Exploration mode affects visuohaptic integration of surface orientation

Myrthe A. Plaisier
Faculty of Human Movement Sciences, Research Institute MOVE, VU University, Amsterdam, The Netherlands
Cognitive Neuroscience Department and Cognitive Interaction Technology–Center of Excellence, Bielefeld University, Bielefeld, Germany

Loes C. J. van Dam
Cognitive Neuroscience Department and Cognitive Interaction Technology–Center of Excellence, Bielefeld University, Bielefeld, Germany

Catharina Glowania
Cognitive Neuroscience Department and Cognitive Interaction Technology–Center of Excellence, Bielefeld University, Bielefeld, Germany

Marc O. Ernst
Cognitive Neuroscience Department and Cognitive Interaction Technology–Center of Excellence, Bielefeld University, Bielefeld, Germany

We experience the world mostly in a multisensory fashion using a combination of all of our senses. Depending on the modality we can select different exploration strategies for extracting perceptual information. For instance, using touch we can enclose an object in our hand to explore parts of the object in *parallel*. Alternatively, we can trace the object with a single finger to explore its parts in a *serial* fashion. In this study we investigated whether the exploration mode (parallel vs. serial) affects the way sensory signals are combined. To this end, participants visually and haptically explored surfaces that varied in roll angle and indicated which side of the surface was perceived as higher. In Experiment 1, the exploration mode was the same for both modalities (i.e., both parallel or both serial). In Experiment 2, we introduced a difference in exploration mode between the two modalities (visual exploration was parallel while haptic exploration was serial or vice versa). The results showed that visual and haptic signals were combined in a statistically optimal fashion only when the exploration modes were the same. In case of an asymmetry in the exploration modes across modalities, integration was suboptimal. This indicates that spatial-temporal discrepancies in the acquisition of information in the two senses (i.e., haptic and visual) can lead to the breakdown of sensory integration.

Introduction

There are many properties of the external world that can be perceived through vision as well as through touch (haptic perception). These two senses, however, sample the properties of objects in our environment in fundamentally different ways. Vision passively samples the pattern of light that is reflected by objects, whereas haptic perception depen\ds on physical contact with the object that we want to perceive. The limited surface area of the hand in direct contact with an object implies the need to move the hand across an object in order to build up the haptic percept over time (*serial exploration*). For haptic perception, the exploration strategies have been shown to be very stereotypical for the object property under investigation (Lederman & Klatzky, 1987). For instance, to judge surface texture, lateral motion across a surface is used while for judging hardness we press on the surface. For vision, the need

Citation: Plaisier, M. A., van Dam, L. C. J., Glowania, C., & Ernst, M. O. (2014). Exploration mode affects visuohaptic integration of surface orientation. *Journal of Vision*, *14*(13):22, 1–12. http://www.journalofvision.org/content/14/13/22, doi: 10.1167/14.13.22.

doi: 10.1167/14.13.22 Received May 15, 2014; published November 20, 2014 ISSN 1534-7362

for very distinct exploratory movements does not exist to the same extent. Eye movements are made to center the object of interest on the fovea, where the acuity of vision is highest. However, the dynamics of these eye movements (e.g., the speed of saccades) are generally independent of the object property we are estimating. Moreover, even the stationary eyes always sample a large field of view all at once (approximately 180 deg), albeit with lesser acuity in the periphery. In this way, the visual system receives information from a relatively large area in *parallel*, even from locations we are not directly exploring or looking at. The question arises how the information from the different senses is combined, when they are retrieved in such different ways. In this study, we investigated the effects of using different exploratory strategies across the senses on the integration of visual and haptic information.

When different sensory modalities simultaneously provide an estimate for the same object property, thus providing redundant information, the sensory estimates are often integrated into a single percept. In the framework of maximum likelihood estimation (MLE) we consider the integration process to be *statistically optimal* when the weight that each modality estimate receives is fully determined by its relative precision. Thus, according to MLE, the less precise modality receives less weight. Importantly, this way of combining the signals results in the most precise (i.e., minimal variance) combined estimate possible. This combined estimate will be more precise than either the visual or haptic estimate alone. This type of optimal integration has been shown to apply to, for example, the combination of visual and proprioceptive position information (Reuschel, Drewing, Henriques, Roesler, & Fiehler, 2010; van Beers, Sittig, & van der Gon, 1999), visual and auditive speech perception (Alais & Burr, 2004), visual and haptic size and shape perception (Ernst & Banks, 2002; Gepshtein & Banks, 2003; Helbig & Ernst, 2007b), the numerosity of sequences of events (e.g., Bresciani, Dammeier, & Ernst, 2006; Shams, Kamitani, & Shimojo, 2002) as well as visual-vestibular heading perception (Fetsch, DeAngelis, & Angelaki, 2010).

Integration of the sensory signals is only beneficial if the two estimates are related to the same event or object. This means that the sensory system must somehow determine whether or not the two sensory inputs have a common source. This is often referred to as the correspondence problem of multisensory integration (see e.g., van Dam, Parise, and Ernst, 2014). When the discrepancy between the multisensory signals becomes too large, the brain needs to decide whether they belong to two separate sources, rather than one. If the signals were indeed derived from two different sources it would not be useful to integrate these signals at all. Spatial separation between the sensory signals can be a cue that the two signals do not originate from the same external source (object or event). Consequently, for visuohaptic integration of size it has been shown that integration deteriorates with a spatial separation between the origins of the visual and haptic information (Gepshtein, Burge, Ernst, & Banks, 2005). However, when it is clear that both sensory inputs originate from the same object despite a spatial separation (e.g., by seeing your hand via a mirror), integration of the signals may still be promoted (Helbig & Ernst, 2007a). Furthermore, when the spatial separation is bridged by using a tool, optimal integration also still occurs (Takahashi, Diedrichsen, & Watt, 2009). This indicates that solving the correspondence problem can be rather complex.

In principle, a difference in exploration procedure between the visual and the haptic senses should not lead to violations of the unity assumption, because there is no indication that the visual and haptic estimates originate from different objects. In other words, when viewing and touching an object at the same time, the two sensory signals should be combined regardless of whether the haptic estimate is obtained by enclosing the object in the hand (parallel exploration) or a single finger is used to trace the object (serial exploration). On the other hand, the time course of the build up of the percept most likely differs between the parallel and serial exploration modes. When a single finger is used to serially trace an object, it probably takes longer for the haptic percept to build up than the visual percept for which exploration normally occurs in a more parallel fashion. Thus, despite originating from the same object there could be different temporal constraints between the modalities due to the different exploration modes, possibly causing multisensory integration to break down (e.g., Bresciani et al., 2005; Shams et al., 2002).

