

The role of attention on the integration of visual and inertial cues

Daniel R. Berger · Heinrich H. Bühlhoff

Received: 1 October 2008 / Accepted: 3 March 2009 / Published online: 7 April 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract The extent to which attending to one stimulus while ignoring another influences the integration of visual and inertial (vestibular, somatosensory, proprioceptive) stimuli is currently unknown. It is also unclear how cue integration is affected by an awareness of cue conflicts. We investigated these questions using a turn-reproduction paradigm, where participants were seated on a motion platform equipped with a projection screen and were asked to actively return a combined visual and inertial whole-body rotation around an earth-vertical axis. By introducing cue conflicts during the active return and asking the participants whether they had noticed a cue conflict, we measured the influence of each cue on the response. We found that the task instruction had a significant effect on cue weighting in the response, with a higher weight assigned to the attended modality, only when participants noticed the cue conflict. This suggests that participants used task-induced attention to reduce the influence of stimuli that conflict with the task instructions.

Keywords Self-motion perception · Perception of angular displacement · Multisensory integration · Attention · Robust integration · Cue conflicts

Introduction

Humans perceive passive self-motion based on information provided by multiple senses; the most important being the visual sense and senses of inertial motion (the vestibular

system and somatosensation/proprioception). To estimate movements of their own body in space, humans can combine information from these different senses through the process of multimodal integration. This enables them to generate a more accurate and reliable internal representation of the movements, which can then be used for optimal behavior (Ernst and Bühlhoff 2004).

A possible role of attention in multisensory integration

It is currently a controversial issue whether multimodal integration is an automatic process that happens early in the sensory processing hierarchy in the brain and thus, cannot be voluntarily manipulated (i.e., is impenetrable to cognition) (Driver 1996; Bertelson et al. 2000; Helbig and Ernst 2008). Alternatively, it may be possible that top-down effects involving higher cognitive processes can influence multimodal integration (Welch and Warren 1980; Bertelson and Radeau 1981; Andersen et al. 2004; Mozolic et al. 2008), although, the conditions under which top-down control can exert an influence on the process of multimodal integration remain unspecified.

Mathematically, the optimal unbiased estimate of a value that is measured independently by two or more senses is obtained if the single sensory estimates are combined by using a weighted arithmetic mean, with individual weights determined by their respective reliabilities (Maximum Likelihood Estimate, MLE; Ernst and Bühlhoff 2004). Assuming such a model of multisensory integration, we call these weights “cue weights”, with the information provided by each sensory modality being called a “cue”. If the brain combines cues by using MLE, multimodal integration of stimuli should be immune to cognitive influences like task-specified attention. However, it is often assumed that there can be additional beliefs

D. R. Berger (✉) · H. H. Bühlhoff
MPI for Biological Cybernetics, Tübingen, Germany
e-mail: dberger@mit.edu

H. H. Bühlhoff
e-mail: heinrich.buelthoff@tuebingen.mpg.de

about the presented signal which can bias the resulting estimate, for example, if they are incorporated into the integration process as “priors” in a Bayesian framework (Yuille and Bülthoff 1996).

Unconditional multisensory integration may in some cases not be the best way to deal with information from multiple sensory modalities. Sensory signals should only be integrated if they belong to the same source in the world. This is the so-called “assignment problem” (Pouget et al. 2004) or “correspondence problem” (Ernst 2007) of multisensory integration. In order to generate best estimates of sensed variables, the brain has to decide first whether to integrate two sensory cues. To do so, it has to determine whether the two signals have the same source, which might involve higher cognitive functions—for example in the superior colliculus of cats and monkeys, neurons with multisensory response properties emerge only after development of afferent projections from association cortex (Wallace 2004). Including this decision of unity in the integration process makes the estimation more robust under sensory conflict (“robust statistical estimation”, Landy et al. 1995). The belief of whether two cues belong together can also be modeled as a Bayesian prior (Ernst 2006; Körding et al. 2007). Since the awareness of a conflict between two cues could have an effect on the belief of whether the two cues belong together, this might also influence the proposed prior. Also, the instruction to use one cue for the response and to ignore the other could in principle influence the integration.

