Published in final edited form as:

Exp Brain Res. 2011 April ; 210(1): 67–80. doi:10.1007/s00221-011-2603-5.

Visual-haptic cue integration with spatial and temporal disparity during pointing movements

Sascha Serwe

Department of Psychology, Giessen University, Otto-Behaghel-Str. 10F, 35394 Giessen, Germany julia.trommershaeuser@nyu.edu

Department of Psychology, FernUniversität in Hagen, Universitätsstr. 11, 58084 Hagen, Germany

Konrad P. Körding

Rehabilitation Institute of Chicago, Northwestern University, 345 E Superior Street, Chicago, IL 60611, USA konrad@koerding.com

Julia Trommershäuser

Department of Psychology, Giessen University, Otto-Behaghel-Str. 10F, 35394 Giessen, Germany julia.trommershaeuser@nyu.edu

Department of Psychology and Center for Neural Science, New York University, 4 Washington Place, Room 809, New York, NY 10003, USA

Abstract

Many perceptual cue combination studies have shown that humans can integrate sensory information across modalities as well as within a modality in a manner that is close to optimal. While the limits of sensory cue integration have been extensively studied in the context of perceptual decision tasks, the evidence obtained in the context of motor decisions provides a less consistent picture. Here, we studied the combination of visual and haptic information in the context of human arm movement control. We implemented a pointing task in which human subjects pointed at an invisible unknown target position whose vertical position varied randomly across trials. In each trial, we presented a haptic and a visual cue that provided noisy information about the target position half-way through the reach. We measured pointing accuracy as function of haptic and visual cue onset and compared pointing performance to the predictions of a multisensory decision model. Our model accounts for pointing performance by computing the maximum a posteriori estimate, assuming minimum variance combination of uncertain sensory cues. Synchronicity of cue onset has previously been demonstrated to facilitate the integration of sensory information. We tested this in trials in which visual and haptic information was presented with temporal disparity. We found that for our sensorimotor task temporal disparity between visual and haptic cue had no effect. Sensorimotor learning appears to use all available information and to apply the same near-optimal rules for cue combination that are used by perception.

Keywords

Multisensory integration; Hand movement control; Motor learning; Cue integration; Vision; Haptics

Introduction

Recent work has emphasized that the processing of noisy sensory cues is one of the central problems the brain needs to overcome to generate a robust percept. Bayesian statistics provide a rule of how to combine sensory information (for review, see e.g., Ernst and Bühlhoff 2004; Landy et al. 1995). According to this rule, two redundant signals affected by Gaussian noise are integrated linearly, whereas the weight of each modality is proportional to the inverse variance of each signal. Many perceptual studies have shown that behavioral results closely match these optimal integration rules. For example, visual depth cues are integrated to form a consistent percept of depth in a scene (Hillis et al. 2002, 2004; Landy et al. 1995) or the slant of a surface (Knill and Saunders 2003; Louw et al. 2007; Sousa et al. 2009; Knill 1998). Close-to-optimal integration has also been observed when information is integrated across sensory modalities, e.g. during size estimation under multisensory conditions (Ernst and Banks 2002) or by estimating the origin of a signal using visual and auditory cues (Alais and Burr 2004). In the motor domain, optimal integration has been demonstrated for estimating hand position by visual and proprioceptive cues (van Beers et al. 1999, 2002) and for the integration of prior information with new sensory feedback in a reaching task (Körding and Wolpert 2004b). Optimal integration rules can also successfully describe many cognitive effects (Griffiths and Tenenbaum 2006) and can thus be seen as a prominent phenomenon.

Here, we tested whether visual and haptic cues are combined following similar combination rules in simple reaching tasks. The work presented here extends previous work by Serwe et al. (2009). In that work, subjects judged the direction of force pulses applied to the fingertip during a pointing movement. At the same time, additional directional visual information was presented and subjects were instructed to use this cue when resolving the applied perturbation direction. Unlike in most other cue integration studies, the cues in this task did not belong together naturally, i.e., did not originate from the same object. Nonetheless, there was a reason to integrate the information across senses: subjects were instructed and learned implicitly that the visual cue was a reliable and valid cue for force pulse direction. However, under the task conditions of this previous study, no integration of visual and haptic information was found. Instead, subjects based their decision on a probability mixture of the individual cue estimates without integrating information across modalities, that is, in a single trial, subjects reported either the direction of the visual cue or the direction of the proprioceptive cue. In the present study, we tried to increase the causal structure between the cues to promote cue integration. Still, the cues did not originate from the same object. However, they were closely related because both allowed to infer the location of the same pointing target. Visual and haptic information could be combined to increase pointing accuracy. In each trial, we either presented a haptic cue or a visual cue or a combination of both cues that provided information about the target position during the reach (see method section for further information). We shifted the target position randomly to a new position across trials, forcing our subjects to rely strongly on the two cues provided during the reach. Under these conditions, both cues allow inferring information about the same, but unknown target position. According to Bayesian ideas, they should yield a more reliable estimate about the target position if integrated under multisensory conditions. Visual and haptic information really belong together because they are related to the same target and allow inferring its position. Thus, even under conditions in which visual and haptic cues are spatially separated, we expected cue integration. In addition, we varied the delays between the onset of the visual and haptic cues to test for effects of synchronicity. We measured unimodal and bimodal performance and modeled pointing accuracy in unimodal and bimodal conditions with several alternative models. We tested specifically for modality-specific effects of changes in motor noise. Using the Bayesian Information Criterion (BIC, see Schwarz 1978) as a criterion, we decided on a multisensory decision model as the best

fitting model. With respect to motor noise, the best fitting model includes a motor noise term in all haptic conditions and no motor noise in the visual only trials. Alternative models that include either the same motor noise in all condition or no motor noise at all, perform worse. However, all three motor noise variations of the multisensory decision model clearly outperform two decision models which postulate that visual and haptic information is not integrated, but that only the visual cue, or, alternatively the most reliable cue, determines the estimate about the target position.

Materials and methods

Subjects

Eight right-handed subjects, two men and six women, aged between 21 and 32, participated in this experiment. All (except SS, the first author) were undergraduate students and naïve to the purpose of the experiment. They were paid for their participation. All subjects had normal or corrected-to-normal vision.