In correspondence with the idea that serially acquired estimates might not be integrated, Rosas, Wagemans, Ernst, and Wichmann (2005) reported that there was no optimal integration for visuohaptic perception of surface orientation, when the haptic orientation estimate was obtained in a serial fashion by moving the index finger over a surface. In their study, the visual stimulus was viewed monocularly and consisted of a rendered plane allowing parallel extraction of visual information. Possibly, this means that optimal integration does not occur if there is a serial component involved, making an instantaneous combination of the sensory orientation estimates impossible. It could, however, also mean that visual and haptic information both have to be extracted in the same fashion (i.e., both parallel or both serial) for optimal integration to occur. It should be noted though, that in the study of Rosas et al. (2005), there were no binocular disparity cues to provide a metric visual estimate of the absolute depth of the plane. In other words, the failure to find statistically optimal visuohaptic integration of surface orientation could also have been caused by a mismatch in perceived depth between vision and touch. In any case, the breakdown of optimal visuohaptic integration was not simply due to surface orientation being the property to be estimated. Another study in which the haptic surface orientation estimate was obtained by placing the index finger and thumb simultaneously on the surface (parallel exploration) reported optimal visuohaptic integration (Burge, Girshick, & Banks, 2010). In short, it is still an open question what the prerequisites are for integration of visual and haptic information to be statistically optimal and if the mode of exploration in the separate modalities affects sensory correspondence and thus, sensory integration.

Here we systematically studied the effects of different exploration strategies on the integration of visual and haptic estimates of surface orientation. We chose to use surface orientation as the object property to be judged, because it easily allows for serial and parallel exploration. In the first experiment, visual and haptic explorations were always performed in the same fashion (both serial or both parallel). To limit the possibility for parallel processing of multiple locations for vision, the visual information consisted of relatively small blobs (apertures attached to the surface with fuzzy boundaries such that they provide a window onto the texture-less surface) rendered at the location of the fingertip(s). For serial exploration, participants moved the index finger over the surface and at the same time they would view the surface through a single visual aperture attached to the surface that moved with the finger position. In the case of parallel exploration, two fingers were placed on the surface while visually two apertures were displayed; one around each finger. In a second experiment, we investigated if integration can also occur in case the exploration modes between the senses are different (one parallel and the other serial).

Experiment 1: Same exploration mode

In this experiment we varied the visual and haptic exploration modes to investigate whether visual and haptic estimates of surface orientation are combined in a statistically optimal fashion. More specifically, we compared two exploration modes: SERIAL and PARALLEL. We define the exploration mode to be serial when the surface is sampled by scanning across different spatial locations of the plane over time. In the SERIAL exploration mode, participants moved the finger laterally over a virtual surface in order to make a haptic estimate of the orientation (roll) of the surface. Visual information consisted of an aperture spatially aligned with the (unseen) fingertip. This aperture provided only local visual information about surface orientation. Therefore, for both modalities it was necessary to serially scan the surface in order to obtain a reliable estimate of its orientation. In the PARALLEL exploration session, the information about surface orientation was available at every instance in time by sampling multiple spatial locations simultaneously. For vision this meant that there were two stationary apertures on two spatially distinct parts of the surface. Similarly, for haptic PARALLEL exploration, two fingers were placed statically on the surface. In both cases, surface orientation can be estimated from the two local sources of depth information (stereo and perspective cues in case of vision, and the height difference between the fingers in case of touch). This holds as long as the pitch of the surface is assumed to be zero and the surface is planar. Therefore, participants were informed that the surface was planar and that its orientation would only be varied in terms of roll angle.

Methods

Participants

Ten students of the University of Bielefeld participated in both Experiments 1 and 2 (seven female, mean age 22 ± 2 *SD* years, nine right-handed and one left-handed, according to the Edinburgh handedness questionnaire [Oldfield, 1971]). They received financial compensation (6 euros per hour) for their participation. Informed consent was acquired prior to participation and participants were treated in accordance with the Declaration of Helsinki. All participants had normal or corrected to normal vision and they all had good stereo vision according to the Randot Stereotest (Stereo Optical Co., Inc., Chicago, IL). None of the participants reported any somatosensory deficits.

Setup

Participants were seated with their body midline aligned with the center of a visuohaptic workbench (Figure 1a). On the left and right sides of the workspace a PHANToM force-feedback device (PHANToM premium 1.5, SensAble Technologies, Inc. Woburn, MA) was placed. Participants put the index and middle fingers of their dominant hand in thimble-like holders connected to the PHANToMs. In this way, forces for the index and middle finger could be generated independently by each PHANToM. This enabled the display of a virtual haptic plane that could be explored using touch in both a serial (using only one finger) and