A problem similar to the assignment or correspondence problem is described in vision, called the “binding problem”. Our visual environment usually contains many objects, and each of these has multiple visual features. Therefore, perceptual grouping, or “binding”, of these features into objects has to be solved in visual object perception. It has been proposed that *attention* could be the mechanism through which a decision is reached as to whether the features belong together. The “feature-integration theory of attention” (Treisman and Gelade 1980) assumes that focused attention and/or other top-down processing is necessary to identify objects. Within this framework it is hypothesized that the brain would exclusively attend to all the features of one object at a time, so that all activations in the distributed visual system of the brain would then belong to the same attended object, which makes binding of these features trivial.

There is some experimental evidence that voluntary attention can indeed influence the grouping of visual features in perception. Good examples are bistable percepts, like the face-vase illusion, the Necker cube or binocular rivalry; the perception of which can, to a certain extent, be influenced by volition (Meng and Tong 2004; Paffen et al. 2006). Top-down attention can also influence the

interpretation of visual slope if monocular and binocular cues are set in conflict (van Ee et al. 2002, 2003).

Although the binding problem has been mostly addressed in the visual system, it must also be solved within other modalities and even between different modalities. Therefore, binding is not restricted to visual feature binding, but can also be extended to the case of crossmodal binding, as has recently been addressed by Vatakis and Spence (2007) and Senkowski et al. (2008). Possibly the rules of perceptual grouping also apply for multimodal stimuli (Harrar and Harris 2007), suggesting that voluntary attention could also play a role in solving the assignment/correspondence problem in multisensory integration. It has been known for some time that if discrepancies between cues occur, their integration is influenced by cognitive factors such as awareness of the discrepancy, assumption of unity of the stimuli, “compellingness” of the stimulus situation, and type of response (see the review of Welch and Warren 1980). For example, humans who are required to point towards an auditory probe (that is accompanied by a spatially incongruent visual probe) are more likely to be influenced by the irrelevant visual probe if they have reason to assume that the two probes share the same location (Bertelson and Radeau 1981).

Also some crossmodal illusions are affected by task-defined attention. Andersen et al. (2004) showed that the crossmodal auditory-visual effects in the Shams illusion (Shams et al. 2000) depend strongly on the task and that auditory and visual events are not mandatorily fused into a “unified percept”. Also the McGurk effect (McGurk and MacDonald 1976) is affected by attention (Alsus et al. 2005, 2007), again challenging strong pre-attentive accounts of audiovisual integration. There are also several ERP studies which show that there is an effect of attention on the integration of auditory and visual signals (Li et al. 2007; Talsma and Woldorff 2005).

In particular, like in visual feature binding, attending to two information sources at the same time could promote their integration, whereas focusing attention to one of them and ignoring the other could inhibit the integration of the two cues (Talsma et al. 2007; Mozolic et al. 2008). Top-down attention would therefore be a means by which cognitive assumptions about whether two stimuli have the same source could be used to make multisensory integration more robust in situations where sensory conflicts occur. The same mechanism could influence stimulus processing to solve a given task. The task would define which sensory sources to use for a response and which to ignore, and top-down attention would impose these task-defined constraints on the sensory processing to optimize the estimation of behaviorally relevant signals.

The term “attention” is used for a large variety of different information selection processes in the brain

(Corbetta and Shulman 2002). First, there can be different sources inducing attention, which can be either the presented stimuli (“bottom-up”, “stimulus-driven” attention or “saliency”), or internal, possibly cognitive processes (“goal-directed” or “top-down” attention), which might be under voluntary control. Also, attention can select information with respect to different stimulus parameters, like spatial location (“spatial attention”), the type of information (feature-based or object-based attention; for example a specific color, shape, etc.) or a sensory modality.

In this study we are interested specifically in *top-down attention* to a *sensory modality*.

Experimental questions and hypotheses

In this study we investigated whether there is a mandatory, complete integration of signals from the visual and the inertial (vestibular/somatosensory/proprioceptive) senses for self-motion perception, or whether participants can, when instructed to, use information from one modality for their response while ignoring information from the other modality. We thus investigated the effect of *voluntary top-down attention to a sensory modality, induced by task instructions*, on multimodal integration of the attended modality with an unattended one. Additionally, we tested whether the influence of any task-induced attention on the response differs as a function of whether the participant notices a cue conflict or not.

