside of a planar plastic panel. Subjects could see the ellipse on the frontside while simultaneously touching a spatially corresponding (invisible) ellipse on the backside of the panel. All subjects were informed that they see and feel the top and bottom caps of one cylinder (with elliptical cross-section) to create a strong impression of unity. In reality, seen and felt shape could differ slightly. Observers were asked to judge the shape of the object, i.e., to indicate whether the ellipse appeared horizontally or vertically elongated. They explored the stimulus either visually or haptically alone or by using both sensory modalities simultaneously. To test for optimal integration, we applied an experimental procedure that has become common practice in the recent past (e.g., Alais and Burr 2004; Ernst and Banks 2002; Knill and Saunders 2003). We first determined the reliability of the shape judgments for each modality alone and from these measurements we derived parameter-free predictions for optimal integration using the maximum-likelihood approach. Two predictions can be made: The model predicts, first, the relative visual and haptic cue weights, i.e., the relative contribution of the unimodal shape estimates to the bimodal percept. Secondly, the model predicts the reliability of the bimodal visual–haptic estimate. We then measured bimodal discrimination performance from which we empirically deduced the bimodal reliability and the degree to which vision and touch contribute to the bimodal percept. This was done by introducing small conflicts between the visual and haptic cues to shape. In order to show that cue weighting changes with the reliability of the signals, the reliability of the visual stimulus was manipulated using blurring methods.

Given the paradigm and experimental setup used, there are two aspects that make the problem of integrating visual and haptic cues for shape perception of real three-dimensional (3D) objects particularly interesting for studying optimal integration: first,

previous studies that tested quantitative predictions of the MLE model to determine whether cue integration is statistically optimal used virtual displays to present the test stimuli (e.g., Alais and Burr 2004; Ernst and Banks 2002; Gepshtein and Banks 2003; Hillis et al. 2004; Knill and Saunders 2003; Landy and Kojima 2001). For example, computer displays were used to present disparity or texture cues to slant (e.g., Hillis et al. 2004; Knill and Saunders 2003) in order to test for optimality in integration of depth information. Using such virtual displays was always a point of critique because it may be that some very specific aspects of the perceptual system are investigated using the virtual displays, which differ from the perception in the natural environment. This critique is substantiated by a number of studies, which observed differences in the results when tested with real as opposed to virtual stimuli (e.g., Buckley and Frisby 1993; van Ee et al. 1999). For example, van Ee et al. (1999) examined the slant-contrast illusion to study how different slant estimators are combined. When a frontoparallel test surface is surrounded by a larger slanted surface (an inducer), the frontoparallel surface is perceived as slanted in the direction opposite to the inducer. This slant-contrast illusion occurs when stereo cues alone specify the slant of the inducer and is diminished when consistent non-stereoscopic cues to slant (texture cues) are available. The illusion does, however, not occur when real planes are used as inducer and test stimulus (van Ee et al. 1999). This suggests that the illusion in the “consistent” condition is most likely a consequence of inappropriate cues arising from presentation on a CRT that provides conflicting information (e.g., blur-gradient, accommodation and the phosphor grid of the CRT display provide information that the inducer surface is flat). Similarly, Buckley and Frisby (1993) observed effects of viewing CRT displays versus real objects. Buckley and Frisby presented subjects with vertical parabolic ridges and asked them to judge the depth of the ridges. Both disparity and texture cues to depth were manipulated independently. In this way they could determine the weight given to each individual cue. The experiment was conducted twice; stimuli were either stereograms presented on a CRT display or real 3D ridges (bearing a print-out of a regular texture which could provide conflicting depth information). In the first experiment, the weights given to disparity and texture cues were dependent on the depth of the stimulus (as specified by disparity), whereas in the real-ridge experiment disparity cues dominated the percept (independent of the depth of the stimulus). Again, cue conflicts arising from inap-

appropriate screen cues (blur-gradient, accommodation, phosphor grid of the CRT display) might account for these inconsistent findings with virtual as opposed to real objects (for more details refer to Buckley and Frisby 1993). Likewise, in studies that quantitatively test for optimality in cue integration (e.g., Hillis et al. 2004; Knill and Saunders 2003), there is the potentiality that such conflicting screen cues have affected the variance of the “single-cue” estimates or introduced a bias and thus have affected the model predictions. These examples show that inappropriate cues that might be present in virtual experimental setups can have profound effects on the results of studies on optimal cue integration. Therefore, here we used real objects to study multisensory integration in a natural full-cue environment.

Secondly, previous research has shown that bimodal integration depends on spatial proximity. Gepstein et al (2005) observed that integration breaks down with increasing spatial separation between the signals. In the present study, visual object information was presented via mirrors. In this way, haptic and visual shape information is presented at discrepant locations, while subjects still know that both signals arise from one and the same object. The question is whether this knowledge promotes integration even though there is a spatial discrepancy between the location at which the visual and haptic shape information is provided. Recently it has been shown that multimodal visual and haptic signals have a mutual biasing effect on the shape percept (Helbig and Ernst 2007). This effect was independent of whether visual and haptic signal are collocated or spatially discrepant as long subjects have a strong assumption that they see and feel the same object (because they have sight of their hand touching the object). This finding is a first hint that integration may still occur even when signals are presented at different location. However, Helbig and Ernst (2007) did not quantitatively test whether the predictions of the MLE model still hold (i.e., whether signals are integrated statistically optimal) under such conditions. Therefore, in the present study, we quantitatively tested whether the predictions of the MLE model still hold when multimodal information is provided at two different spatial locations via mirrors while subjects could see their hand exploring the stimulus and thus knew that what they see is what they feel. Optimal integration under such spatially discrepant stimulus presentation conditions would provide strong evidence that multisensory integration is enabled by the subjects’ assumption that visual and haptic information emanates from the same object.

Methods

Participants

Ten right-handed subjects (eight male/two female) with normal or corrected-to-normal vision participated in the experiment for payment. All participants were naïve to the purpose of the experiment. The average age was 24.5 years (range 18–30). Participants gave their informed consent before taking part in the experiment, which was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Stimuli

Stimuli consisted of elliptical ridges with different length-to-width ratios (see Fig. 1a). Two elliptical ridges (thickness 2.0 mm) were mounted to opposite sides of a planar plastic panel of 58.8 mm × 50.0 mm (thickness of the panel was 2.0 mm). Ellipses on both sides of the panel were aligned back to back so as to simulate a composite cylinder protruding through a hole. The stimuli were created using 3D Studio Max 4.3 and were printed in 3D using the Dimension™ 3D-Printer (Stratasys®, Inc.) that builds up 3D objects, layer-by-layer by depositing filaments of a thermoplast (Acrylnitril–Butadien–Styrol, ABS). The printed objects are hard, white, and opaque.