Apparatus

Participants sat in front of a visuo-haptic setup in a dimly lit, quiet room. The apparatus consisted of a PHANToM 3.0L haptic force-feedback device (temporal resolution = 1,000 Hz, spatial resolution = ~0.03 mm, force feedback in the three translatory directions) and a 22" computer screen (Iiyama Vision Master Pro 514, 120 Hz, 960 × 1,280 pixels). The right index finger was connected to the PHANToM via a thimble-like holder. The haptic workspace of the PHANToM was spatially aligned with the visual scene. However, movement in depth was ignored here, that is, we just used the left–right and top–down coordinates of the PHANToM device relative to the observer. The visual scene was a flat monitor surface that projected onto a mirror that reflected the images of the computer screen. Head position was fixed relative to the screen using a chin rest. Subjects listened to music via headphones to mask possible sources of noise created by the mechanics of the PHANToM. The experiment was run on a PC (2.8 Ghz; 1 GB RAM) using C++ code to control the apparatus, present the stimuli, and track the finger.

Task

Subjects were required to make rightward pointing movements from the starting point to a target located on a line which was 32 cm to the right (see Fig. 1a, target line). The exact height (y position; distance from starting point upwards) of the target varied from trial to trial (for each condition, 80 unique targets evenly distributed between 6 and 21 cm were presented randomly). We instructed the subjects to point as accurately as possible within the overall time limit of 2,000 ms after a visual go signal. If the total trial duration (reaction time + movement time) took longer than 2,000 ms, the words “too late” appeared on the screen and the trial was repeated later.

Subjects pointed open loop. That is, they did not see their own hand during the pointing movement. Between successive trials, a small circle (diameter 4 mm) represented the moving index finger in the virtual environment so that subjects could maintain a reliable estimate of their finger position. As soon as the subject started the trial, the dot was rendered invisible until the movement ended. At the beginning of each trial, subjects cannot know the vertical position of the target, but information was provided during the movement as follows: once the finger had covered a distance of 1 cm and after a subsequent temporal delay of either 50 or 250 ms, a visual or a haptic stimulus provided the information about the vertical position of the target. In each trial, we either presented the visual cue, the haptic cue, or a combination of both (see text below and Fig. 1 for the exact stimulus description).

The movement was complete, once the finger crossed the target line. Following movement completion, subjects received feedback about the true target position (represented by a small red circle, diameter 4 mm) and about their final finger position (represented by a small green circle, diameter 4 mm). In addition, we provided feedback about their pointing accuracy. Hits to the center of the target were rewarded with 100 points. For 1 mm distance between target center and the middle of the fingertip, 3 points were subtracted. Therefore, subjects started to lose points when there was a small overlap between the target and the virtual fingertip. The penalty increased linearly with increasing target-fingertip distance, and for distances larger than 33 mm, zero points were earned. The points obtained were presented on every trial and converted into a monetary bonus, awarded at the end of the experiment (1 € for 10,000 points).

Stimuli

Haptic trials—A force pulse was applied to the fingertip for 100 ms using the PHANTOM device. The force pulse increased and decreased in a sinusoidal fashion within this interval to ensure a gradual pulse on- and offset (Fig. 1b). The applied force amplitude was scaled linearly with the vertical target position: 1 N corresponded to a target that was presented 6 cm above the starting position, 3.5 N corresponded to a target position that was presented at a height of 21 cm above the starting position. Thus, the information given by the haptic cue is indirect and the perceived force pulse strength had to be transferred into the corresponding target height by the subject. After movement onset, we presented the haptic cue early (50 ms, *HE*) or late (250 ms, *HL*). Movement onset was detected online using a 1 cm threshold on the distance between start position and finger position.

Visual trials—In contrast to the haptic cue, the visual cue provided direct information about the target location. The visual cue consisted of three small blue dots, presented for 100 ms and drawn from a Gaussian distribution centered on the true target position (Fig. 1c; two-dimensional Gaussian, standard deviation in *y*-direction 4 cm, standard deviation in *x*-direction 1 cm). This is a version of standard stimuli used in previous studies of uncertainty in sensorimotor integration (Körding and Wolpert 2004b; Tassinari et al. 2006). After sampling from this distribution, the dots were shifted in the *x*-direction (horizontal) until the centroid fell onto the target line. Only the variation in the *y*-direction (vertical) was informative. Due to the sampling procedure, the vertical position of the centroid did not match the correct target position perfectly, but varied around the target position in the vertical direction with a standard deviation of about 2.3 cm ($4/\sqrt{3}$ cm). Thus, aiming at the centroid in vision only trials maximized the probability of a hit, but was not very reliable. To increase the relative benefit of cue integration, the reliability was adjusted such that the visual trials were approximately as reliable as the haptic trials.

Matching the conditions of the haptic cue, we presented the visual cue early during the movement (50 ms after movement onset, *VE*) or late (250 ms after movement onset, *VL*).

Visual-haptic trials—Stimuli in the visual-haptic trials were presented as described above with all possible combinations of visual and haptic cue onset: visual and haptic information presented early (50 ms after movement onset, both early, *BE*), visual and haptic information presented late (250 ms after movement onset, both late, *BL*), visual information presented early and haptic information presented late (stimulus onset delay of 200 ms, visual first, *VF*), and haptic information presented early and visual information presented late (stimulus onset delay of 200 ms, haptic first, *HF*).

Experimental sessions and design

Training session—Subjects first completed 160 visual trials to practice the visual task (*VE* and *VL* intermixed) followed by 160 haptic practice trials (*HE* and *HL* intermixed). Following these unimodal practice trials, they practiced the task under multisensory conditions in 160 trials with all possible trial types randomly intermixed. This last sequence matched the experimental condition except that our subjects did not earn money for the points they collected during these practice trials.

Experimental sessions—Following the 480 practice trials, we collected data during two experimental sessions. Each session consisted again of 80 training trials with only haptic cues (*HE*, *HL* randomly intermixed). Subjects then trained again on 80 trials that were fully equivalent to the experimental conditions (all trial types randomly intermixed) except that they did not earn money for the points they collected. This was followed by 320 experimental trials (4 blocks with 80 trials) in each of the two sessions, resulting in a total of 640 trials (all 8 conditions of cue onset randomly intermixed).