Since we manipulated the focus of attention by using a task defined by the instruction to attend to one modality and ignore the other while responding, we could not distinguish whether the effects we find are implicitly induced by the task or imposed by the instructions. This is reflected in the term “task-induced attention” used throughout this paper, meaning that the effects we find depend on the instruction to attend to one or the other modality.

We investigated these questions by means of a turn reproduction task, where passively presented rotations of a participant about the earth-vertical axis had to be actively rotated back to the origin by the participant. During the return rotation, we introduced “gain factors” between visual and platform rotations. Specifically, the visual scene was rotated slower or faster than the platform, with a constant factor per trial. Thus, different turns had to be performed to turn back the platform or the visual scene to its original orientation. Participants were instructed to use one of the two cues to complete the return while ignoring the other cue. If participants put a high weight on the visual cue, it would be expected that they would return the visual scene more or less correctly to the original orientation while under- or overturning the platform. The opposite would be true if they put a high weight on the inertial cue.

From their return angle we can then deduce how much weight they put on each of the cues to perform their return.

There are several possible ways in which multisensory integration and attention could relate to each other in the brain. One possibility is that visual and inertial sensory information for the estimation of a turn angle are automatically integrated into a unitary combined percept of the rotation, and that this integration is immune to task-instructed voluntary attention to one of the modalities. If participants only have access to this combined percept and not to estimates of the visual rotation and the inertial rotation separately, then the cue weights in the responses should not be affected by the instructions. If, on the other hand, we find that instructing participants to use one of the modalities for their response and to ignore the other modality has an effect on the cue weights in the responses, this would argue against an exclusive automatic integration of visual and inertial cues in the perception of these rotations.

If participants respond only according to the instructed modality, the results would suggest that participants have access to independent rotation estimates from individual modalities and are able to select one of them for their response, according to task instructions. If, however, the task instruction has a significant, but incomplete effect on the cue weights, this would indicate that the responses are based on combined information from both modalities and that task-induced attention to one modality influences the relative cue weighting. Such a response pattern would suggest that multisensory integration processes are occurring, and that they are influenced by the task to attend to one or the other modality while returning.

For optimal robust integration, it is necessary that the information from different sensory modalities is only integrated if the information provided by the different modalities belongs to the same entity (is not conflicting). Therefore, we also investigated whether cue weights and the influence of task-induced attention to a particular modality differ as a function of cue conflicts. In order to determine in which trials participants were aware of the cue conflict they were asked after each trial to rate whether they had detected a conflict. Based on these ratings, the trials were then divided into “cue conflict” and “no cue conflict” conditions. If we find an effect of task-induced attention only in the “cue conflict” condition, this would be evidence for “robust” multisensory integration, in which task-induced attention can be used to resolve the conflict (to prefer the to-be-attended-modality over the to-be-ignored one).

Integration of cues from two modalities should lead to an increase of the reliability of the sensory estimate, and in turn to a reduction of the variance of the responses in

comparison to the responses if only one cue is available (Ernst and Bühlhoff 2004). To test this, we ran all participants also in single-cue conditions with only visual information and only inertial information. In the “visual-only” condition the platform was off during presentation and return, so participants had to solve the task using only visual information. In the “body-only” condition, rotations were presented in darkness and had to be returned in darkness.

Methods

Twenty volunteers (10 females), aged 20–33 years (mean 24), participated in this experiment. They were paid for participation. None of them had a known history of neurological or vestibular disorders. All participants gave their informed consent prior to the experiment according to the declaration of Helsinki. They were asked to read an instruction sheet and were additionally instructed verbally and by sample trials before the experiment to make sure that they had understood the task.