Participants saw the ellipse on the frontside (visual stimulus) and/or palpated the elliptical ridge on the backside of the panel (haptic stimulus) using the index finger of their right hand (see Fig. 1b). They were told that they see and feel the top and bottom caps of one elliptical cylinder that is embedded in the panel. Therefore, it was plausible for the observers to assume that both sensory modalities provide information about the same elliptical shape. In reality, the elliptical shapes on front- and back-side could differ, so that participants could be presented with conflicting visual and haptic shape information. The length of the long axis of the ellipses was always 10 mm. The long axis could be oriented horizontally or vertically. The length of the short axis varied between 6.0 and 9.4 mm. Long and short axis were perpendicular. To describe the elliptical shapes parametrically, we calculate the difference of the vertical (a_{ver}) and the horizontal (a_{hor}) axes. The difference of the elliptical axes ($EA_{\text{diff}} = a_{\text{ver}} - a_{\text{hor}}$) is a measure of the elongation of the ellipse (see Fig. 1c). Ellipses with a negative EA_{diff} are horizontally elongated, while positive values are indicative of vertically elongated ellipses. The higher the absolute value, the more elongated the ellipses are

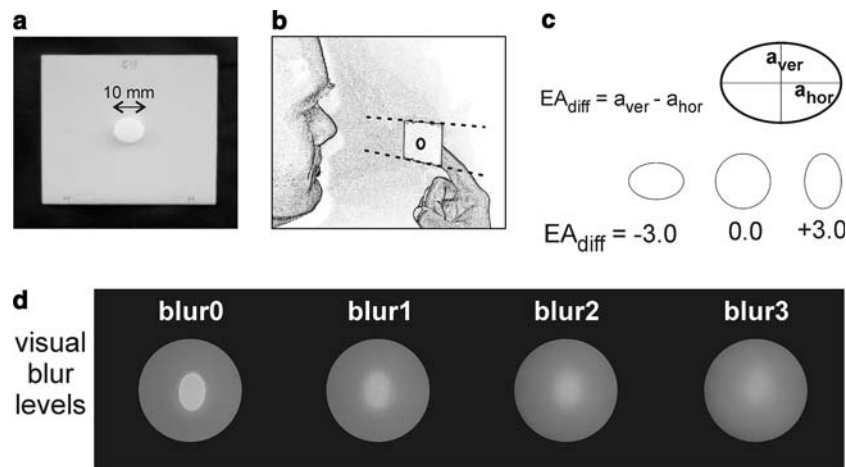


Fig. 1 Stimuli. **a** Example of a visual–haptic stimulus. Stimuli were real 3D objects made of plastic (Acrylnitril–Butadien–Styrol). Stimuli consist of extruded ellipses raised 2 mm from the background. Two ellipses were mounted to front and backside of a planar panel. The ellipses were laid back to back so as simulate one elliptical cylinder embedded in the panel. Front and back ellipse could differ in aspect ratio (EA_{diff}). In this way, we could create small conflicts between vision and touch. **b** Participants saw the ellipse on the frontside of the panel (visual stimulus) and touched the ridge of the ellipse on the back (haptic stimulus). **c** Elliptical shapes were parametrically described by the difference

of the vertical (a_{ver}) and the horizontal (a_{hor}) axes. The longer of the two axes was always 10 mm. The difference of the elliptical axes ($EA_{diff} = a_{ver} - a_{hor}$) is a measure of the ellipses' elongation. The higher the EA_{diff} the more elongated the ellipses are. Ellipses with a negative EA_{diff} are horizontally elongated while positive values are indicative of vertically elongated ellipses. **d** Effect of the blurring lens on the visual stimulus. Photographic images of the stimuli are shown. The amount of blur increases from *blur0* (undegraded vision) to *blur3* (highest level of visual blur)

[$EA_{diff} \in (-10, 10)$]. Note that this parametric description is unambiguous because the long axis of all stimuli was always 10 mm long.

Conditions

Observers explored the stimuli for 5.0 s either visually or haptically alone (V, H) or by using both sensory modalities simultaneously (VH). They were asked to judge whether the ellipse was horizontally or vertically elongated. The experiment comprised a total of 17 conditions: haptic-alone, visual-alone measured repeatedly at four different blur levels, visual–haptic with three different amounts of conflict ($\Delta = \pm 1.0$ or 0.0 mm) between visual and haptic input each measured repeatedly at four different blur levels.

In the unimodal conditions (V, H), the EA_{diff} of the presented stimuli ranged from -3.0 to 3.0 mm. Table 1 shows which stimuli we used. In the bimodal conditions (VH), visually and haptically presented ellipses could be either consistent ($EA_{diff,V} - EA_{diff,H} = \Delta = 0$ mm) or there could be a conflict ($EA_{diff,V} - EA_{diff,H} = \Delta$) between the elliptical shapes on front and backside (that is, vision and touch could provide conflicting shape information). We used small conflicts of $\Delta = -1.0$ mm or $\Delta = +1.0$ mm. That is, in one conflict condition the $EA_{diff,H}$ of the haptic stimulus (back ellipse) always differed by $+1.0$ mm from the $EA_{diff,V}$ of the

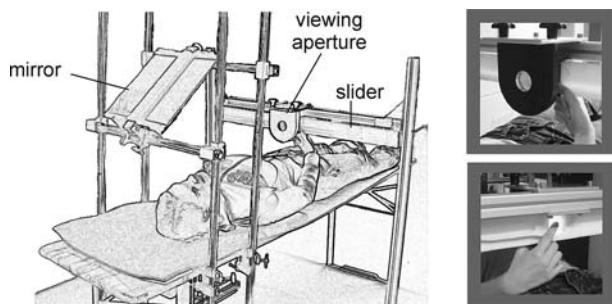
visual ellipse (front ellipse); in the other condition it always differed by -1.0 mm. The visual–haptic stimulus pairs presented in the bimodal conditions with different amounts of conflict are listed in Table 1.