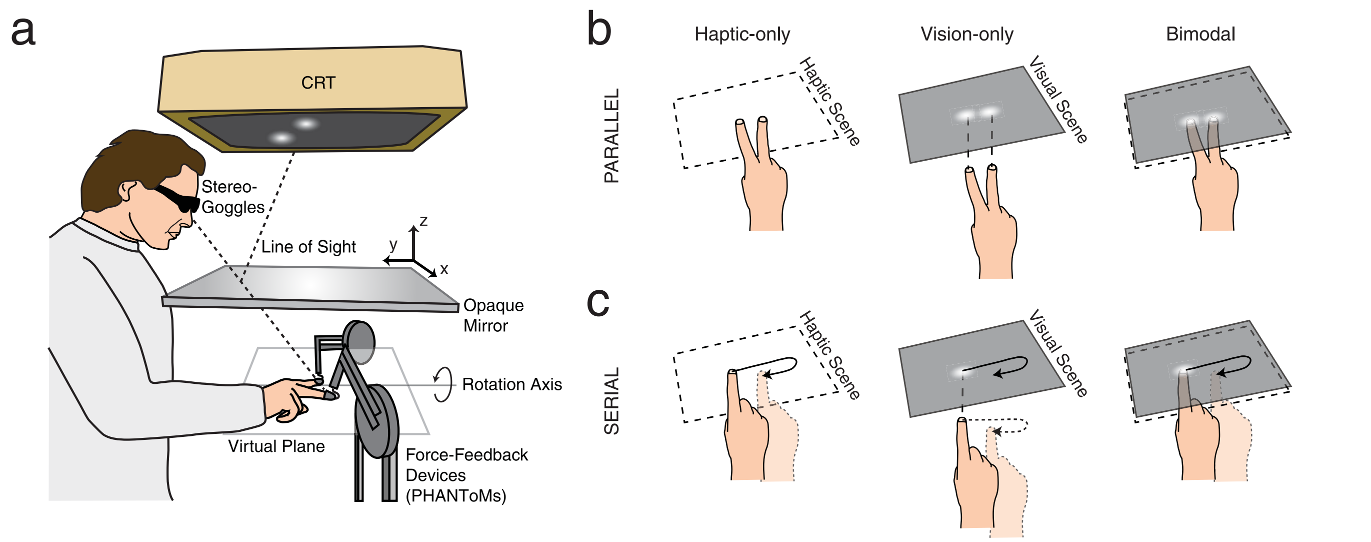

Figure 1. Set-up and stimuli. (a) The visuohaptic workbench with a PHANToM force-feedback device on each side of the workspace for haptic rendering of the plane. Participants viewed the plane through a pair of stereo goggles. (b) In the PARALLEL exploration mode, participants placed the index and middle fingers on the virtual plane and kept static contact. In the haptic-only trials no visual information was displayed. In the vision-only trials two Gaussian aperture blobs were visually rendered, one around each finger, but there was no force feedback. In bimodal trials, both haptic and visual renderings of the plane were switched on. (c) In the SERIAL exploration mode, the index finger was used to trace the surface. In haptic-only trials there was no visual rendering. In the vision-only trials, a Gaussian aperture blob was rendered at the location of the index finger. No force-feedback was provided, so participants moved the finger laterally in midair to move the Gaussian aperture blob across the surface. In bimodal trials haptic and visual renderings were both switched on and participants moved the finger laterally over the surface.

a parallel (using both fingers) fashion. Visual information was presented on a CRT screen (Sony CPD G500/G500J, Sony Europe Limited, Weybridge, UK) mounted upside-down and set to a refresh rate of 140 Hz. The image of the screen was projected via a mirror onto the workspace such that haptic and visual information were spatially aligned. The viewing distance to the visually and haptically rendered surface was 55 cm Shutter glasses (RealD Pro CrystalEyes 4S VRLOGIC Gmbh, Dieburg, Germany) were used to present the left and right eye images with 70 Hz each, alternating between the eyes. This way, perspective and binocular cues were provided for the visual surface. The system delay for the visual representation was determined using the method described in Di Luca (2010). The system delay of the visuohaptic workbench was 34.5 ms (Rohde & Ernst, 2012).

Stimuli and procedure

Visual stimuli consisted of a white plane without texture seen through a circular aperture with fuzzy edges attached to the surface. The background was always black. Thus, the visual information corresponded to circular Gaussian aperture blobs (the standard deviation was 10.6 mm corresponding to ca. 1.1 deg visual angle). These were displayed at the locations corresponding to the finger positions (note that the fingers themselves were not visible to the participant due to the use of the mirror setup). Visual surface orientation was defined by projecting the apertures onto the surface using stereo and perspective (elongation) cues. This gave the impression that the surface was illuminated from above with a flashlight (or two in the case of PARALLEL exploration). The haptic surface was rendered using force-feedback with the setup described in the previous section. Static contact with the surface from a single finger did not provide any information about surface orientation, because the thimble-like holder provides a single contact point only. In the visual case, however, a single static Gaussian aperture blob projected onto the surface does provide some limited information about surface orientation due to stereo and perspective cues from the fuzzy outline of the aperture blob. In order to approximate the information available from the haptic static cue, we minimized these visual orientation cues from one static aperture blob by rendering a surface without texture, keeping the aperture blob size small (roughly the same as the finger size), and blurring the edges of the aperture. Therefore, the properties of the Gaussian aperture blobs were such that the reliability of the visual information was comparable to the reliability of the haptic information, preventing visual capture. This was confirmed during pilot testing. Because a single static contact point (haptic) or a single static aperture blob (vision) contained no or very limited cues as to the orientation of the surface, movements were necessary to estimate surface orientation in both modalities.

In Experiment 1, the exploration mode was always the same for both modalities (i.e., both serial or both parallel; see also Figure 1b, c). Each participant completed two sessions, one each for the SERIAL and PARALLEL exploration modes. The order of the sessions was counterbalanced across participants. Each session contained three types of trials (vision-only, haptic-only, or bimodal), randomly intermixed. To notify the participants what the available modalities in the upcoming trial would be, a message was displayed on the screen before each trial. Participants started a trial by lifting the hand at least 35 mm above area where the virtual surface would be rendered. When the hands were lifted high enough the notification text disappeared. Exploration time was limited to 2 s, starting from the moment participants first touched the virtual plane. After the 2 s exploration the haptic and/or visual rendering of the plane ceased such that the surface disappeared. Next, two virtual buttons appeared asking participants to indicate which side of the plane was perceived as higher. In order to help participants to find the correct button during this stage in the trial, the finger positions were visually indicated using small cursors in the shape of spheres (radius 5 mm). Note, however, that these cursors only appeared after lifting the fingers at least 35 mm above the surface to prevent any additional visual cue to the orientation of the surface. The cursors disappeared as soon as the answering button was pressed and the next trial started.

The three trial types (haptic-only, vision-only, bimodal) in the PARALLEL exploration session were performed in the following way. In the haptic-only trials, the participants placed the index and middle fingers of the dominant hand on the plane and kept static contact. No visual information was displayed (Figure 1b). In the vision-only trials, participants also lowered their fingers towards the plane, but force-feedback was switched off such that participants reached through the plane. The visual Gaussian aperture blobs were projected onto the surface at the x and y positions of each finger from the moment the fingers crossed the virtual surface (projection was always on the surface, independent of the z-position of the finger). In the bimodal trials, force-feedback was switched on and visual aperture blobs were displayed at the finger positions from the moment the fingers touched the haptically rendered plane. Again, participants kept their fingers in static contact with the haptic plane.

In the SERIAL exploration session, participants used the index finger only. Again there were three trial types (haptic-only, vision-only, and bimodal). In the haptic-only trials, participants lowered their finger onto the haptically rendered plane and made lateral movements across the surface without any visual information (Figure 1c). Movements were restricted to 10 cm in the x-direction (5 cm to the left and right of the body midline). Otherwise there were no restrictions on the exploratory movements. In the vision-only trials, again, force-feedback was switched off so there was no haptic information about the surface. Participants lowered the index finger towards the surface and from the moment participants moved through the plane of the virtual surface, a Gaussian aperture blob was projected onto the plane at the x and y positions of the index finger. Again, the projection was on the surface independent of the height (z-position) of the finger and no force-feedback was provided. Participants performed lateral movements in midair to move the Gaussian aperture blob across the virtual surface in order to estimate the surface orientation. Also, in this case, movements were restricted to 10 cm in the x-direction. In the bimodal trials, force-feedback was provided and a Gaussian aperture blob was displayed at the location of the fingertip from the moment participants touched the haptically rendered surface. We checked that there were no effects of trial type (vision-only, haptic-only, or bimodal) on the number of turns from left to right or the extent of the movements.