Participants were seated on a Stewart motion platform equipped with a flat projection screen (Fig. 1). In each trial, they first experienced a passive rotation around an earth-vertical axis (“yaw” rotation, change of heading). They then had to return actively back to the origin, using a joystick. We measured their return angle in four different conditions: visual-only rotations (rotation of the visual scene but stationary platform during passive and during active rotations, “visual-only”, VO), platform-only rotations (passive and active whole-body rotations in darkness by movement of the Stewart platform, “body-only”, BO), and combined visual and platform rotations, in which during passive and active rotations both the visual scene and the platform rotated at the same time (“combined-cue conditions”).

For the return in the combined-cue conditions, participants were instructed to either use the visual cue and to ignore the inertial cue (“visual attention” condition, VA), or to use the inertial cue and to ignore the visual cue (“body attention” condition, BA). These tasks were introduced to focus voluntary attention to either visual or to body cues during self-motion perception. Participants were told that during the passive presentation of the rotation, both visual and platform rotations would always be concordant, but that there might be cue discrepancies during their active return. Thus we instructed them to attend to only one cue in particular during their active return, whereas they could choose to attend to one cue or both during the passive presentation. During the VA condition, participants were instructed to reproduce the rotation of the visual scene by turning within the visual scene back to the

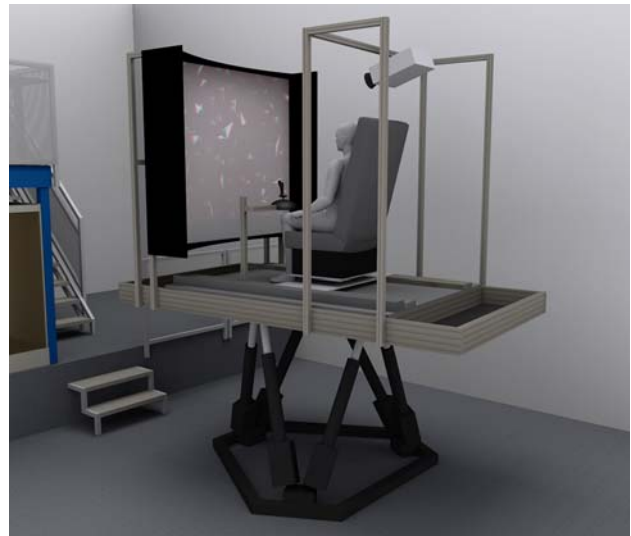


Fig. 1 Rendering of the Stewart motion platform with projection screen, showing the random anaglyph triangle scene used in this experiment. During the experiment, the *top part* of the platform was completely enclosed by a *black* curtain so that subjects could not see the surrounding laboratory

visual origin. During this condition the platform was also rotated; however, participants were told to ignore the platform movement. During the BA condition, participants were instructed to reproduce the body rotation by turning the platform back to its origin. In this case, even though visual rotations were also presented, participants were told to ignore them. Note that the only difference between the VA and BA conditions was the task, which influenced which modality had to be attended and which ignored during the active return. The presented stimuli were the same in both conditions.

Figure 2 shows the rotation trajectories during an example trial in more detail. Each trial started with a “Presentation phase”, during which participants were passively rotated for either 10°, 15° or 25° in 3 s, with a raised cosine profile. These passive turns were always 3 s long, independent of the turned angle, as a way of preventing subjects from estimating the size of a rotation by using its duration. Thus, larger rotations also involved higher velocities and accelerations. In the “combined-cues” conditions, these passively presented visual and platform turns were always consistent (i.e. they had the same angle, velocity profile, etc.).

After the presentation phase, a short delay of 1 s was inserted before the participant was allowed to turn back. During the delay the auditory instruction to “turn back the visual scene” or “turn back the platform” was played back in the headphones to remind the participant of the current task condition. We kept the delay short to prevent decay of the memory of the presented rotation.