Conflicts ($EA_{diff,V} > EA_{diff,H}$, $EA_{diff,V} < EA_{diff,H}$) were counterbalanced and randomly presented from trial-to-trial during the experiment to prevent visual–haptic adaptation. When debriefing, all participants reported that they were not aware of the conflicts. In the no-conflict case, we presented participants with the same visual and haptic ellipses as in the unimodal conditions (see Table 1).

In order to manipulate the reliability of the visual stimulus we used blur at four different levels, which degraded the visual information. To blur the stimulus subjects looked through a circular viewing aperture (diameter 40.0 mm, see Fig. 2). The aperture was fitted with a transparent plastic cylinder (thickness 12 mm) on which we glued a plastic foil (Leitz®, translucent polypropylen folders, thickness 0.16 mm). The foil had a microbumpy surface scattering light randomly in all directions and thereby inducing blur. The amount of blur was manipulated by varying the distance between the aperture and the visual stimulus (6.0–12.0 cm). It is difficult to parametrically describe the exact physical effects of the blurring foil on the visual shape information. However, note that we were not interested in the physical effects that caused the blur but rather in

Table 1 Visual and haptic stimuli (difference of the elliptical axis EA_{diff}) used in the unimodal visual, unimodal haptic and in the bimodal condition with different amounts of conflict ($\Delta = \pm 1.0$ or 0.0 mm)

Unimodal								
$EA_{diff, V}$	-3.0	-2.0	-1.2	-0.6	0.6	1.2	2.0	3.0
$EA_{diff, H}$	-3.0	-2.0	-1.2	-0.6	0.6	1.2	2.0	3.0
Bimodal								
$\Delta = 0.0$								
$EA_{diff, V}$	-3.0	-2.0	-1.2	-0.6	0.6	1.2	2.0	3.0
$EA_{diff, H}$	-3.0	-2.0	-1.2	-0.6	0.6	1.2	2.0	3.0
$\Delta = -1.0$								
$EA_{diff, V}$	-3.0	-2.0	-1.2	-0.6	0.6	1.2	2.0	3.0
$EA_{diff, H}$	-2.0	-1.0	-0.2	0.4	1.6	2.2	3.0	4.0
$\Delta = +1.0$								
$EA_{diff, V}$	-3.0	-2.0	-1.2	-0.6	0.6	1.2	2.0	3.0
$EA_{diff, H}$	-4.0	-3.0	-2.2	-1.6	-0.4	0.2	1.0	2.0

**Fig. 2** Apparatus. Subjects were lying on their back looking via mirrors through a translucent viewing aperture onto the stimulus. The stimuli are presented using a slider and were manually put in place by the experimenter. Subjects had their hands behind the slider to feel the object from the back while looking at the front

the effect of the blur on the reliability of the perceptually relevant visual shape estimate (JND). This effect was measured psychophysically (see “[Results and discussion](#)”, unimodal visual shape discrimination). Photographs of the visual stimuli at different blur levels are depicted in Fig. 1d. We labeled the different visual blur levels blur0, blur1, blur2 and blur3, where blur0 denotes a condition of non-degraded vision (transparent aperture) and blur1–blur3 denote conditions of increasingly degraded vision. The visual-alone as well as the bimodal visual–haptic conditions were repeatedly measured at four different levels of visual blur. In the haptic-alone condition, we used an opaque aperture to obviate vision.

Apparatus

While performing the task participants lay comfortably on a stretcher on their back and viewed the stimuli through a mirror (see Fig. 2). This was the most com-

fortable position for the subjects to perform the task and to touch the stimuli for a sustained period. The stimuli were presented via a slider with a $40.0 \text{ mm} \times 40.0 \text{ mm}$ window in its center (see Fig. 2). Through the window participants could see and touch one stimulus at a time. Participants touched the stimulus through the window from the backside of the slider (Fig. 2) and they could see it through the window at the front. The experimenter manually presents the stimuli one after another by sliding them by the window. When a new stimulus is inserted into the slider, all stimuli are moved one step further, so that the current stimulus is replaced by the next stimulus. Stimulus timing was controlled by auditory signals that indicated the experimenter when to insert the next stimulus panel into the slider. As mentioned before, the visual blur level was controlled by a translucent aperture set in front of the presentation slider. The distance between viewing aperture and stimulus could be adjusted to modulate the blur level.

Procedure

Stimuli were presented for 5 s (in all conditions). The presentation time of 5 s was chosen to allow subjects sufficient time to haptically extract shape information. Exploration times of 3–5 s yield good haptic performance. Each trial started with an auditory ‘go’ signal instructing the participant to start exploring the stimulus. After the exploration phase of 5 s, an auditory ‘stop’ signal instructed participants to stop exploring the stimulus immediately. After a response period of another 5 s directly following exploration, the next ‘go’ signal indicated the beginning of the next trial. During the response period, participants decide whether the ellipse appeared horizontally or vertically elongated. They provided their answer by pressing one of two response buttons with their left hand. Subjects kept their eyes closed during the response period and the experimenter replaced the current stimulus by the next one.

Trials were blocked by condition (V, H and VH). Stimuli were presented in random order within one block. Bimodal conflict and non-conflict trials randomly occurred within the same block. Each block consisted of either 32 trials (unimodal conditions) or 48 trials (bimodal conditions, 16 trials of each of the conflict conditions $\Delta = +1.0$ mm, $\Delta = -1.0$ mm and $\Delta = 0.0$ mm). Several blocks of each condition were measured repeatedly (unimodal: 6 repetitions, bimodal: 12 repetitions) in random order such that subjects performed a total of 192 trials in each of the 17 conditions (8 different stimuli each presented 24 times), i.e., a total of 3,264 trials per subject.

Data analysis

The proportion of trials in which the ellipse is perceived as being vertically elongated is plotted as a function of the difference of the elliptical axes ($EA_{\text{diff}} = a_{\text{ver}} - a_{\text{hor}}$) (see Fig. 3). To obtain a psychometric function the data were fitted with cumulative Gaussians free to vary in position (PSE) and slope (JND) using the software package *psignifit* (see <http://www.bootstrap-software.org/psignifit/>; Wichmann and Hill 2001). From the fitted psychometric functions, we determined the point of subjective equality (PSE) and the difference threshold [just-noticeable difference (JND)]. The PSE corresponds to the EA_{diff} at which the psychometric function reaches 0.5 (see Fig. 3). That is, the PSE is the point at which the ellipse is equally often judged as being elongated vertically or horizontally. In other words, it corresponds to the stimulus that is perceived as being circular. The JND is defined as the difference between the PSE and the 0.84 point on the psychometric function ($=EA_{\text{diff}}$ of the stimulus that is judged to be vertically elongated 84% of the time) (see Fig. 3). The 0.84 point is usually used to define the JND because then the JND corresponds to the standard deviation σ of the cumulative Gaussian fitted to the data. The JND is the

difference threshold, i.e., it corresponds to the EA_{diff} of an elongated ellipse that can be reliably discriminated from an ellipse perceived as being circular. PSE and JND were determined for each of the 17 conditions separately and for all 10 participants individually.