For both exploration modes there was no force-feedback in the vision-only trials. This means that participants placed or moved the fingers somewhere in midair. Thus, proprioceptive information from the position of the fingers was not coupled to the orientation of the virtual plane. Therefore, proprioceptive information was not a cue to surface orientation in the vision-only trials. Nonetheless, participants received unconstrained proprioceptive information during exploration, which might have added noise in the vision-only trials of both exploration modes.

On any given trial, the orientation of the virtual plane varied in roll to the left or right by −15, −10, −5, −4, −3, −1, 0, 1, 3, 4, 5, 10, or 15 deg. The task was to indicate whether the left or right side of the plane was higher (i.e., whether the surface was rolled towards the right or the left). Each roll angle was repeated 15 times for each of the unimodal and bimodal trials leading to 585 trials per session (3 trial types × 13 orientations × 15 repetitions).

Analysis

To analyze the data we fitted psychometric curves (cumulative Gaussians) to participants' left/right responses as a function of surface orientation. The fits were obtained using the Psignifit toolbox for Matlab (Wichmann & Hill, 2001). The "just noticeable

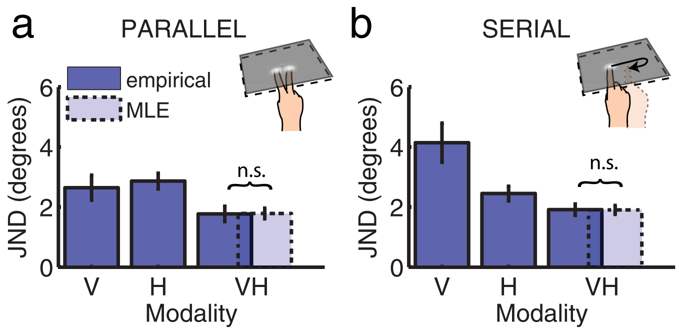

Figure 2. The JNDs for the unimodal and bimodal surface orientation estimates averaged across participants. Error bars indicate the between-subjects standard errors (*SE*). The transparent bar indicates the MLE prediction. (a) Results for parallel surface exploration. (b) Results for serial surface exploration.

difference" (JND) was taken as the difference between the 84% and the 50% cutoffs of the cumulative Gaussian. In the case of statistically optimal integration, the bimodal JND can be predicted from the unimodal JNDs using maximum likelihood estimation. This prediction is given by:

$$JND_{vh}^2 = \frac{JND_v^2 \cdot JND_h^2}{JND_v^2 + JND_h^2} \tag{1}$$

Here JND_v and JND_h represent the visual and haptic unimodal JNDs, respectively. If the unimodal signals are combined in a statistically optimal fashion, the bimodal JND should correspond to the MLE prediction and therefore be smaller than either of the two unimodal JNDs. To test for optimal integration, empirical bimodal JNDs were compared to the MLE prediction, as well as to the best unimodal JND.

Results

The JNDs averaged across participants are shown in Figure 2 for the PARALLEL and the SERIAL exploration modes. The transparent bar indicates the MLE prediction for the bimodal JNDs in the case that the unimodal signals would be integrated in a statistically optimal fashion. It can be seen that the empirical value of the bimodal JNDs were close to the values predicted by MLE. The distribution of the JNDs did not deviate from a normal distribution (Kolmogorov-Smirnov test $p > 0.1$); therefore, parametric tests were used for the statistical analysis. To test whether the bimodal JNDs were indeed smaller than either of the unimodal JNDs, the bimodal JNDs were compared to the smallest unimodal JNDs (i.e., for some participants this would be the haptic JND, and for others the visual JND). A pair-wise t test showed that

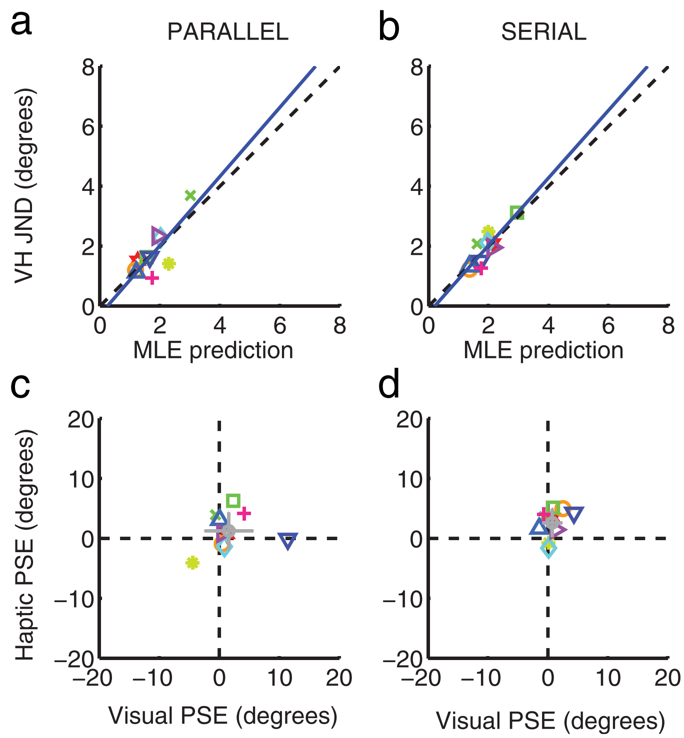

Figure 3. The individual participant bimodal JNDs are shown as a function of the MLE predictions for the PARALLEL exploration mode (a) and the SERIAL exploration mode (b). The dashed line indicates the unity line and the solid line represents the linear regression to the data. In the lower panels, the haptic PSEs are shown as a function of the visual PSEs for the PARALLEL exploration mode (c) and the SERIAL exploration mode (d). The gray dot indicates the mean of all participants and error bars indicate the standard error. Unbiased responses would fall on the intersection of the two dashed lines. Differently shaped and colored symbols correspond to the different participants in all panels.

for both the PARALLEL as well as the SERIAL exploration mode the bimodal JNDs were significantly smaller than the lowest unimodal JND, $t(9) = 2.9$, $p = 0.02$ and $t(9) = 2.7$, $p = 0.02$, respectively. Furthermore, the bimodal JNDs did not differ significantly from the MLE predicted values for either of the two exploration modes, parallel, $t(9) = 0.08$, $p = 0.93$; and serial, $t(9) = -0.09$, $p = 0.93$.