Model of optimal cue combination

We compared the behavioral data in our experiment to the predictions of an optimal multisensory decision model. The model computes predictions about the optimal finger position based on Maximum Likelihood Estimation (MLE) as we explain next. In the conditions in which only a visual cue is provided, a reasonable strategy could be to aim at the centroid of the three dots. However, this strategy would ignore the information subjects may have about the distribution of targets. Though we used a boxcar distribution for target selection, subjects tended to judge the center of the configuration to be more likely with decreasing likelihoods for higher or lower targets. This finding matches previous results which have demonstrated that subjects tend to treat non-Gaussian distributions as if they were Gaussian (Körding and Wolpert 2004a). We therefore fitted the data obtained assuming a Gaussian prior. In the conditions in which only a visual cue is provided, the predicted hit position $\hat{x}_V$ is therefore defined by the weighted average between the prior position P and the position of the centroid C . The optimal weights are the normalized inverse variances of prior and centroid. This yields the following prediction for the mean hit position in visual trials:

$$\hat{x}_V = \frac{1/\sigma_p^2}{1/\sigma_p^2 + 1/\sigma_v^2} P + \frac{1/\sigma_v^2}{1/\sigma_p^2 + 1/\sigma_v^2} C. \quad (1)$$

Using the Bayesian framework, we simultaneously fitted the total variance in each condition as an optimal combination of the variances σ_p^2 and σ_v^2 :

$$\hat{\sigma}_{T_V}^2 = 1 / \left(\frac{1}{\sigma_p^2} + \frac{1}{\sigma_v^2} \right). \quad (2)$$

We similarly modeled the trials in which only haptic information was provided, by using the corresponding estimate for the haptic variance σ_h^2 and by replacing C , with the position represented by the strength of the force pulse F . In contrast to the visual conditions, we modeled different haptic variances for *HE* when compared to *HL* trials (see below that variances of *HE* and *HL* trials indeed differ significantly whereas there is no difference in variances of *VE* when compared to *VL* trials). We further assume that the prior position P

and the variance of the prior σ_p^2 remain constant across trials. Therefore, the prediction for the mean hit position in *HE* trials is

$$\hat{x}_{HE} = \frac{1/\sigma_p^2}{1/\sigma_p^2 + 1/\sigma_{HE}^2} P + \frac{1/\sigma_{HE}^2}{1/\sigma_p^2 + 1/\sigma_{HE}^2} F. \quad (3)$$

However, in haptic trials, we expected that the force itself introduces some additional uncertainty to the movement and therefore affects motor execution. As a result, the accuracy in haptic conditions is reduced by a noticeable amount whereas it is negligible in the visual conditions. This is modeled by an additional motor error term σ_M^2 that is part of the total variance in all haptic trials and cannot be reduced further. The total variance in *HE* trials is therefore:

$$\sigma_{T_{HE}}^2 = 1 / \left(\frac{1}{\sigma_p^2} + \frac{1}{\sigma_{HE}^2} \right) + \sigma_M^2. \quad (4)$$

The only difference for *HL* trials is that σ_{HE}^2 is replaced by σ_{HL}^2 in Eqs. 3 and 4.

For bimodal trials, we modeled optimal integration of visual and haptic information and the prior distribution. Again, we have different variances for the *HE* and *HL* conditions whereas the *VE* and *VL* are expected to be the same. The predicted optimal hit position in trials where the force pulse is presented early (*BE* and *HF*) is therefore defined by

$$\begin{aligned} \hat{x}_{BE} &= \hat{x}_{HF} \\ &= \frac{1/\sigma_p^2}{1/\sigma_p^2 + 1/\sigma_V^2 + 1/\sigma_{HE}^2} P \\ &\quad + \frac{1/\sigma_V^2}{1/\sigma_p^2 + 1/\sigma_V^2 + 1/\sigma_{HE}^2} C + \frac{1/\sigma_{HE}^2}{1/\sigma_p^2 + 1/\sigma_V^2 + 1/\sigma_{HE}^2} F. \end{aligned} \quad (5)$$

Since there is a force pulse present in all bimodal trials, the total end point variance around the optimal hit position includes the motor error term σ_M^2 :

$$\sigma_{T_{BE}}^2 = \sigma_{T_{HF}}^2 = 1 / \left(\frac{1}{\sigma_p^2} + \frac{1}{\sigma_{HE}^2} + \frac{1}{\sigma_V^2} \right) + \sigma_M^2. \quad (6)$$

For *BL* and *VF* trials, the force pulse is presented late and σ_{HE}^2 is exchanged by σ_{HL}^2 in Eqs. 5 and 6.

Model evaluation—For every subject, we individually computed predictions for mean end point position and end point variance by simultaneously fitting the data from all 8 experimental conditions (640 trials). We used 6 free parameters (σ_{HE}^2 , σ_{HL}^2 , σ_V^2 , μ_P , σ_P^2 , σ_M^2) and Eqs. 1–6 (including haptic variations of Eqs. 3–6). Model parameters were fit to match precision and bias at the same time. The fitted parameters were then used to simulate 10,000 repetitions of each experimental trial. We measured the standard deviation of pointing error for each condition in the distribution of simulated data and compared it with the observed standard deviation of pointing error for each subject.

Comparison with alternative models—In addition, we tested four alternative models and compared the overall performance of these models to the model described above. We used the Bayesian Information Criterion (BIC, see Schwarz 1978) as a criterion to select the best fitting model. The BIC provides a measure that allows comparing model predictions for models with different numbers of parameters by computing a trade-off between goodness of fit and number of used parameters by penalizing the use of additional parameters.

Here, we decided to compare the predictions of our model to four alternative models. The first alternative model tests the hypothesis that motor noise affects motor execution in general and not particularly in the force pulse conditions (*same motor noise*). Therefore, the motor noise term σ_M^2 is added in the visual conditions as well and Eq. 2 is modified to

$$\hat{\sigma}_{T_V}^2 = 1 / \left(\frac{1}{\sigma_p^2} + \frac{1}{\sigma_v^2} \right) + \sigma_M^2, \quad (7)$$

, whereas Eqs. 1 and 3–6 stay unmodified. The second alternative model tests the hypothesis that the motor noise in all conditions is negligible (*no motor noise*). Therefore, the motor noise term σ_M^2 is omitted not only in the visual conditions but in all condition. Therefore, it is deleted from Eqs. 4 and 6. Here, Eqs. 1–3 stay unmodified.

The third alternative model (“take visual”) tests the null hypothesis that visual and haptic information is not integrated at all. Instead, in the bimodal trials the subject always chooses as if only the visual cue is presented. Thus, this model consists of Eqs. 1 and 2 for all visual and bimodal trials and Eqs. 3 and 4 for haptic trials. However, the motor noise parameter of Eq. 4 is excluded here. This model therefore predicts that pointing accuracy does not increase under bimodal conditions when compared to the visual condition.

The fourth alternative model (“take best unimodal”) also tests the null hypothesis that visual and haptic information is not integrated. However, here it is assumed that in the bimodal trials the subject chooses the cue that provides the most reliable estimate about the target position, i.e., either the visual cue or the haptic cue, ignoring the other. Thus, this model consists of Eqs. 1–3 and 4 without motor noise. For all bimodal conditions, the corresponding unimodal equations were chosen depending on the most reliable cue in a given condition (mostly, but not always, the visual cue). This model therefore predicts that pointing accuracy does not increase under bimodal conditions when compared to the most accurate unimodal condition (the unimodal condition with the lowest variance).