Results and discussion

The experiment aimed at testing whether visual and haptic shape information is integrated in a statistically optimal fashion. To this end, we determined shape discrimination reliabilities for each modality alone to make predictions for optimal integration behavior. One important prediction of optimal integration is that subjects should give progressively more weight to haptic shape information when vision becomes less reliable. To test this model prediction, we manipulated the reliability of the visual stimulus by blurring the visual stimulus and measure how weighting changes with the reliability of the signals. A further prediction of optimal integration is that the integrated estimate is more reliable than either unimodal estimate. We compared the observed data (bimodal JNDs and relative visual and haptic cue weights) to the predictions of an optimal integrator

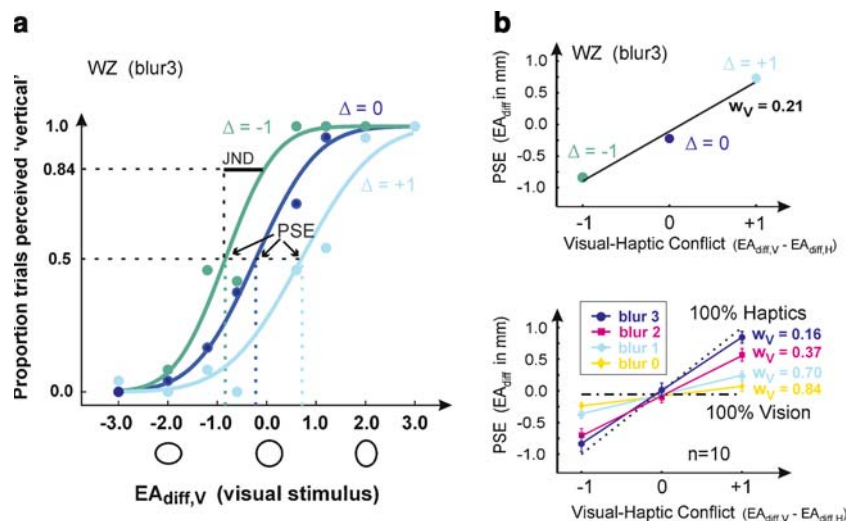


Fig. 3 Illustrates how the points of subjective equality (*PSE*) were used to determine the empirical weights. **a** Psychometric functions for bimodal shape discrimination with different amounts of conflict ($\Delta = \pm 1$ and 0) between visually and haptically specified shape. Data are for one observer (WZ) at noise level blur3. The psychometric functions plot the proportion of trials in which the ellipse is perceived as being vertically elongated as a function of the difference of the elliptical axes ($EA_{\text{diff},V} = a_{\text{ver},V} - a_{\text{hor},V}$) of the visual stimulus (ellipse on the frontside). In conflict conditions, the $EA_{\text{diff},H}$ of the haptic stimulus is either slightly larger ($EA_{\text{diff},V} - EA_{\text{diff},H} = \Delta = -1$) or smaller ($EA_{\text{diff},V} - EA_{\text{diff},H} = \Delta = +1$) than the visual $EA_{\text{diff},V}$, and congruent in non-conflict trials. **b** PSE for bimodal shape

discrimination as a function of visual–haptic conflict. The PSEs are derived from the psychometric functions (illustrated for one observer, WZ, in the left panel). The upper panel shows the data for observer WZ. PSEs are plotted as function of visual–haptic conflict. The lower panel shows average data of ten observers. Error bars represent the standard errors of the mean computed by averaging subjects' individual PSEs. The four colored lines represent the four different blur levels (yellow: blur0, light blue: blur1, magenta: blur2, dark blue: blur3). The continuous line is a linear fit to the data. The slope is a direct measure of the relative visual and haptic cue weights. The relative haptic weight w_H equals 1 if the slope equals 1 ($w_H = 1 - w_V$) and equals 0 if the slope is 0 (indicated with dashed lines in the lower panel)

(MLE model, described in more detail below) to assess whether human observers do indeed integrate visual and haptic shape information in a statistically optimal manner.

Unimodal visual and haptic shape discrimination

Estimates of the variances of the unimodal visual and haptic shape estimates (σ_V^2 , σ_H^2) are obtained by fitting a cumulative Gaussian distribution to the data of the unimodal visual and haptic shape estimate, respectively (see Fig. 3). The JND is derived from the psychometric functions and corresponds to the standard deviation σ . The JND is a measure of the reliability of the shape estimate. The reliability is inversely proportional to the variance:

$$r = \frac{1}{\sigma^2} = \frac{1}{\text{JND}^2} \quad (1)$$

For all unimodal conditions, the PSE was near 0. The dashed line in Fig. 5 indicates the unimodal haptic shape discrimination threshold (JND_H). The solid lines show the unimodal visual shape discrimination thresholds (JND_V) as a function of the visual blur level. Visual JNDs increased with increasing blur level (see Fig. 5). That is, observers' visual shape discrimination ability deteriorates when the visual stimulus is progressively more blurred and thus, visual shape information becomes less reliable. When the visual stimulus is not degraded (blur0), observers can reliably decide whether the visual stimulus is horizontally or vertically elongated when the ellipses differ in their $\text{EA}_{\text{diff},V}$ by $\text{JND}_V = 0.42$ mm from the $\text{EA}_{\text{diff},V}$ of a shape being perceived as circular, whereas at the highest visual blur level (blur3), ellipses need to differ by $\text{JND}_V = 1.76$ mm to be reliably discriminated from a circular stimulus.

Shape discrimination ability in the haptics-alone condition (JND_H) is independent of the visual blur level. Exploring the stimuli using haptics alone, observers could reliably discriminate ellipses from a circular stimulus that differed in their $\text{EA}_{\text{diff},H}$ by $\text{JND}_H = 1.02$ mm. Thus, the haptic JND_H is well within the range of the observed visual JND_V , so that we should be able to find a mutual influence of touch and vision in the combined conditions. The unimodal JNDs were now used to construct a maximum likelihood Integrator in order to predict the observers' bimodal performance.