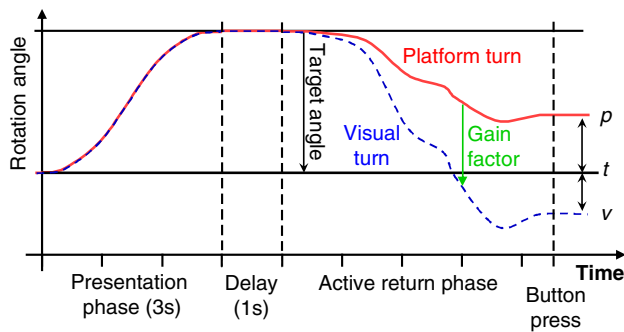


Fig. 2 Schematic representation of the rotation trajectories in one combined-cue trial. The visual weight w_v can be computed from the response as $w_v = (t - p)/(v - p)$

Subsequently participants were given control of the platform and/or visual rotation, and they rotated back to the original orientation they experienced at the start of the trial (Fig. 2, “Active return phase”). In the single-cue conditions the mean and variance of their active return angle for the different presented rotation angles were measured.

During the active return in the combined-cue conditions, one of five gain factors was introduced between visual and platform rotation (0.35, 0.707, 1.0, 1.41, or 2.85). This meant that the visual scene was rotated at either faster or slower rate than the platform, with a constant factor, thus resulting in a visual rotation angle that was larger or smaller than the platform rotation angle (Fig. 2, “Active return phase”). For all gain factors $\neq 1$ we could compute the relative weighting of visual and inertial cues that participants used to compare the passive and active rotations in the trial. Given a passively presented rotation angle t , an actively performed platform rotation p and an actively performed visual rotation v (see Fig. 2), we computed the visual weight w_v to be $w_v = (t - p)/(v - p)$.

The visual weight describes how much the participant relied on visual information for the return. For gain factors $\neq 1$, participants cannot return the platform and the visual scene correctly at the same time. For example, if the gain factor during the return is 2.85, the visual scene rotates 2.85 times as fast as the platform. Therefore, to turn the platform back correctly, participants would have to turn the visual scene more than twice as far as during the presentation phase. If they would turn the visual scene correctly, they would have to turn the platform only about 35% as far as during the presentation phase. Consequently, participants must find a compromise. Either they turn the platform correctly, or the visual scene correctly, or they do something in between. The visual weight w_v describes the influence of the visual rotation on the response. A visual weight of $w_v = 1$ indicates that participants matched the visual return rotation perfectly to the presented rotation and

ignored the platform rotation completely during the return. Conversely, a visual weight of $w_v = 0$ indicates that participants matched the platform return rotation to the presented rotation and ignored the visual rotation during the return.

Participants were explicitly instructed before the combined-cue conditions that the rotation angles of visual and platform rotations during active returns could be different and that they should pay attention to one modality and ignore the other. They were told that when, for example, their task was to attend to the platform rotation, the concurrent visual rotation during their active returns could be different and should be ignored (but without closing the eyes). After rotating back in the combined-cue conditions they had to indicate with a button press whether they had experienced the visual rotation as being faster, the platform rotation as being faster or that they had not noticed a cue conflict while turning back. Participants did not receive any feedback about the accuracy of their return angle, nor about the correctness of their button presses.

VO, BO, VA, and BA conditions were presented in separate blocks. The VO and BO conditions were always measured first because they were easier to handle for the participant. (In conditions VO and BO, participants only had to turn back to the origin by using the joystick. They did not have to ignore one of the two modalities while returning, and they also did not have to report cue conflicts. This made the overall task easier.) Apart from that, the order of conditions was counterbalanced across participants. Each of the combinations of 5 gain factors and 3 target turning angles was measured 8 times, resulting in 120 trials for each block (480 trials per participant).

The order of trials in each block was randomized, but within the dynamic constraints defined for each trial by the current position of the platform within its yaw rotation range ($-55^\circ \dots 55^\circ$). As subjects turned the platform actively, it was impossible to predict the end position of the platform after a trial. To minimize the number of platform repositionings needed during the experiment, the next trial to be presented was selected online depending on the current position of the platform. The trial selection algorithm also balanced the conditions presented during the experiment as much as possible, so that not all small rotations would be presented at the beginning of the experimental block. Only during circumstances under which none of the remaining trials could be presented was the platform repositioned.