Figures 3a and b show the empirical bimodal JNDs for the individual participants as a function of their JNDs as predicted by MLE. The solid line represents the linear regression to the individual participant JNDs. It can be seen that the fitted line is close to the unity line (dashed line) for both the SERIAL and PARALLEL exploration modes. The fitted slopes did not differ significantly from 1 ($p > 0.16$). This indicates that participants performed close to statistically optimal for both exploration modes. To check whether there were systematic biases in the visual or haptic orientation estimates, the PSEs for the unimodal trials are shown in Figures 3c and d. Here it can be seen that most points are scattered around a roll angle of zero for both vision and touch. This means that on average the participants both haptically and visually perceived the surface as not inclined when it indeed had a zero-deg roll. The visual estimates were unbiased in both exploration modes. Only the haptic PSEs for serial exploration showed a significant deviation from zero roll, $t(9) = 3.6$, $p = 0.006$. This indicates there was a small bias in the haptic estimates, but only in the SERIAL exploration mode. The perceptual system can deal with such biases by, for instance, decreasing the weight that the biased estimate receives or by recalibrating the estimates (Ernst & Di Luca, 2011). Since the system has no access to the physical ground truth it has to rely on, for instance, prior experience to determine which signal is biased and there might be situations in which there is no way of knowing this. It has, however, been shown that there can be circumstances under which it is even beneficial to integrate biased signals over the option of ignoring the biased estimate (Scarfe & Hibbard, 2011). Even when sensory conflicts are introduced experimentally, integration of the sensory estimates according to MLE describes human behavior (e.g., Ernst, 2012; Ernst & Banks, 2002; Helbig & Ernst, 2007b; Knill & Saunders, 2003). Correspondingly, the current small haptic bias for serial exploration did not lead to a breakdown of integration.

Taken together, the results of Experiment 1 show that the bimodal JNDs were lower than the unimodal JNDs for both the serial and PARALLEL exploration modes. Furthermore, the bimodal JNDs did not differ from the MLE predictions. This indicates that the unimodal signals were combined in a statistically optimal fashion for the parallel as well as SERIAL exploration mode. That is, when the exploration modes for vision and haptics were the same, optimal bimodal integration occurred. Note that since optimal integration was also found for the SERIAL exploration mode, this suggests that the surface orientation estimates from the separate modalities do not have to be instantaneous for integration to occur, but can be acquired over time.

Experiment 2: Different exploration modes

In Experiment 1 we found optimal integration when the exploration mode was the same for both modalities (i.e., when visual and haptic exploration occurred both in a serial or both in a parallel fashion). To investigate whether optimal integration breaks down when the exploration mode differs between vision and haptics we conducted Experiment 2.

Methods

All 10 participants from Experiment 1 returned to participate in Experiment 2 (the maximum time after finishing Experiment 1 was four months). The setup was the same as for Experiment 1. There were two experimental sessions, one each for the different combinations of serial and parallel exploration: VISION PARALLEL – HAPTIC SERIAL and VISION SERIAL – HAPTIC PARALLEL. The sessions were presented in blocks and counterbalanced across participants. Again, there were three trial types (vision-only, haptic-only, and bimodal) randomly interleaved in each session.

In the VISION PARALLEL – HAPTIC SERIAL exploration session, all three trial types were started by lowering the index finger towards the surface (Figure 4a). In the haptic-only trials the surface was rendered using force-feedback and participants explored it by moving the index finger laterally over the surface. So, these trials were performed in exactly the same way as the haptic-only trials from the SERIAL exploration session of Experiment 1. The vision-only trials were initiated by moving the finger towards the surface, but since force-feedback was switched off, participants reached through the surface instead of touching it. At the moment the finger reached the virtual plane of the surface visual rendering in the form of a Gaussian aperture blob was switched on. In this case the visual information was not related to the location of the participant's finger and they just kept their finger statically in midair. Instead, the visual information from a vision-only trial from the PARALLEL exploration session of Experiment 1 was displayed. We took care that the presented visual information recorded from Experiment 1 matched in terms of surface orientation for the current trial of Experiment 2, and had been recorded from the same participant. Furthermore, we took care that all recordings from vision-only trials of Experiment 1 were used, to prevent possible singular behavior on individual trials in Experiment 1 from strongly influencing the results. This means that the vision-only trials from the PARALLEL exploration session of Experiment 1 were all played back in random order. Finally, in the bimodal trials, force-feedback was switched on and participants lowered the finger onto the haptically rendered surface. They moved the finger laterally across the surface (SERIAL exploration). At the same time visual information from a PARALLEL exploration bimodal trial from Experiment 1 was displayed. The surface orientation was always the same for the visual and haptic modalities. That is, we took care that the displayed visual information had been recorded from a trial in Experiment 1 with the same surface orientation. Like in the vision-only trials, participants were always

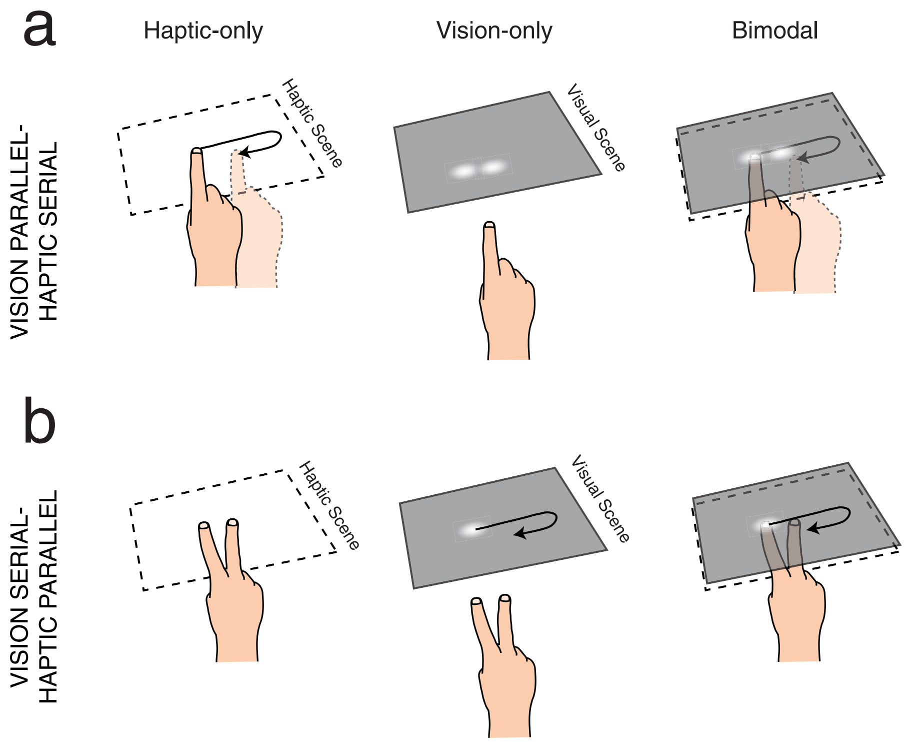

Figure 4. Exploration sessions of Experiment 2. a) In the VISION PARALLEL – HAPTIC SERIAL exploration session, participants moved their index finger over the haptically rendered surface. In the haptic-only trials, no visual information was rendered. In the vision-only trials, there was no force-feedback so participants reached through the plane of the surface. From that moment on, a visual rendering of a parallel exploration vision-only trial recorded from Experiment 1 was displayed. In the meantime, they just held their finger in midair. In the bimodal trials, both haptic and visual rendering was switched on. Participants moved the index finger over the haptically rendered surface. Visual rendering was decoupled from the finger position and the recorded visual information from a bimodal trial from the PARALLEL exploration session of Experiment 1 was shown. b) For the VISION SERIAL – HAPTIC PARALLEL exploration session, participants placed the index and middle fingers on the surface and kept static contact. In the haptic-only trials, no vision information was rendered. In vision-only trials, participants reached through the plane and held the fingers in midair. At the same time, visual information from a vision-only trial from the SERIAL exploration session of Experiment 1 was shown. In bimodal trials, participants placed the fingers on the haptically rendered surface. Visual information recorded from a bimodal trial of the SERIAL exploration session of Experiment 1 was shown, so in both exploration sessions, the visual renderings were recordings of the participants' own behavior in Experiment 1.