Predicted and observed cue weights

In a situation in which there are two cues to shape (visual and haptic), the statistically optimal strategy for cue integration is a weighted average

$$\hat{S}_{VH} = w_V \hat{S}_V + w_H \hat{S}_H \quad (2)$$

where the optimal weights are:

$$w_V = \frac{\frac{1}{\sigma_V^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}} = \frac{\sigma_H^2}{\sigma_H^2 + \sigma_V^2}, \quad (3)$$

and likewise for w_H . That is, the optimal relative visual and haptic weights (w_V , w_H) are given by the inverse variance, normalized to sum up to one (e.g., Clark and Yuille 1990; Yuille and Bülthoff 1996):

$$w_V + w_H = 1. \quad (4)$$

Thus, higher weight is attributed to the more reliable of the two cues. This predicts that subjects should give progressively more weight to haptic shape information as the visual blur level increases. To assess whether participants integrate visual and haptic shape information in a statistically optimal way, we first compared the observed relative visual weight at different visual blur levels to the optimal weights predicted by the MLE rule (Eqs. 3 and 4).

The relative visual and haptic weights (w_V , w_H) can be obtained empirically by introducing conflicts between visually and haptically specified cues to shape. The integrated shape percept, i.e., the PSE_{VH} should lie in between the shapes specified by vision (S_V) and touch (S_H). The relative position of the PSE_{VH} on the shape axis in between S_V and S_H is a measure of the relative visual and haptic cue weights. Figure 3 illustrates how the PSE_{VH} was used to determine the empirical visual and haptic weights. If the PSE_{VH} is consistent with the visually specified circular stimulus ($\text{EA}_{\text{diff},V} = 0$), vision dominates, that is the relative visual weight equals 1.0. If, however, the haptic modality dominates the percept (relative haptic weight equals 1.0), the PSE_{VH} shifts towards the haptically specified circular stimulus $\text{EA}_{\text{diff},H} = 0$ (illustrated in Fig. 3a for one observer WZ at the visual blur level 3). We measured PSEs in the non-conflict and in the two conflict conditions and plotted the PSEs as a function of the visual–haptic conflict (illustrated in Fig. 3b). In order to get a more reliable empirical estimate of the relative visual and haptic cue weights, we fitted a line ($y = w_H \cdot \text{conflict} + \text{bias}$) to this data. The slope of this line corresponds to the haptic weight w_H . The dashed lines in the lower panel of Fig. 3b show the predictions if vision or haptics were to dominate completely. PSEs would lie along a line of slope 1, if haptics dominates completely and along a horizontal line (slope 0) if vision dominates. Cue weights were determined for each subject individually.

The predicted versus observed relative visual weights are depicted in Fig. 4 for the four blur levels. The upper panel shows data for four representative observers. The predictions are shown by grey lines and fall close to the actual data (black line). The lower panel shows average data for all ten subjects. A repeated-measures ANOVA was performed on visual weights with the within-subjects factors prediction (predicted, observed) and blur level (blur0, blur1, blur2, blur3). The analysis revealed a significant main effect of blur level [$F(3,27) = 123.585$, $MSE = 0.14$, $P < 0.001$]. As predicted, the visual weights decrease when visual blur is added. More importantly, the

ANOVA revealed that the observed relative visual weight does not differ significantly from the predicted value [$F(1,9) = 2.169$, $MSE = 0.34$, $P > 0.17$] and there is no significant interaction between the factors prediction and blur level ($P > 0.26$). We conducted post hoc t tests comparing predicted and observed visual weight for each of the four blur levels separately. The α level was adjusted using a Bonferroni correction (so that the α level was set to 0.0125 instead of 0.05). None of the four two-tailed paired-sample t tests did reveal significant differences (blur0: $P = 0.57$, blur1: $P = 0.23$, blur2: $P = 0.29$, blur3: $P = 0.05$). Thus, the weight data are consistent with the MLE predictions, which is a

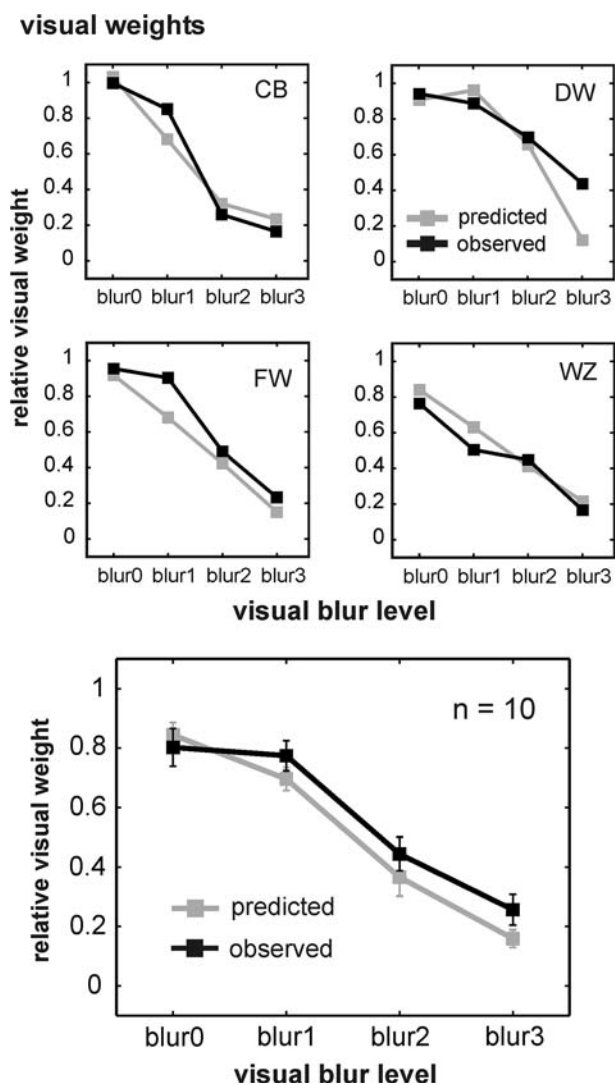


Fig. 4 Predicted and observed relative visual weights. Grey and black lines show the predicted and observed visual weights, respectively. The upper panels show data for four representative observers. The lower panel shows average data for ten observers. Error bars represent standard error of the mean across the ten participants