Participants saw a visual scene on a projection screen located in front of them (86° horizontal \times 63° vertical, viewing distance 0.65 m). A 3D visual scene was shown by using red-cyan viewers. It consisted of a field of random, limited-lifetime triangles in anaglyphic 3D. We used

random triangles instead of random dots to facilitate stereo fusion. Approximately 150 triangles were visible at the same time, at distances between the screen distance and up to 2–3 m. Triangle brightness was attenuated with distance to enhance the impression of depth. Triangle lifetime was 1.2 s. We chose this rather long interval to reduce motion in the image, so that participants could accept the scene as stable and not moving by itself. The limited lifetime was necessary so that participants could not solve the task of turning back visually by remembering a static pattern of triangles, but had to rely on visual motion cues. A fixation cross at eye height was located approximately 10 cm in front of the screen. Participants were told to fixate on the fixation cross during all turns.

We used a 3D visual stimulus to increase the visual immersion for the participants. It has been shown that the impression of self-motion (vection) is more compelling when stereoscopic cues are used (Palmisano 2002) and if the moving scene is farther away from the observer than an attended object (Nakamura and Shimojo 1999; Kitazaki and Sato 2003). In our case both the frame of the screen and the fixation cross were stationary with respect to the observer and closer to the observer than the moving 3D scene. The rotation did not introduce any motion parallax in the image, as the camera was turned around the optic center of the camera projection.

Participants used a joystick that was mounted in front of them to control active rotations. Deflecting the joystick sideways controlled angular rotation velocity such that, the more it was pushed sideways, the faster the platform and scene rotated. When the joystick was centered the rotation stopped and when released, the joystick re-centered automatically. The translation of joystick deflection to rotation speed was randomized for each trial by using a translation factor between 1 and 2 to prevent motor learning effects. This meant that for higher translation factors a small joystick tilt caused a fast rotation, whereas for lower translation factors the joystick had to be tilted much more for the same effect. Participants could use corrections if they thought that they had turned too far.

Participants wore noise-cancellation headphones playing noise (a mixture of recorded sounds of the platform itself and a river) to mask noises of the motion platform. Shakers agitating the seat and foot plate were used so that the participant could not sense vibrations of the motion platform legs.

Results

All statistical tests were made using repeated-measures analyses of variance (ANOVA). Post hoc comparisons

using Neuman–Keuls tests ($P < 0.05$) were performed when necessary.

Performance in the single-cue conditions

The mean return errors for every participant and condition in the single-cue conditions BO (body-only or platform-only; whole-body rotations in darkness) and VO (visual-only; rotations of the visual scene while the platform was off) are shown in Fig. 3. They were analyzed using a 3×2 [target angle (10° , 15° , 25°) \times cueing condition (BO, VO)] ANOVA. The ANOVA revealed a main effect of target angle [$F(2,38) = 88.5$, $P < 0.001$], the return errors being significantly different between the three target angle levels (see Fig. 3). On the other hand, there was no significant effect of the cueing condition. There was a significant interaction between cueing condition and target angle [$F(2,38) = 9.53$, $P < 0.001$].

These results show that participants could solve the task using either of the modalities alone, and were more accurate in the VO than the BO condition.

Effect of attending to a modality in the combined-cue conditions

In the combined-cue conditions VA (“visual attention”) and BA (“body attention”), the participant could use both visual and inertial sensory information to estimate the size of the rotations. We were interested in the relative weight that the visual rotation and the inertial rotation had on the return angle for different target rotation magnitudes. To compute cue weights and the tendency to over- or underestimate rotations during the active return, we fitted linear functions to the measured rotation angles over gain factors, individually for different participants, target angles, and attention tasks (VA, BA). Two parameters were fitted: a

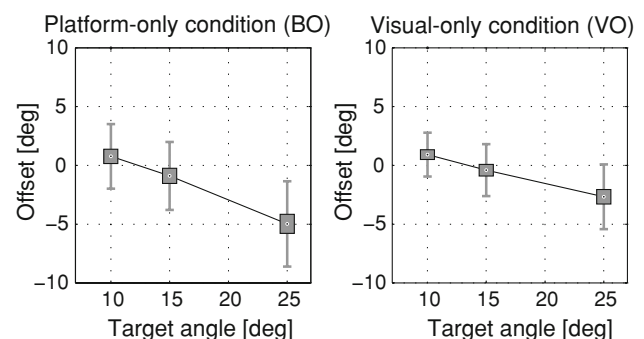


Fig. 3 Responses in the single-cue conditions, shown as response offset (or “response error”; returned angle minus presented rotation angle). Values above zero indicate that participants turned back too far, whereas values below zero indicate that participants stopped short when turning back. Error bars represent standard deviations (*whiskers*) and standard errors (*bars*) across all 20 participants