presented with visual information recorded from their own experimental sessions.

In the haptic-only trials of the VISION SERIAL – HAPTIC PARALLEL exploration session (Figure 4b), participants placed the index and middle fingers on the haptically rendered surface and kept static contact (PARALLEL exploration). No visual information was displayed and these trials were performed in the same way as the haptic-only trials from the PARALLEL exploration session of Experiment 1. In the vision-only trials again force-feedback was switched off and participants reached through the surface. From that moment, visual information recorded from a vision-only trial in the SERIAL exploration session of Experiment 1 was displayed. This means they held their fingers statically in midair while they saw the Gaussian aperture blob move across the surface. Force-feedback was switched on in the bimodal trials. Participants placed the fingers on the haptically rendered surface and kept static contact (PARALLEL exploration). At the same time, visual information from a bimodal trial of the SERIAL exploration session of Experiment 1 was displayed. So while participants rested their fingers statically on the surface, they saw a Gaussian aperture blob move laterally across the surface.

Results

Figure 5 shows the JNDs for the VISION PARALLEL – HAPTIC SERIAL and the VISION SERIAL – HAPTIC PARALLEL exploration sessions. The opaque bars correspond to the vision-only, haptic-only, and bimodal JNDs. The transparent bar corresponds to the MLE prediction for the bimodal JNDs. As for Experiment 1, the distribution of the JNDs did not

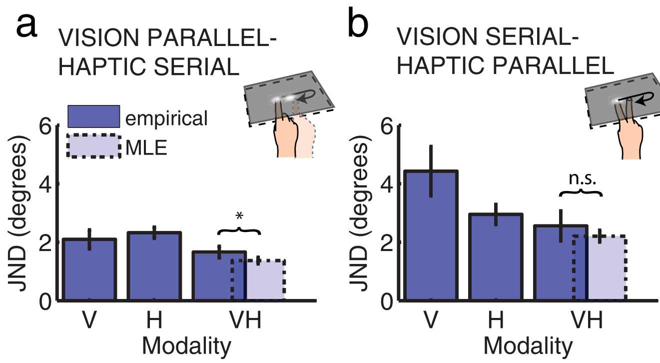

Figure 5. The JNDs for the unimodal and bimodal surface orientation estimates averaged over participants. The error bars indicate the standard error. (a) Results for the VISION PARALLEL – HAPTIC SERIAL session. (b) Results for the VISION SERIAL – HAPTIC PARALLEL sessions.

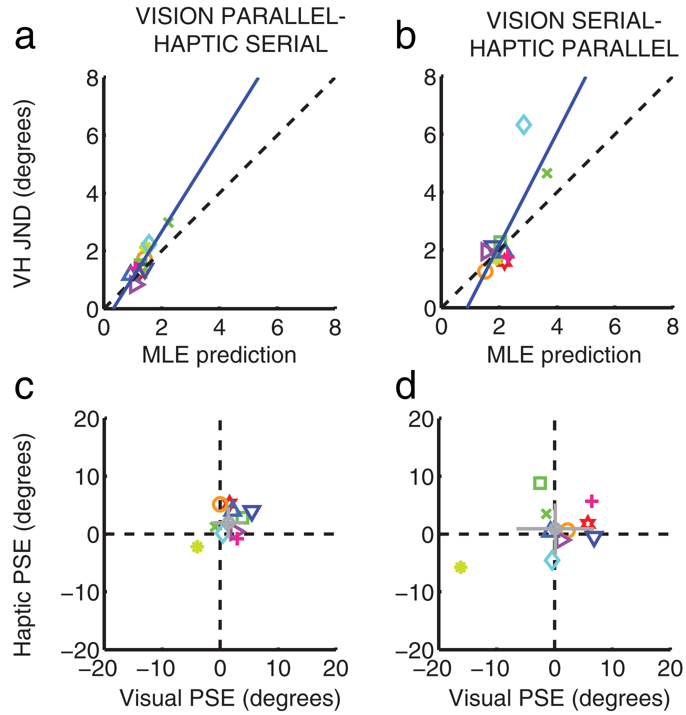

Figure 6. The individual participant bimodal JNDs as a function of the MLE prediction are shown for the VISION PARALLEL – HAPTIC SERIAL exploration session (a) and the VISION SERIAL – HAPTIC PARALLEL exploration session (b). The dashed line indicates the unity line and the solid line represents a linear regression to the data. In the lower panels, the visual PSEs are shown as a function of the haptic PSEs for the VISION PARALLEL – HAPTIC SERIAL exploration session (c) and the VISION SERIAL – HAPTIC PARALLEL exploration session (d). Colored symbols indicate individual participants. The gray dot indicates the mean across all participants and error bars indicate the standard error. Unbiased responses would fall on the intersection of the two dashed lines. Different plot symbols indicate the different participants in all panels, and are consistent with the symbols and colors used in Figure 3.

deviate from a normal distribution (Kolmogorov-Smirnov test, $p > 0.1$). It can be seen that the bimodal JNDs were larger than the MLE prediction for both exploration sessions. Particularly, for the VISION PARALLEL – HAPTIC SERIAL exploration session the bimodal JNDs did not differ significantly from the best unimodal JNDs, $t(9) = -0.2$, $p = 0.9$. Furthermore, the bimodal JNDs were significantly larger than the JNDs predicted by MLE, $t(9) = -2.8$, $p = 0.02$. For the VISION SERIAL – HAPTIC PARALLEL exploration session, the bimodal JNDs also did not differ from the best unimodal JND, $t(9) = 0.06$, $p = 0.9$, but the bimodal JNDs did not differ significantly from the MLE prediction either, $t(9) = -0.9$, $p = 0.3$.