Results

In our experiment, subjects were instructed to estimate the position of a hidden target based on noisy visual and proprioceptive cues provided midway during the reach. We tested whether human subjects combine cues in a near optimal way as typically found during perceptual multisensory estimation tasks. We also tested whether the temporal delays between visual and proprioceptive cues affected integration performance.

Performance during training

We first tested whether human subjects learned to efficiently perform in our multisensory reaching task. During training, visual and haptic trials were presented in separate blocks. During these unimodal practice trials, pointing performance improved only slightly over time, which resulted in a relatively constant standard deviation of the pointing error. This indicates that our participants understood and learned the task almost instantaneously and did not improve much with increased practice (Fig. 2). In the visual early condition, all

subjects performed very close to the predicted optimum for the visual only trials. Assuming that a subject hits the centroid perfectly in every trial, the standard deviation of pointing errors would be 2.3 cm ($4/\sqrt{3}$ cm). Accordingly, we found little variance across subjects and the standard error of the mean across subjects was very small compared to the other conditions (where the variance across subjects is higher).

During the initial training (shown in Fig. 2), subjects practiced visual trials and haptic trials in successive blocks (2 visual blocks with 80 trials per block, then 2 haptic blocks with 80 trials per block). Following this initial training, subjects repeated the training trials, this time with all 8 possible unimodal and bimodal conditions randomly intermixed (20 trial per condition, 2 blocks with 80 trials per block). This random mixture was consistent with the conditions of the main experiment that followed subsequently. At that stage, performance clearly had converged and did not change remarkably throughout the experiment (Fig. 3).

To test this statistically, we tested whether the variance of pointing errors in the first half of trials in one condition differed from the variance of pointing errors in the second half of trials in the same condition, using a two-sample *F*-test; 5 out of 8 subjects did not show any significant change in any of the eight conditions (all $p < .05$). Each of the remaining three subjects improved in only 1 out of 8 conditions: two subjects reduced their variance in the haptic late condition, while one subject showed a reduction in variance in the both early condition. We interpret this result as evidence that pointing performance had stabilized following initial training.

Effects of visual and haptic cue onset during the reach

Subjects made use of both haptic and visual information presented during the reach. Both visual and haptic cue onset biased the pointing movement toward the direction indicated by the cue (see also Schmidt 2002, for similar effects found during a visual motor priming task). The trajectories confirm that subjects use the cue information to adjust their movement during the reach in the direction of the (hidden) target position (see Fig. 4 for data from one representative subject, RZ). For analysis, we grouped the targets into three equally sized groups of high (ranging from 16.2 to 21.0 cm), middle (ranging from 11.1 to 15.9 cm), and low targets (ranging from 6 to 10.8 cm). Trajectories aimed at different target heights differed significantly with respect to the mean finger position by the time the finger crossed the target line. Figure 4 also illustrates that, in general, subjects performed well in both unimodal and bimodal conditions as indicated by the small difference between the mean (hidden) target position and the mean final finger position. In all conditions, subjects exhibited a slight tendency to hit too low for high and too high for low targets, indicating a bias introduced by a central prior. Subjects hit high targets too low in 69% of the trials (subjects range between 57% and 91%) whereas subjects hit low targets too high in 71% of the trials (subjects range between 60% and 80%). The tendency to hit middle targets too high was 54% (with 5 subjects hitting more often too high and 3 subjects hitting more often too low). This obvious tendency to aim for the middle likely reflects an initial movement plan to aim for the middle of all possible locations. From there, any necessary adjustment during the movement would, on average, be minimized, yet cause a “middle bias”, i.e., a remaining tendency to aim for the middle, due to the initial movement plan which was also already partly executed.

In all trials, it took some time until cue information was processed. In all conditions containing haptic information, the force pulse (represented by a black bar on the *x*-axis in Fig. 4) had a direct and immediate effect on the pointing movement. This is revealed by deviations of up to 2 cm between the mean trajectory of strong force pulses (high targets) when compared to the mean trajectory of weak force pulses (low targets). However, the

distance between trajectories toward high, middle, and low targets stayed largely constant for some time following the force pulse suggesting that initially the force cue resulted in a passive displacement only. For the representative subject shown in Fig. 4, trajectories for the different target heights in the haptic early condition (solid line) started to diverge from this initial distance around 450 ms after movement onset (400 ms after force pulse onset). For the force pulse in the haptic late condition (dashed line), the trajectories started to diverge around 550 ms after movement onset (300 ms after force pulse onset). This suggests that force pulse corrections to planned reach direction occurred faster for late presentation when compared to early presentation. This effect is more pronounced in the conditions in which we presented the visual cue. The trajectories in the visual conditions start to diverge from each other earlier in the reach for *VE* trials (350 ms after movement onset, 300 ms after stimulus onset), and again occur faster for cues provided late in the reach (500 ms after movement onset, 250 ms after cue onset in *VL*). The delay between movement plan updating following presentation of the cue in the bimodal trials was as short as in the visual conditions, but we did not find evidence for a further reduction. One reason for the faster update in movement plan for visual cues when compared to haptic cues is that no transfer of sensory information about the target position was needed in the visual conditions. As described in the stimulus section, the visual cue provided direct information about the target position whereas the information given by the haptic cue was indirect and the force pulse strength had to be translated into target height. But note that due to extensive learning this translation is still a very fast and automatic process.

We emphasize that there was no time pressure to begin or complete the movement. Subjects were instructed to complete their reach within 2,000 ms following presentation of the go-signal. The mean reaction time (time between go-signal and movement start) across all subjects and all trials was 418 ms (individual mean reactions times across all trials ranged from 373 to 518 ms). The mean total trial duration (reaction time + movement time) across all subjects and conditions was 1,357 ms (individual mean trial durations times ranged between 1,174 and 1,592 ms). This suggests that subjects did not exceed the time limit and were still able to make very careful and slow reaches to allow for processing of both cues.

Model predictions

We finally compared pointing performance in unimodal and bimodal conditions to the predictions of the multisensory decision model (Eqs. 1–6). Fig. 5 shows the comparison between the simulated standard deviation of pointing error and the observed standard deviation of pointing error. Data for each subject are shown individually. As shown in Fig. 5, the model fits the data well. Since our model assumes optimal integration of information in the bimodal trials, this result suggests that subjects indeed performed close to optimal and integrated all available information.