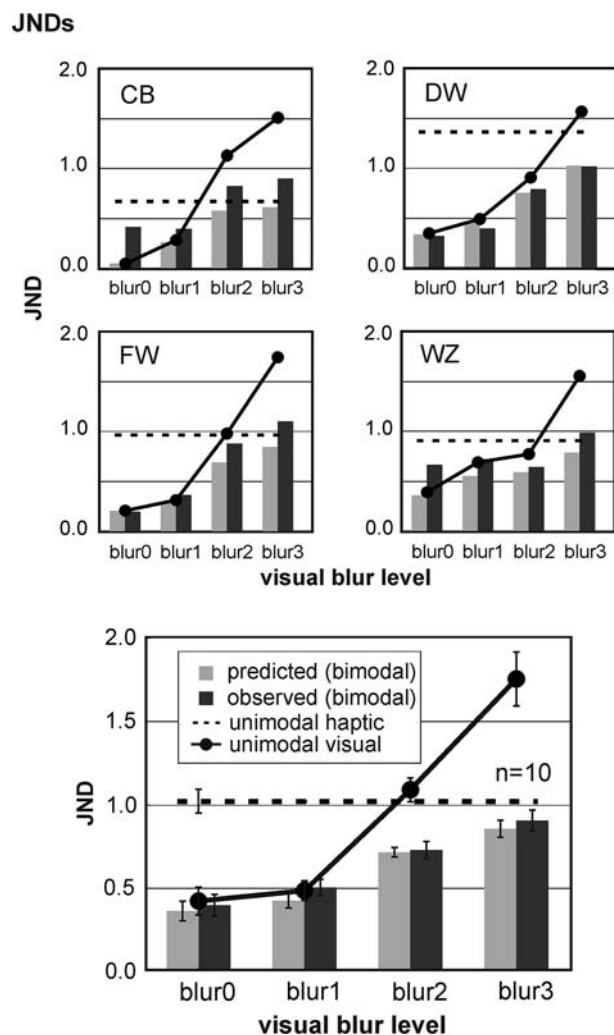


Fig. 5 Predicted and observed just-noticeable differences (JNDs) for four different blur levels. Combined cue JNDs predicted from the single cue JNDs are shown as grey bars. Black bars are the observed bimodal JNDs. Unimodal visual and haptic JNDs are depicted as continuous and dashed lines, respectively. The upper panels show data for four representative observers. The lower panel shows average data for ten observers. Error bars denote standard error of the mean across the ten subjects

first indication that observers combine visual and haptic cues to shape in a statistically optimal fashion. To finally prove that subjects integrate visual and haptic shape information in an optimal way, however, we have to demonstrate a reduction in variance when the visual and haptic cues are presented simultaneously as compared to unimodal presentation. This reduction in variance is the signature of integration.

Bimodal shape discrimination

Combining estimates using the MLE approach, the variance of the statistically optimal bimodal estimate is

$$\frac{1}{\sigma_{\text{VH}}^2} = \frac{1}{\sigma_{\text{V}}^2} + \frac{1}{\sigma_{\text{H}}^2} \quad \text{or} \quad \sigma_{\text{VH}}^2 = \frac{\sigma_{\text{H}}^2 \cdot \sigma_{\text{V}}^2}{\sigma_{\text{H}}^2 + \sigma_{\text{V}}^2}, \quad (5)$$

where σ_{V} , σ_{H} and σ_{VH} are the standard deviations of the visual, haptic and combined shape estimate. The variance of the combined estimate is always less than the variance of either unimodal estimate

$$\sigma_{\text{VH}}^2 \leq \min(\sigma_{\text{V}}^2, \sigma_{\text{H}}^2). \quad (6)$$

In terms of reliability, Eq. 5 can be written as $r_{\text{VH}} = r_{\text{V}} + r_{\text{H}}$. That is, by combining several sources of information the reliability is increased. The solid and dashed lines in Fig. 5 show the unimodal visual and haptic JNDs and the grey and black bars are predicted and observed bimodal JNDs. Again, the upper panels show data for four representative observers. JNDs for combined cue stimuli were on average lower than or equal to the JNDs measured for the individual unimodal stimuli and nicely followed the predictions (Eq. 5). The lower panel in Fig. 5 shows average data for all ten participants. To test whether the integrated estimates are indeed more reliable than either unimodal estimate, we conducted one-tailed paired-sampled t tests comparing unimodal visual and haptic JNDs with bimodal JNDs at each of the four visual blur levels (blur0–blur3). In the following, we used visual–haptic JNDs obtained with non-conflict stimuli only. Note however, that bimodal JNDs do not vary significantly across conflicts ($P > 0.74$). Bimodal JNDs at all visual blur levels were significantly lower than the unimodal haptic JND (blur0: $P < 0.0001$, blur1: $P < 0.0001$, blur2: $P < 0.01$, blur3: $P < 0.043$). Comparing bimodal JNDs with unimodal visual JNDs, we found that the bimodal JNDs were significantly lower than the visual JNDs at high levels of visual blur (blur2: $P < 0.01$, blur3: $P < 0.001$). There was also a reduction in JND in the two low visual blur conditions (blur0 and blur1), but this reduction in variance from vision-alone to

visual–haptic did not reach significance (blur0: $P < 0.396$, blur1: $P < 0.298$). This is most likely due to a lack of statistical power because the predicted improvement in these two cases is very small. The predicted improvement of the bimodal JND relative to both the visual alone and haptic alone JNDs is maximal when the variance of both unimodal signals is the same. In such a case when the variances of both unimodal signals are equal the predicted improvement in discrimination performance relative to both unimodal conditions is a factor of $\sqrt{2}$. However, when the performance in one unimodal condition is clearly better than the performance in the other sensory modality the predicted improvement relative to the better of the two conditions is very little. Given our data, the uncertainty in estimating the thresholds is relatively large compared to the small predicted improvement from JND_{V} to JND_{VH} in the blur0 and blur1 conditions. Therefore, this improvement is not easy to confirm statistically.