For each of the two bimodal exploration conditions the empirical JNDs as a function of the JNDs predicted by MLE are shown for all individual participants in Figure 6a and b, together with a linear regression line. As can be seen, in the VISION PARALLEL – HAPTIC SERIAL exploration session the slope of the fitted line is larger than 1. This deviation from unity was statistically reliable, $t(9) = 2.4$, $p = 0.001$, which confirms that the signals were not combined in an optimal fashion. In Figure 6b it can be seen that, also for the VISION SERIAL – HAPTIC PARALLEL exploration session, the slope of the empirical versus predicted JND is larger than 1. Linear regression to the JNDs yielded indeed a slope that was significantly different from 1, $t(9) = 1.6$, $p = 0.04$. Furthermore, Figure 6c and d show that the PSEs were, on average, not different from an orientation of zero in both exploration sessions. There was no difference between the haptic-only and the vision-only PSEs ($p > 0.5$) in either of the exploration sessions. Note that the participants with PSEs deviating furthest from veridical were not the same as the participants showing the largest deviations from the MLE prediction for the JND (Figure 6b). This indicates that a large deviation from MLE in terms of JND cannot be explained in terms of a conflict between the senses due to systematic biases.

In Experiment 2 we introduced a difference in exploration mode between the two modalities. For the VISION PARALLEL-HAPTIC SERIAL exploration session there was a clear deviation from statistically optimal integration since the bimodal JNDs were larger than the MLE predictions. Moreover, the bimodal JNDs were not smaller than the best unimodal JNDs. For the VISION SERIAL – HAPTIC PARALLEL exploration session, it was not entirely clear whether, on a population level, visual and haptic orientation estimates were integrated in a statistically optimal fashion. The bimodal JND was neither smaller than the best unimodal JND, nor did it differ significantly from the MLE prediction. Linear regression to the individual participant JNDs, however, yielded a slope that was significantly larger than 1. This shows that larger JNDs deviated more from the MLE prediction than smaller JNDs. This is clear evidence that integration was not optimal and indicates that the nonsignificant difference from the MLE prediction on the population level was the result of noise. Thus, we can conclude that there was no statistically optimal integration when the exploration modes differed across the sensory modalities in both exploration sessions.

General discussion

In daily life humans often integrate signals from the separate senses to maximize perceptual precision. Therefore, the perceptual system needs to determine whether or not the signals correspond to the same object or event and then determine the variance of each signal in order to optimally weigh the different sources of information. Here we investigated how differences in exploration modes between the modalities affect visuohaptic integration of surface orientation. Two types of exploration modes were investigated. In the PARALLEL exploration mode, a signal was instantaneously informative of surface orientation by sampling two locations of the surface at the same time. For SERIAL exploration, the surface was explored by scanning the surface over time, while at each instant, only receiving information from one spatial location. This means that for SERIAL exploration, the perceptual estimate was acquired from sequentially sampled locations. It was hypothesized that integration might not be optimal in this situation because the orientation estimates are not acquired instantaneously. We found that this was not true as the results from Experiment 1 clearly show that for both parallel *and* serial visuohaptic exploration the multisensory JNDs were close to statistically optimal. Thus, whether visuohaptic integration is optimal does not depend on the exploration mode per se. The fact that, even for the SERIAL exploration mode, integration was close to statistically optimal furthermore indicates that the sensory signals need not be instantaneously informative about surface orientation for optimal integration to occur. This is in agreement with results from audiovisual integration of duration (Hartcher-O'Brien, Di Luca, & Ernst, 2014).

Clearly, optimal integration is possible for both parallel and serial exploration. In Experiment 2 we found that when the exploration modes differed between the modalities, the precision of the bimodal estimates deviated from statistically optimal. It is surprising that there was breakdown of integration in the VISION PARALLEL – HAPTIC SERIAL exploration session, since that session most closely resembles daily life exploration. Under most natural conditions, the field of view is much larger than the area that can be touched by a single hand. Therefore, it is often the case that a surface or object completely fits within the visual field, while the object does not fit completely in the hand. This makes it necessary to serially explore the object to obtain a haptic estimate. Because it is a very common situation to have such asymmetry between visual and haptic exploration, it might be expected that the perceptual system has learned to integrate parallel visual with serial haptic information. The results from Experiment 2 clearly show that this was not the case. One might wonder, though, why the deviation from optimality was least clear in the VISION SERIAL – HAPTIC PARALLEL exploration session, which mimics a very uncommon exploration condition. Possibly, the single Gaussian aperture blob for vision still allowed for some extraction of surface orientation information even when static. Although we tried to minimize the orientation cues provided by a single static Gaussian aperture blob, this could not be completely prevented.

The VISION PARALLEL – HAPTIC SERIAL exploration session was similar to the exploration used in the study of Rosas et al. (2005). In that study, integration was found to be suboptimal. This is in agreement with the current study because integration was also not optimal in the VISION PARALLEL – HAPTIC SERIAL exploration session. In contrast, the study of Burge et al. (2010) did show optimal integration of visuohaptic surface orientation. The experiment by Burge and colleagues, however, most closely resembled the parallel visuohaptic exploration mode from our Experiment 1, for which we also found optimal integration. Therefore, these two previous studies are in agreement with our current findings that integration is close to optimal whenever the exploration mode is the same in both modalities, but not when there is an asymmetry between the exploration modes.

Now the question arises why the sensory signals were not integrated in a statistically optimal fashion when the exploration modes differed between the modalities. Differences in exploration mode could, in principle, break the unity assumption due to either spatial or temporal discrepancies. The haptic and visual surfaces were always spatially aligned and, by providing stereo cues for vision, we took care that even in terms of the perceived metric depth of the surface, the visual and haptic estimates matched. This is in contrast to the study of Rosas et al. (2005) in which the visual surfaces were presented only monocularly, thus leading to an ambiguous visual depth estimate. Note, however, that in Experiment 1 of the present study, visual and haptic information always came from the exact same location of the surface. This means that visual and haptic information were always spatially and temporally aligned. In Experiment 2, this was not the case because the parallel modality only sampled two regions of the surface while the serial modality sampled the surface from left to right over time. Spatial and temporal misalignment between the visual aperture blob(s) and the position of the finger(s) might give a strong cue that the two sources of information may not belong together. Note that this is different from the situation in which breakdown is caused by a spatial offset between visual and haptic information as introduced in other studies (e.g., Gepshtein et al., 2005; Helbig & Ernst, 2007a). In those studies the haptic and visual stimuli were physically presented in different locations, and thus could more easily have been interpreted as two separate objects instead of one and the same. In our study the haptically and visually presented surfaces were completely aligned and congruent. Therefore, the spatiotemporal conflict existed only in the form of the location on the surface from which each modality extracted the orientation information at each instance in time.