Figure 6 shows the averaged results across subjects. For all conditions, the mean observed SD of the pointing error matches the predictions for the mean predicted SD of the pointing error. In addition, Fig. 6 illustrates several other important findings. Our model includes the assumption that timing does not matter for the visual cue (*VE* = *VL*) whereas it does matter for the haptic cue (*HE* ≠ *HL*). The results shown in Fig. 6 are in line with this assumption. Across subjects, we found a significantly different standard deviation of pointing error between *HE* and *HL* conditions, $t(7) = 3.40$, $p < .05$, and no significant difference between *VE* and *VL* condition, $t(7) = 0.16$, $p = .88$. The correspondence between model and data in the bimodal conditions with temporal disparity (*HF*, *VF*) suggests that sequential presentation had no influence on the amount of cue integration. The model predicts the same performance for both sequential (*HF*, *VF*) versus simultaneous (*BE*, *VL*) cue onset. Consistent with this hypothesis, a 2 (force pulse: early vs. late) × 2 (timing: simultaneous vs. sequential) repeated measures ANOVA shows that there is no significant main effect for

force pulse and no significant interaction, but most importantly no significant main effect for timing of cue onset (all $p > .05$ in all conditions). To test for improvements in performance under bimodal conditions, we compared performance in the bimodal conditions to performance in the unimodal conditions. Across subjects, the mean standard deviation of pointing errors in the four bimodal conditions ($M_{\text{bimodal}} = 2.29$) was significantly smaller than the standard deviation of pointing errors in the four unimodal conditions ($M_{\text{unimodal}} = 2.67$), $t(62) = 3.38$, $p < .05$. However, to clearly demonstrate an improvement in the bimodal conditions, we compared each bimodal condition to its best unimodal component (and not to the mean unimodal performance). In addition, the unimodal reference point for each bimodal condition differed: bimodal conditions consist of a combination of a visual cue (presented either early or late) and a haptic cue (presented either early or late). Thus, we compared the standard deviation of pointing error in all bimodal conditions to the standard deviation of pointing error in its best unimodal component (i.e., *BE* and *VF* are compared to *VE*, *BL* and *HF* are compared to *VL*). The standard deviations of pointing error in all bimodal conditions were smaller than the standard deviations of pointing errors in the corresponding unimodal conditions and all but one of these differences were significant (all $p < .05$ for all conditions except the comparison between *VE* and *VF*, $p = .49$). This indicates that subjects increased their accuracy if there was more than one cue available.

Comparison to alternative models

Finally, we compared the explanatory power of this model to the predictions made by several alternative models using the Bayesian Information Criterion (BIC, see Schwarz 1978). The BIC helps to find a trade-off between optimal model fit and parsimony of the number of parameters used by the different models. The absolute BIC value is meaningless. Therefore, we computed the difference between the BIC value of the standard model and the BIC value of the alternative models. Table 1 shows the relative BIC values of the different models, the parameter count and the fitted parameter values. A lower BIC value suggests that the model with the lower BIC value should be preferred. The *same motor noise* model uses the exact same 6 parameters as the model defined by Eqs. 1–6, but this alternative model includes motor noise even in the visual conditions where the force pulse is absent. However, the model fit to the data is worse ($\Delta\text{BIC} = 15.29$), and therefore we rejected the *same motor noise* model. However, completely omitting motor noise in all conditions, as in the *no motor noise* model, is not a good alternative either. The *no motor noise* model reduces the number of parameters to 5. However, the model comparison indicates that this model is significantly worse ($\Delta\text{BIC} = 11.68$). Most importantly, both models that assume no integration (*take visual*, *take best unimodal*) fit clearly worse than any of the other alternative models (*take visual*: $\Delta\text{BIC} = 38.40$, *take best unimodal*: $\Delta\text{BIC} = 83.10$); this demonstrates that subjects neither choose the most reliable cue nor choose the visual cue in all bimodal trials.

Discussion

The aim of our experiment was to study the integration of visual and haptic information in the context of human arm movement control. We created a multisensory cued pointing paradigm in which human subjects pointed at an invisible target, presented at an unknown target location. Our subjects were instructed to infer the target position by relying on a visual or a haptic cue that was presented after movement onset during the reach. We measured pointing accuracy as a function of haptic and visual cue onset and compared the measured pointing performance to the predictions of an optimal multisensory decision model. Our model predicts the optimal finger position and the pointing accuracy as a function of the variability introduced by the visual and haptic cues, the overall motor noise and prior assumptions about the distribution of possible target locations.

Our main result is that subjects can rely on visual and haptic cues for online correction of hand movements in a fashion that is close to optimal. No matter whether information about the target position was presented visually by dots scattered around the target position, or haptically by a force pulse whose strength indicated target height, the participants in our experiment reacted to this information midflight during the reach and corrected their movement to minimize the distance to the target. This finding matches results of earlier work by Liu and Todorov (2007). In this experiment by Liu and Todorov, subjects pointed toward a target that occasionally jumped to a new location after movement onset and adjusted their movement online to correct for the induced shift in target position (see also Schmidt 2002; Song and Nakayama 2007, 2008). In these experiments, the target was indicated by a clearly visible stimulus. In contrast, in our experiment, the target was invisible, such that our subjects had to rely on the visual and/or haptic cue to infer the unknown target position. Our results indicate that subjects can solve this task.

Our model makes the assumption that subjects integrate prior information about the distribution of possible target locations with the (multisensory) information about the actual target information provided by the cue. We added this additional constraint based on the observation that all subjects tend to judge the center of the configuration to be more likely than the far ends and therefore, in general, overshoot low targets and undershoot high targets (see e.g. Fig. 4). Thus, though in fact the targets in our experiment were sampled from a boxcar distribution, fitting the Gaussian prior distribution could account for the measured end point distribution. This finding is consistent with previous results that suggest that subjects tend to treat non-Gaussian distributions as if they were Gaussian (Körding and Wolpert 2004a).

It is very likely that subjects used the training trials to adjust this prior. This is in line with previous work in which it was shown that stochastic prior information about target location can be learned implicitly across only a few hundred trials (Baddeley et al. 2003; Seydell et al. 2008; Körding and Wolpert 2004b). However, a simple fit of a Gaussian distribution to the boxcar distribution used in our experiment would have resulted in slightly different parameters than those estimated in our experiment (parameters of an unbiased Gaussian prior: $M = 13.5$ cm and $SD = 4.38$ cm). The fit of the model to the measured end point data yielded a set of estimates (see Table 1) which suggest that subjects rely on a prior that is slightly biased toward higher target positions, reflecting a general tendency to overshoot the target. In addition, the estimated width of this fitted prior distribution was narrower than the default Gaussian prior, indicating a larger relative contribution of the prior when compared to the sensory evidence.

The model comparison of the multisensory decision model with possible alternative models indicates that all models which assume a combination of visual and haptic information in the bimodal conditions (standard model, same motor noise, no motor noise) clearly outperform the models that assume no integration at all (take visual, take best unimodal). All of these models are also better predictors of the performance in bimodal trials when compared to unimodal trials. The multisensory decision model results in the lowest BIC value and predicts performance in unimodal and bimodal trials successfully.