weighting factor between the visual and the inertial target rotation expressed as visual weight w_v (the inertial weight being $1 - w_v$), and a constant offset c . The offset c is equivalent to the “response offset” found in the single-cue conditions, as shown in Fig. 3. Any tendency to over- or under-rotate for a given target rotation angle independent of the gain factor will show up in c .

For a given presented target angle α , gain factors g_i , visual target rotation angle $t_v(\alpha, g_i) = \alpha \cdot g_i$, body target rotation angle $t_b(\alpha, g_i) = \alpha$ and response angle r , the visual weight w_v and the offset c were derived by minimizing the error e ,

$$e = \sum_i ((w_v \cdot t_v(\alpha, g_i) + (1 - w_v) \cdot t_b(\alpha, g_i) + c) - r(\alpha, g_i))^2.$$

We then computed two separate 3×2 [target angle (10° , 15° , 25°) \times attention condition (VA, BA)] ANOVAs, for response offsets and visual weights, respectively.

Figure 4 shows the resulting data of all participants. The response offset was significantly affected by the target angle [$F(2,38) = 96.1$, $P < 0.001$], participants returning small rotations quite accurately but larger rotations not far enough (Fig. 4, left). There was also a main effect of the attention condition. The participants returned farther when the platform was attended than when the visual scene was attended [$F(1,19) = 5.2$, $P < 0.05$]. The interaction between target angle and attention condition was not significant for response offsets [$F(2,38) = 2.1$, $P > 0.05$].

The cue weights also depended on the target angle and on the attention task. The visual weight was significantly higher for small rotations than for large rotations [$F(2,38) = 40.3$, $P < 0.001$], the visual weights differing significantly between all pairs of target angles. We also found a main effect of attention. The visual weight was higher if vision was to be attended, and lower if the inertial rotation was to be attended. That said, the ignored modality still had a strong influence on the responses. The difference was significant [$F(1,19) = 19.4$, $P < 0.001$]. Attending to

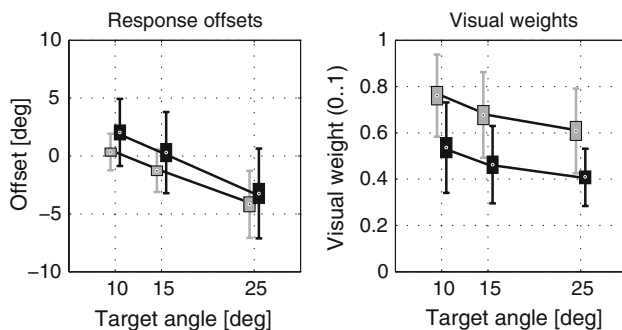


Fig. 4 Offsets and visual weights for different conditions and target angles. *Black* attend to physical rotation (BA), *gray* attend to visual rotation (VA). *Error bars* and *whiskers* show standard error and standard deviation over participants

the visual rotation thus increased the visual weight in the response, and attending to the physical rotation decreased it: The mean visual weight over all participants was between 0.76 (for 10° rotations) and 0.61 (for 25° rotations) in the VA condition, and between 0.53 and 0.41 in the BA condition. The interaction between target angle and attention condition was also not significant for cue weights [$F(2,38) = 0.3$, $P > 0.05$].

Even though attending to one modality gave that modality a higher weight in the response, participants could not ignore the other modality completely in either case. The cross-modal influence of vision in the BA condition (visual weights 0.53–0.41) is higher than the cross-modal influence of the inertial rotation in the VA condition (inertial weights 0.24–0.39). We compared these weights by using a separate 3×2 (target angle, attention condition) ANOVA and found that the difference is significant [$F(1,19) = 7.48$, $P < 0.05$].