An alternative explanation for the breakdown of integration in Experiment 2 is that the temporal buildup of the percept likely differs between the SERIAL and PARALLEL exploration modes. Presumably, the gathering of information is faster using a PARALLEL exploration mode, since the information from the simultaneously sampled locations provides information about surface orientation at each instance in time. For SERIAL exploration, a single sampled location does not provide orientation information and thus movement is required to explore the surface and the percept has to be built up over time. This means that possibly the slant estimate from the modality using PARALLEL exploration has already finalized when the estimate for the SERIAL modality is still being formed. This temporal mismatch could also have caused optimal integration to break down because it causes a temporal discrepancy in the acquisition of the two unimodal estimates. This view would be consistent with studies that found that integration of the sensory signals decreases with increasing time delay between the modalities for visuohaptic or audiohaptic sequences of events (Bresciani et al., 2006; Bresciani et al., 2005). It is not likely, however, that in the current study a temporal discrepancy in the acquisition of the estimates caused the visual and haptic information to not be linked to the same source or object, given the relatively long exploration time (2 s). It could, however, still be the case that two estimates are only integrated in an optimal way if they are both acquired within a certain time window, regardless of the fact that they clearly originate from the same source.

Active control of object exploration strategies is a hallmark of haptic perception. Recently it has been shown that exploratory movements are adapted to optimize the haptic percept (Drewing, 2012). In daily life, however, we mostly combine serial haptic exploration with parallel visual exploration. Our study clearly shows that when these two exploration modes are used in combination, the resulting bimodal estimate is worse than it would be when using similar exploration modes in both modalities. Although there have been many accounts of situations in which multimodal estimates are combined in an optimal way, in daily life visuohaptic perceptual estimates might often not be optimal at all due to the differences in exploration mode between vision and touch.

Keywords: surface orientation, vision, haptics, multisensory integration, maximum likelihood estimation

Acknowledgments

This work was supported by a Netherlands Organisation for Scientific Research (NWO) grant and a NWO VENI grant awarded to MAP and by the European Commission with the Collaborative Project Wearable Haptics for Humans and Robots (WEARHAP) (EU FP7/2007-2013 project n 601165 WEARHAP) for MOE. We thank the Max Planck Institute for Biological Cybernetics for their support with the experimental equipment. We acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Funds of Bielefeld University Library.

Commercial relationships: none
Corresponding author: Myrthe Plaisier.
Email: M.A.Plasier@vu.nl.
Address: Faculty of Human Movement Sciences, Research Institute MOVE, VU University, Amsterdam, The Netherlands.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262.

Bresciani, J., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision, 6*(5)2, 554–564, http://www.journalofvision.org/content/6/5/2, doi:10.1167/6.5.2. [PubMed] [Article]

Bresciani, J., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: Auditory signals can modulate tactile tap perception. *Experimental Brain Research, 162*(2), 172–180.

Burge, J., Girshick, A. R., & Banks, M. S. (2010).

Visual-haptic adaptation is fetermined by telative teliability. *Journal of Neuroscience, 30*(22), 7714–7721. doi:10.1523/JNEUROSCI.6427-09.2010.

Di Luca, M. (2010). New method to measure end-to-end delay of virtual reality. *Presence, 19,* 569–584.

Drewing, K. (2012). After experience with the task humans actively optimize shape discrimination in touch by utilizing effects of exploratory movement direction. *Acta Psychologica, 141*(3), 295–303, doi: 10.1016/j.actpsy.2012.09.011.

Ernst, M. (2012). Optimal multisensory integration: Assumptions and limits. In B. Stein (Ed.), *The new handbook of multisensory processes* (pp. 1084–1124). Cambridge, MA: MIT Press.

Ernst, M., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433.

Ernst, M., & Di Luca, M. (2011). Multisensory perception: From integration to remapping. In J. Trommershaeuser & M. S. Landy (Eds.), *Sensory cue integration* (pp. 224–250). New York: Oxford University Press.

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2010). Visual-vestibular cue integration for heading perception: Applications of optimal cue integration theory. *European Journal of Neuroscience, 31*(10), 1721–1729, doi:10.1111/j.1460-9568.2010.07207.x.

Gepshtein, S., & Banks, M. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology, 13*(6), 483–488, doi:10.1016/S0960-9822(03)00133-7.

Gepshtein, S., Burge, J., Ernst, M., & Banks, M. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision, 5*(11):7, 1013–1023, http://www.journalofvision.org/content/5/11/7, doi:10.1167/5.11.7. [PubMed] [Article]

Hartcher-O'Brien, J., Di Luca, M., & Ernst, M. O. (2014). The duration of uncertain times: Audiovisual information about intervals is integrated in a statistically optimal fashion. *PLOS ONE, 9*(3), e89339, doi:10.1371/journal.pone.0089339.

Helbig, H. B., & Ernst, M. (2007a). Knowledge about a common source can promote visual-haptic integration. *Perception, 36*(10), 1523–1533.

Helbig, H. B., & Ernst, M. O. (2007b). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179*(4), 595–606.

Knill, D., & Saunders, J. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43*(24), 2539–2558, doi:10.1016/S0042-6989(03)00458-9.

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology, 19,* 342–368.

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113, doi:10.1016/0028-3932(71)90067-4.

Reuschel, J., Drewing, K., Henriques, D. Y. P., Roesler, F., & Fiehler, K. (2010). Optimal integration of visual and proprioceptive movement information for the perception of trajectory geometry. *Experimental Brain Research, 201*(4), 853–862, doi:10.1007/s00221-009-2099-4.

Rohde, M., & Ernst, M. (2012). To lead and to lag—Forward and backward recalibration of perceived visuo-motor simultaneity. *Frontiers in Psychology, 3*(599).

Rosas, P., Wagemans, J., Ernst, M., & Wichmann, F. (2005). Texture and haptic cues in slant discrimination: Reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A—Optics, Image Science, and Vision, 22*(5), 801–809, doi:10.1364/JOSAA.22.000801.

Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision, 11*(7):12, 1–17, http://www.journalofvision.org/content/11/7/12, doi:10.1167/11.7.12. [PubMed] [Article]

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*(1), 147–152.

Takahashi, C., Diedrichsen, J., & Watt, S. J. (2009). Integration of vision and haptics during tool use. *Journal of Vision, 9*(6):3, 1–13, http://www.journalofvision.org/content/9/6/3, doi:10.1167/9.6.3. [PubMed] [Article]

van Beers, R., Sittig, A., & van der Gon, J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology, 81*(3), 1355–1364.

van Dam, L., Parise, C., & Ernst, M. (2014). Modeling multisensory integration. In D. Bennett & C. Hill (Eds.), *Sensory integration and the unity of consciousness*. (pp. 209–229). Cambridge, MA: MIT Press.

Wichmann, F., & Hill, N. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293–1313.