In addition to introducing spatial disparity between visual and haptic cues, we presented the cues sequentially instead of simultaneously in half of the bimodal trials. There is growing evidence that humans can efficiently integrate noisy sensory information over time. For example, it has been shown that humans use Kalman filter type algorithms to estimate the position of their hand (Burge et al. 2008; Saunders and Knill 2004; Wolpert et al. 1995) and integrate sensory noisy information over time for decision making (Gold and Shadlen 2007; Shadlen and Newsome 2001). On the other hand, it has been reported that co-occurrence of

two cues is a crucial factor for crossmodal cue integration (Bertelson and Aschersleben 1998; Lewald et al. 2001; Radeau and Bertelson 1987; Slutsky and Recanzone 2001; Bresciani et al. 2005). In all of these studies, the two cues were interpreted by the participants as belonging together, and the judgment about one of the perceptual signals was biased toward the other one (or a common characteristic, like spatial location, was merged from both sensory modalities).

In our task, we provided cues which conveyed information about the target location which would be available at a later point in time. Thus, there was no need to “integrate” the two cues (i.e., to interpret them to be related to each other) nor to “think” that the two cues belong together (i.e., originate from a common object). Much like in the experiments by Shadlen and colleagues (Gold and Shadlen 2007; Shadlen and Newsome 2001), the cues in our task could be treated as two separate pieces of evidence which needed to be integrated to reach a decision criterion. Interestingly, in our task, performance did not suffer during sequential presentation of visual and haptic cues when compared to simultaneous presentation of both cues. This suggests that in a situation where the correspondence of two cues is obviously neither spatial nor temporal, but simply determined by the fact that the two cues (which originate from different modalities) can both be taken into account to estimate a characteristic of a third object (the target location), these cues are integrated with Bayesian weights.

One might argue that subjects still perceive the sequential presentation as being simultaneous. We used a large stimulus onset asynchrony of 200 ms and it has been reported that subjects are able to perceive visual-haptic asynchrony as soon as the delay is bigger than 45 ms (Vogels 2004). This temporal window of perceived simultaneity is task specific; in a similar task with visual and auditory stimuli, a delay of 100 ms reduced simultaneity perception significantly (Slutzky and Recanzone 2001). However, in these previous studies subjects were instructed to detect asynchrony and, thus, likely paid attention to the timing itself. In our experiment, the asynchrony occurred without any instruction or comment about timing of cue presentation. In addition, none of the subjects in our experiment reported to have noticed any difference in timing, even though we cannot exclude that a larger disparity would produce different results.

What is the reason for cue integration? Obviously the brain should only integrate signals from different modalities if and only if each modality provides information about the same object or source of signals (Körding et al. 2007). In support of this hypothesis, it has been shown that the spatial offset between where an object is seen and where an object is felt affects the integration of sensory information (Gepshtein et al. 2005); similarly, spatial distance between visual and auditory stimuli has been shown to decrease the ventriloquist effect (Wallace et al. 2004).

Causal inference models of cue combination (Körding et al. 2007) provide a conceptual framework that successfully explains behavior in these tasks. Models such as the model by Körding et al. (2007) are built on the assumption that the likelihood of a common cause is inferred online by the nervous system. The result of this computation determines in a single trial whether two cues are integrated, partially integrated or processed separately. The additional assumption that increasing spatial distance decreases the likelihood of a common cause explains a continuous shift from optimal integration to full separation of auditory and visual cue locations with increasing spatial differences between the cues (Körding et al. 2007).

Consistent with the model predictions, perceptual-motor coherence combined with knowledge about a joint causal structure can reduce the effect of spatial disparity. Cue

integration despite spatial disparity has been demonstrated for spatial disparity induced by a mirror (Helbig and Ernst 2007) or by a tool (Takahashi et al. 2009). These findings suggest that there are situations where a causal relationship between two cues is obvious for other reasons and therefore need not be inferred from the cues themselves. In our task, visual and haptic cues do share an obvious common consequence. They both point to the same target position and there is a benefit of integrating all the available information. Specifically, this implies that there is no need for our subjects to infer the likelihood of a common cause on a trial-by-trial basis. Thus, neither spatial nor temporal synchrony is necessary. The key to overcoming spatial disparity between sources of sensory information is therefore any kind of evidence that indicates that the two signals belong together.

The integration of multiple sensory cues is obviously important in many behavioral contexts. For both humans and non-human primates, several multisensory areas have been identified, including the lateral intraparietal area, the parietal reach region, the ventral intraparietal area, the ventrolateral prefrontal cortex, and the superior temporal sulcus (for reviews see Stein and Stanford 2008; Calvert and Thesen 2004). For human subjects, the importance of temporal synchrony for multisensory integration has also been demonstrated by recent EEG work (Calvert 1998; Spence and Squire 2003). From a functional perspective, one requirement for optimal inference is that the reliability of a signal is encoded by the brain. These probability distributions could, for example, be encoded directly (Beck and Pouget 2007; Rao 2004), via probabilistic population codes (Deneve et al. 2001; Ma et al. 2006) or via synaptic integration in single neurons (Deneve 2008a, b). However, with probabilistic population coding, multisensory neurons would be additive or subadditive, but not superadditive (Ma and Pouget 2008). This suggests that superadditivity might not be a strict criterion for neurons involved in multisensory processing though superadditive neurons might play a special role. A brain area likely involved in crossmodal integration of sensory information is the mid-brain superior colliculus of cats (Meredith and Stein 1983). Some of the crossmodal neurons identified in this area were superadditive: that is, they showed a larger response under multimodal stimulation than the simple sum of both unimodal responses (Wallace et al. 1998; Stein and Meredith 1993). This enhanced multimodal response was strongest for temporal coincidence (Meredith and Stein 1983) and spatial proximity (Meredith and Stein 1996).

The growing body of experimental evidence underlines the necessity for computational models that incorporate and test assumptions about the postulated neuronal architecture and neuronal pathways of information processing. In parallel, models of Bayesian Inference are just beginning to be applied to modeling the formation of multisensory perceptual estimates (Ma et al. 2006, 2008; Hossfeld et al. 2007).

Acknowledgments

We thank Nathalie Wahl for help with data collection. Funded by the Deutsche Forschungsgemeinschaft (Emmy-Noether-Programm, TR, 528/1-2, 1-3).