Lastly, we directly compared the observed bimodal size discrimination thresholds (JND_{VH}) with the predictions of the optimal integration model as specified in Eq. 5. Figure 5 shows the predicted (grey bars) and observed (black bars) bimodal JNDs. The observed JNDs are in close agreement with the model predictions. A repeated-measures ANOVA was performed on bimodal shape discrimination thresholds (JNDs) with the within-subjects factors prediction (predicted, observed) and blur level (blur0, blur1, blur2, blur3). The ANOVA revealed a significant main effect of the visual blur level [$F(3,27) = 45.410$, $\text{MSE} = 2.356$, $P < 0.001$], indicating a significant increase of the bimodal JNDs with increasing visual blur level. Most importantly, the analysis revealed that the measured combined JNDs do not differ significantly from those predicted from the unimodal JNDs using the MLE model [$F(1,9) = 1.751$, $\text{MSE} = 2.264$, $P > 0.21$; no significant interaction of factors prediction and blur level, $P > 0.827$]. For consistency across observers see Fig. 6 which shows a more detailed representation of the JND data. For all ten participants, we plotted bimodal JNDs as a function of the JND predicted from their unimodal data (Eq. 5). Different blur levels are highlighted in different colors. The empirical JNDs follow the JNDs predicted by an optimal integrator.

Overall, the results show that shape discrimination performance improves when simultaneously seeing and touching the stimuli. This reduction in variance of the shape estimates is the signature of integration. The observed bimodal shape discrimination thresholds (JNDs) are consistent with the model predictions which provide further evidence that the brain makes optimal use of the sources of visual and haptic shape

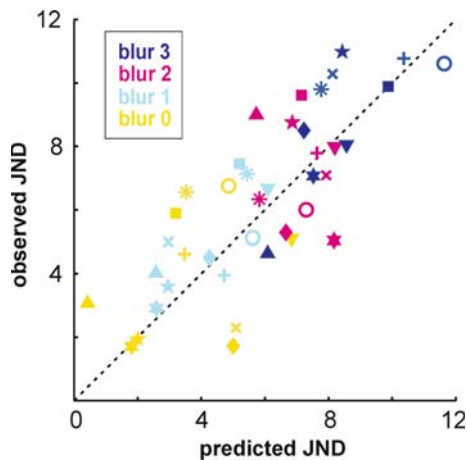


Fig. 6 Scatter plot of measured JNDs as a function of predicted JNDs. Different *symbols* represent data obtained from different observers. *Colors* represent the four different blur levels (yellow: blur0, light blue: blur1, magenta: blur2, dark blue: blur3). The *dashed line* indicates where the data would lie if the predicted and measured JNDs are identical

information. This seems to be true, even though the visual shape information is presented via mirrors and therefore at a spatial location that differed from the spatial location at which the haptic signals were presented. This finding confirms and extends previous research (Helbig and Ernst 2007) which showed that visual and haptic shape signals arising from discrepant locations could introduce a mutual bias in the shape percept when observers had knowledge that the visual and haptic signals belong to the same object, such as here, when the information is presented via a mirror.

General discussion

In the present study, we tested whether redundant shape information from vision and touch is integrated into a unified percept in a statistically optimal fashion. In agreement with this hypothesis, we observed that the bimodal data are well in agreement with the predictions of the optimal integration model (MLE). Visual and haptic shape information is weighted according to its reliability. Most importantly, bimodal shape estimates are more reliable than shape estimates that rely on either vision or touch alone. These results indicate that observers integrate visual and haptic shape information of real 3D objects in a statistically optimal manner and thereby reduce the variance of the integrated visual–haptic shape estimate.

A number of previous studies used real objects (e.g., plastic rectangles, wooden blocks) to investigate integration of visual and haptic shape (or size) information

(e.g., Hershberger and Misceo 1996; Klein 1966; McDonnell and Duffett 1972; Miller 1972; Power and Graham 1976; Rock and Victor 1964). In these studies, conflicts were created between visually and haptically specified shape (or size). This was mostly done by means of a lens that optically distorts the visual image along one axis while the tactual object is unaffected. Some studies (Miller 1972; Power and Graham 1976; Rock and Victor 1964) observed that vision dominates the bimodal shape percept, whereas others observed a considerable contribution of touch to the bimodal percept (Hershberger and Misceo 1996; Klein 1966; McDonnell and Duffett 1972). However, these experiments could not shed light on the mechanisms underlying the differential contribution of vision and touch. In those studies, the reliabilities of the unimodal visual and haptic estimates were not determined and thus it was not possible to elucidate the relations between the reliability of a cue and the relative weight assigned to this cue. In addition, quantitative predictions for optimal cue weighting could not be derived. Therefore we developed an experimental design allowing for quantitative predictions based on presentation of real objects to establish whether visual and haptic shape information is integrated according to an optimal integrator (MLE).

At present, there are several studies demonstrating optimal integration for a variety of object properties. For example, Knill and Saunders (2003) as well as Hillis et al. (2004) investigated intra-modal (visual–visual) integration of disparity and texture cues to surface slant and found that observers use a statistically optimal strategy for combining these signals. Landy and Kojima (2001) found close to optimal integration for the perception of visual texture boundaries. For the perception of visual-auditory location, optimal integration was found by Alais and Burr (2004). More closely related to the present study, Ernst and Banks (2002) as well as Gepshtein and Banks (2003) found optimal integration for the perception of visual and haptic size of an object. In the present study, we extended this list of features (i.e., slant, size, location and texture boundary) to shape information from vision and haptics. This is interesting because shape, as studied here, is a 2D feature (height-to-width ratio of a planar shape that varies in two dimensions), whereas the other features investigated so far (slant, size, location and texture boundary) were 1D only. Hence, we have demonstrated with this study that also more complex features (visual and haptic shape) for which the perceptual system has to use information from more than one spatial dimension are also integrated in a statistically optimal fashion.

While previous research mainly used computer generated (virtual) stimuli to test for optimal integration, we studied integration behavior with natural stimuli (real 3D objects). In this way, we avoided conflicts between the cues of interest and second-order cues that may arise from rendering the stimuli with a computer display. Such inappropriate cues could possibly interfere with the cues of interest. That is, using a more ecologically valid context we here confirmed that human observers adopt statistically optimal integration strategies for the perception of visual–haptic shape information. It has been a matter of debate for a long time whether complex naturalistic stimuli are required to study theories of multisensory integration or whether it is sufficient to use simple minimalistic and easier to be controlled stimuli. The question is whether we can extrapolate the principles of multisensory integration found with simple (computer generated) stimuli to understand more complex naturalistic perceptual situations (for a review see Bertelson and de Gelder 2004; De Gelder and Bertelson 2003). In this study, we took an effort towards more environmentally valid full-cue stimuli while at the same time our stimuli were parametrically controllable. We therefore think that in this context the stimuli used in the present study provide a good compromise in order to bridge this gap between complexity and naturalness.