References

- Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*. 2004; 14:257–262. [PubMed: 14761661]
- Baddeley RJ, Ingram HA, Miall RC. System identification applied to a visuomotor task: near-optimal human performance in a noisy changing task. *J Neurosci*. 2003; 23:3066–3075. [PubMed: 12684493]
- Beck JM, Pouget A. Exact inferences in a neural implementation of a hidden Markov model. *Neural Comput*. 2007; 19:1344–1361. [PubMed: 17381269]

- Bertelson P, Aschersleben G. Automatic visual bias of perceived auditory location. *Psychon Bull Rev.* 1998; 5:482–489.
- Bresciani J, Ernst MO, Drewing K, Bouyer G, Maury V, Kheddar A. Feeling what you hear: auditory signals can modulate tactile tap perception. *Exp Brain Res.* 2005; 162:172–180. [PubMed: 15791465]
- Burge J, Ernst MO, Banks MS. The statistical determinants of adaptation rate in human reaching. *J Vis.* 2008; 8(4):1–19. [PubMed: 18484859]
- Calvert GA. Crossmodal identification. *Trends Cogn Sci.* 1998; 2:247–253. [PubMed: 21244923]
- Calvert GA, Thesen T. Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris.* 2004; 98:191–205. [PubMed: 15477032]
- Deneve S. Bayesian spiking neurons I: inference. *Neural Comput.* 2008a; 20:91–117. [PubMed: 18045002]
- Deneve S. Bayesian spiking neurons II: learning. *Neural Comput.* 2008b; 20:118–145. [PubMed: 18045003]
- Deneve S, Latham PE, Pouget A. Efficient computation and cue integration with noisy population codes. *Nat Neurosci.* 2001; 4:826–831. [PubMed: 11477429]
- Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature.* 2002; 415:429–433. [PubMed: 11807554]
- Ernst MO, Bühlhoff HH. Merging the senses into a robust percept. *Trends Cogn Sci.* 2004; 8:162–169. [PubMed: 15050512]
- Gepshtein S, Burge J, Ernst MO, Banks MS. The combination of vision and touch depends on spatial proximity. *J Vis.* 2005; 5(11):1013–1023. [PubMed: 16441199]
- Gold JI, Shadlen MN. The neural basis of decision making. *Annu Rev Neurosci.* 2007; 30:535–574. [PubMed: 17600525]
- Griffiths TL, Tenenbaum JB. Optimal predictions in everyday cognition. *Psychol Sci.* 2006; 17:767–773. [PubMed: 16984293]
- Helbig HB, Ernst MO. Knowledge about a common source can promote visual-haptic integration. *Perception.* 2007; 36:1523–1533. [PubMed: 18265835]
- Hillis JM, Ernst MO, Banks MS, Landy MS. Combining sensory information: mandatory fusion within, but not between, senses. *Science.* 2002; 298:1627–1630. [PubMed: 12446912]
- Hillis JM, Watt SJ, Landy MS, Banks MS. Slant from texture and disparity cues: optimal cue combination. *J Vis.* 2004; 4(12):967–992. [PubMed: 15669906]
- Hospedales T, Cartwright J, Vijayakumar S. Structure inference for Bayesian multisensory perception and tracking. *Proc Int Joint Conf Art Intell (IJCAI '07).* 2007:2122–2128.
- Knill DC. Discrimination of planar surface slant from texture: human and ideal observers compared. *Vis Res.* 1998; 38:1683–1711. [PubMed: 9747503]
- Knill DC, Saunders JA. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vis Res.* 2003; 43:2539–2558. [PubMed: 13129541]
- Körding KP, Wolpert DM. Advances in neural information processing systems 16. In: Thrun, S.; Saul, L.; Schölkopf, B., editors. *Advances in neural information processing systems 16.* MIT Press; Cambridge, MA: 2004a. p. 1327–1334.
- Körding KP, Wolpert DM. Bayesian integration in sensori-motor learning. *Nature.* 2004b; 427:244–247.
- Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. *PLoS ONE.* 2007; 2(9):e943. doi:10.1371/journal.pone.0000943. [PubMed: 17895984]
- Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vis Res.* 1995; 35:389–412. [PubMed: 7892735]
- Lewald J, Ehrenstein WH, Guski R. Spatio-temporal constraints for auditory-visual integration. *Behav Brain Res.* 2001; 121:69–79. [PubMed: 11275285]
- Liu D, Todorov E. Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *J Neurosci.* 2007; 27:9354–9368. [PubMed: 17728449]

- Louw S, Smeets J, Brenner E. Judging surface slant for placing objects: a role for motion parallax. *Exp Brain Res.* 2007; 183:149–158. [PubMed: 17632707]
- Ma WJ, Pouget A. Linking neurons to behavior in multisensory perception: a computational review. *Brain Res.* 2008; 1242:4–12. [PubMed: 18602905]
- Ma WJ, Beck JM, Latham PE, Pouget A. Bayesian inference with probabilistic population codes. *Nat Neurosci.* 2006; 9:1432–1438. [PubMed: 17057707]
- Ma WJ, Beck JM, Pouget A. Spiking networks for Bayesian inference and choice. *Curr Opin Neurobiol.* 2008; 18:217–222. [PubMed: 18678253]
- Meredith MA, Stein BE. Interactions among converging sensory inputs in the superior colliculus. *Science.* 1983; 221:389–391. [PubMed: 6867718]
- Meredith MA, Stein BE. Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol.* 1996; 75:1843–1857. [PubMed: 8734584]
- Radeau M, Bertelson P. Auditory-visual interaction and the timing of inputs. *Psychol Res.* 1987; 49:17–22. [PubMed: 3615744]
- Rao RPN. Bayesian computation in recurrent neural circuits. *Neural Comput.* 2004; 16:1–38. [PubMed: 15006021]
- Saunders JA, Knill DC. Visual feedback control of hand movements. *J Neurosci.* 2004; 24:3223–3234. [PubMed: 15056701]
- Schmidt T. The finger in flight: Real-time motor control by visually masked color stimuli. *Psychol Sci.* 2002; 13:112–117. [PubMed: 11933993]
- Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978; 6:461–464.
- Serwe S, Drewing K, Trommershauser J. Combination of noisy directional visual and proprioceptive information. *J Vis.* 2009; 9:1–14. [PubMed: 19757906]
- Seydell A, McCann BC, Trommershauser J, Knill DC. Learning stochastic reward distributions in a speeded pointing task. *J Neurosci.* 2008; 28:4356–4367. [PubMed: 18434514]
- Shadlen MN, Newsome WT. Neural basis of a perceptual decision in the parietal cortex (Area LIP) of the Rhesus monkey. *J Neurophysiol.* 2001; 86:1916–1936. [PubMed: 11600651]
- Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport.* 2001; 12:7–10. [PubMed: 11201094]
- Song J, Nakayama K. Automatic adjustment of visuomotor readiness. *J Vis.* 2007; 7(5):1–9.
- Song J, Nakayama K. Target selection in visual search as revealed by movement trajectories. *Vis Res.* 2008; 48:853–861. [PubMed: 18262583]
- Sousa R, Brenner E, Smeets JB. Slant cue are combined early in visual processing: Evidence from visual search. *Vis Res.* 2009; 49:257–261. [PubMed: 19027776]
- Spence C, Squire S. Multisensory integration: maintaining the perception of synchrony. *Curr Biol.* 2003; 13:R519–R521. [PubMed: 12842029]
- Stein, BE.; Meredith, MA. *The merging of the senses.* MIT Press; Cambridge, MA: 1993.
- Stein BE, Stanford TR. Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci.* 2008; 9:255–266. [PubMed: 18354398]
- Takahashi C, Diedrichsen J, Watt SJ. Integration of vision and haptics during tool use. *J Vis.* 2009; 9(6):1–15. [PubMed: 19761294]
- Tassinari H, Hudson T, Landy MS. Combining priors and noisy visual cues in a rapid pointing task. *J Neurosci.* 2006; 26:10154–10163. [PubMed: 17021171]
- van Beers RJ, Sittig A, Denier van der Gon J. Integration of proprioceptive and visual position-information: an experimentally supported model. *J Neurophysiol.* 1999; 81:1355–1364. [PubMed: 10085361]
- van Beers RJ, Wolpert DM, Haggard P. When feeling is more important than seeing in sensorimotor adaptation. *Curr Biol.* 2002; 12:834–837. [PubMed: 12015120]
- Vogels IMLC. Detection of temporal delays in visual-haptic interfaces. *Hum Factors.* 2004; 46:118–134. [PubMed: 15151159]
- Wallace MT, Meredith MA, Stein BE. Multisensory integration in the superior colliculus of the alert cat. *J Neurophysiol.* 1998; 80:1006–1010. [PubMed: 9705489]