Even though in this study the stimuli sensed visually and haptically were presented at spatially discrepant locations via a mirror, observers adopted a statistically optimal integration strategy. In contrast, Gepshtein et al (2005) found that integration declines when signals were spatially discrepant. Gepshtein et al (2005) created a spatial discrepancy by laterally shifting the visual stimulus relative to the haptic stimulus (both rendered in a virtual environment) by different amounts. Subjects could not see their hand touching the object. They clearly perceived the discrepancy in location between the visual and haptic stimulus and there was no obvious cause for this spatial discrepancy apparent to the subject. In contrast to Gepshtein et al., we found optimal integration despite the spatial discrepancy between vision and touch which is most likely due to the fact that in our study observers apparently knew that the signals come from the same object because they were familiar with mirrors and therefore knew that the felt object is the one seen in the mirror. This knowledge about object identity was further supported by seeing the lower part of the right hand in the mirror while touching the stimulus. The present finding confirms results of Helbig and Ernst (2007) and may be practically relevant for functional magnetic resonance imaging (fMRI) studies of multisensory

integration that often rely on presenting visual stimuli via mirrors.

One last issue to point out here is the apprehension that was raised with respect to an increase in uncertainty due to additional memory components inherent in 2-interval forced-choice (2-IFC) paradigms. Previous research predominantly used 2-IFC tasks to study optimal cue integration (Alais and Burr 2004; Ernst and Banks 2002; Gepshtein and Banks 2003; Gepshtein et al. 2005; Hillis et al. 2004; Knill and Saunders 2003). 2-IFC tasks involve sequential comparisons and thus, additional memory components. This can lead to increased variance of the perceptual estimate. Moreover, judgments made from memory (e.g., size judgments) can differ systematically from those made perceptually (Kerst and Howard 1978; Moyer et al. 1978). Therefore, both the predictions derived from the single-cue estimates as well as the integration results itself may be affected by memory components when using a 2-IFC task. The task we applied here (judging the elongation of an ellipse) is a 1-IFC task. To perform this task, observers discriminated the elliptical stimulus from an internal circular standard or they compared the two main axes of the elliptical stimulus and judged which one was longer while they are exploring the stimulus. In this way, we avoided additional memory components being involved in the perceptual judgment and it is thus more direct to compare the empirical performance with the model's prediction.

To summarize, the results of the current experiment are consistent with the hypothesis that humans optimally combine visual and haptic shape information of real 3D elliptical objects.

Acknowledgments This work was supported by the Max Planck Society, by the 5th and 6th Framework IST Program of the EU (IST-2001-38040 TOUCH-HapSys & IST-2006-027141 ImmerSense) and the SFB550. We wish to thank the participating volunteers and K. Mayer and A. Reichenbach for help with conducting the experiments.

References

- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257–262
- Bertelson P, de Gelder B (2004) The psychology of multimodal perception. In: Spence C, Driver J (eds) *Crossmodal space and crossmodal attention*. Oxford University Press, Oxford, pp 151–177
- Blake A, Bühlhoff HH, Sheinberg D (1993) Shape from texture: ideal observers and human psychophysics. *Vision Res* 33:1723–1737
- Bresciani JP, Ernst MO, Drewing K, Bouyer G, Maury V, Kheddar A (2005) Feeling what you hear: auditory signals can modulate tactile tap perception. *Exp Brain Res* 162:172–180

- Buckley D, Frisby JP (1993) Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Res* 33:919–933
- Clark JJ, Yuille AL (1990) Data fusion for sensory information processing systems. Kluwer, Dordrecht
- De Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. *Trends Cogn Sci* 7:460–467
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433
- Ernst MO, Bühlhoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8:162–169
- Gepshtein S, Banks MS (2003) Viewing geometry determines how vision and haptics combine in size perception. *Curr Biol* 13:483–488
- Gepshtein S, Burge J, Ernst MO, Banks MS (2005) The combination of vision and touch depends on spatial proximity. *J Vis* 5:1013–1023
- Helbig HB, Ernst MO (2007) Knowledge about a common source can promote visual-haptic integration (submitted)
- Hershberger WA, Misceo GF (1996) Touch dominates haptic estimates of discordant visual-haptic size. *Percept Psychophys* 58:1124–1132
- Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: optimal cue combination. *J Vis* 4:967–992
- Jacobs RA (1999) Optimal integration of texture and motion cues to depth. *Vision Res* 39:3621–3629
- Jacobs RA (2002) What determines visual cue reliability? *Trends Cogn Sci* 6:345–350
- Kerst SM, Howard JH (1978) Memory psychophysics for visual area and length. *Mem Cognit* 6:327–335
- Klein RE (1966) A developmental study of perception under conditions of conflicting sensory cues. *Diss Abstr* 27:2162B–2163B
- Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res* 43:2539–2558
- Landy MS, Kojima H (2001) Ideal cue combination for localizing texture-defined edges. *J Opt Soc Am A Opt Image Sci Vis* 18:2307–2320
- Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res* 35:389–412
- McDonnell PM, Duffett J (1972) Vision and touch: a reconsideration of conflict between the two senses. *Can J Exp Psychol* 26:171–180
- Miller EA (1972) Interaction of vision and touch in conflict and nonconflict form perception tasks. *J Exp Psychol* 96:114–123
- Moyer RS, Bradley DR, Sorensen MH, Whiting C, Mansfield DP (1978) Psychophysical functions for perceived and remembered size. *Science* 200:330–332
- Power RP, Graham A (1976) Dominance of touch by vision: generalization of the hypothesis to a tactually experienced population. *Perception* 5:161–166
- Rock I, Victor J (1964) Vision and touch: an experimentally created conflict between the two senses. *Science* 7:594–596
- van Ee R, Banks MS, Backus BT (1999) An analysis of binocular slant contrast. *Perception* 28:1121–1145
- Wichmann FA, Hill NJ (2001) The psychometric function: I. fitting, sampling, and goodness of fit. *Percept Psychophys* 63:1293–1313
- Yuille AL, Bühlhoff HH (1996) Bayesian theory and psychophysics. In: Knill D, Richards W (eds) *Perception as Bayesian inference*. Cambridge University Press, Cambridge