- Wallace MW, Roberson G, Hairston W, Stein BE, Vaughan J, Schirillo J. Unifying multisensory signals across time and space. *Exp Brain Res*. 2004; 158:252–258. [PubMed: 15112119]
- Wolpert DM, Ghahramani Z, Jordan M. An internal model for sensorimotor integration. *Science*. 1995; 269:1880–1882. [PubMed: 7569931]

Fig. 1.

Task. **a** Subjects had to point across a vertical target line 32 cm to the right of a fixed start position (**a**). They had to hit an invisible target dot on that line that varied in height (between 6 and 21 cm above start position). Information about the correct target height in a given trial was presented during the movement. This information was either haptic or visual and was presented either early (50 ms after movement onset) or late (250 ms after movement onset). **b** Example of a haptic trial. The *solid line* represents one sample trajectory including the small deviation caused by the force pulse. The force pulse itself (*dashed line*) is superimposed on a different scale, such that the *x*-axis matches fingertip position and *y*-axis represents the force pulse strength at that time (total duration 100 ms). Force pulse strength is scaled to match the corresponding target height. **c** Example of a visual trial: *three filled dots* were sampled from a distribution around the invisible target location (*unfilled circle*) and briefly flashed (100 ms)

Fig. 2.

Standard deviation of pointing error during the unimodal training sessions. Subjects first ran two blocks of 80 trials each in the visual training condition, then two blocks of 80 trials each in the haptic training condition. In the visual (and haptic, respectively) training conditions, both early and late trials were randomly intermixed, but modalities were blocked. For each bin of size 20 trials, the standard deviation of errors was computed (error = distance between hit position and target position). Each *graph* shows the average and the standard error of the mean across 8 subjects. Subjects learned both the visual task and the haptic task almost immediately and then did not improve further during the training session

Fig. 3.

Standard deviation of pointing error for the 8 experimental conditions. Before subjects started to run the main compensated experimental session, they ran two training blocks (80 trials each) that were exactly the same as in the later experiment. The experiment itself consisted of 8 blocks of 80 trials each. All conditions were randomly intermixed. For bin sizes of 20 trials, we computed performance across the experiment. Performance was measured by estimating the standard deviation of pointing errors. We averaged the data across 8 subjects and the *bars* indicate the standard error of the mean. The first bin corresponds to the two training blocks (160 trials total, i.e. 20 trials per condition), the bins labeled 1–4 are the experimental sessions (640 trials total, 20 trials per bin). Subjects' performance did not change across the different blocks

Fig. 4.

Mean trajectories for one example subject (RZ) for high, middle and low targets. From top to bottom, the *rows* represent the haptic only condition, visual only condition, visual-haptic simultaneous presentation, and visual-haptic sequential presentation condition, respectively. *Solid lines* represent early presentation (50 ms; *HE, VE, BE*) or respectively haptic early presentation in the sequential case (*HF*), *dashed lines* represent late presentation (250 ms; *HL, VL, BL*) or respectively haptic late presentation in the sequential case (*VF*). The *left part* in each *row* represents the time relative to movement onset. The *right part* of each *row* represents the time relative to reaching the target line

Fig. 5.

Individual model predictions for standard deviation of pointing error. *Each plot* shows the individual standard deviation of errors for one subject. *Vertical lines* indicate 95% confidence intervals for the estimator of the standard deviation. The *solid line* represents the simulated standard deviation of errors using the fitted parameters. Since every single trial was simulated 10,000 times, the confidence interval for the standard deviation of the model is negligibly small and therefore not plotted

Fig. 6. Predicted versus observed standard deviation of pointing error averaged across subjects. *Gray bars* represent the mean performance for each condition averaged across eight subjects; *white bars* represent the predictions by the standard model. Model predictions were first computed individually based on 10,000 simulations per trial and then averaged across subjects. *Solid lines* represent the standard error of the mean

Table 1

Relative BIC values, number of parameters k per model, and mean parameter values [in cm] across subjects for the different models

Model	ΔBIC	k	σ_{HE}	σ_{HL}	σ_V	P	σ_P	σ_M
Standard model	0	6	2.20	2.89	1.86	14.08	3.14	1.22
Same motor noise	15.29	6	2.55	3.19	1.87	14.23	3.23	0.55
No motor noise	11.68	5	2.96	3.79	2.25	14.30	3.85	–
Take visual	38.40	5	2.69	3.50	2.10	14.23	3.68	–
Take best unimodal	83.10	5	2.71	2.71	2.58	14.18	3.81	–

The relative BIC value is the difference between each model's BIC value and the BIC value of the standard model ($\Delta BIC = BIC - BIC_{\text{standard}}$). Any positive ΔBIC value suggests that the model's BIC value is higher than the BIC of the standard model and that therefore the standard model is to be preferred