Do participants respond differently when they notice cue-conflicts than when they do not?

After each active return, participants indicated whether they thought the visual rotation had been faster, the platform rotation had been faster, or whether they had not detected a difference between the visual rotation and the platform rotation. Figure 5 shows that participants were, in general, not very accurate at this task, and that they responded similarly in VA and BA conditions. The discrimination performance is, however, in agreement with a study by Jaekl et al. (2005), in which participants detected visual-vestibular conflicts during rotations only for gains below approximately 0.8 and above approximately 1.4.

These responses enabled us to split the data into trials in which participants reported that they did not notice a cue conflict while returning and trials in which they reported that they did notice a conflict and investigate the effect of the instruction to attend to one of the modalities on the responses independently in the two cases.

Splitting the data reduced the number of data points per condition, and poses the problem that we no longer have complete data sets for all participants and conditions. Typically, participants responded more often “conflict detected” during trials with gain factors far from 1, and more often “no conflict detected” during trials with gain factors close to 1 (see Fig. 5). Some participants have such a strong bias to respond “conflict detected” or “no conflict detected”, that there are very few or no data points in some conditions. For those participants with enough data points for all fits, we fit linear functions with cue weights and offset to the data as described in the previous section, independently for “no conflict noticed” and “conflict noticed” trials. For this analysis we required that all

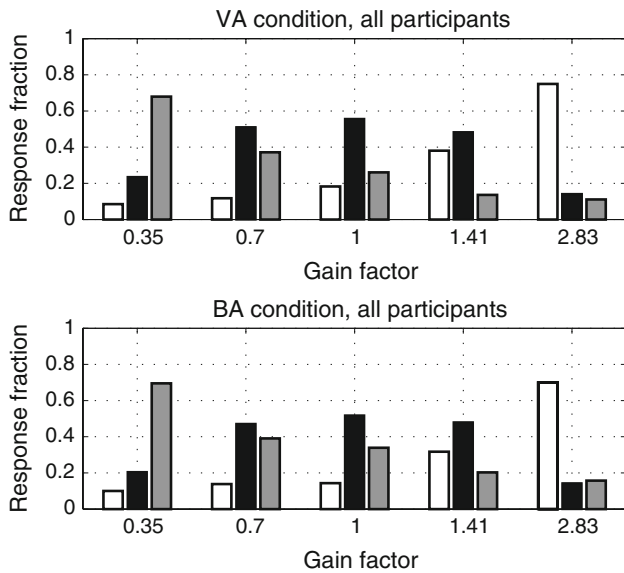


Fig. 5 Histograms of perceiving cue conflicts for different attention conditions and gain factors (visual rotation divided by platform rotation), over all 20 participants. A gain factor of one means that the visual rotation and platform rotation were equivalent. *White* visual rotation was perceived as faster, *black* no conflict between visual and platform rotation was noticed, *gray* platform rotation was perceived as faster

function fits should be based on at least ten data points. 13 of the 20 participants (7 female and 6 male) reached this criterion.

Results for the visual weights as a function of conflict detection are shown in Fig. 6. Offsets are shown in Fig. 7 and indicate that they did not differ significantly as a function of conflict detection and are not analyzed further. We analyzed the visual weights using a $3 \times 2 \times 2$ ANOVA with target angle (10°, 15°, 25°), attention condition (VA, BA), and conflict awareness (no conflict noticed, conflict noticed) as within-subject factors. The interaction of conflict awareness and target angle, and the triple interaction of conflict awareness, target angle, and attention condition were not significant. However, there was a significant interaction between the attention condition and conflict awareness [$F(1,12) = 7.01, P < 0.05$]. Post hoc analysis showed that visual weights in the BA condition were significantly lower if conflicts were noticed compared to both VA conditions. Analysis of the data for “conflict noticed” and “conflict not noticed” in separate ANOVAs showed that the effect of the attention condition on the responses was significant when cue conflicts were noticed [$F(1,12) = 20.70, P = 0.001$], but not when they were not noticed [$F(1,12) = 2.9, P > 0.05$]. This shows that the influence of attention on the cue weights was stronger when cue conflicts were noticed than when they were not noticed.

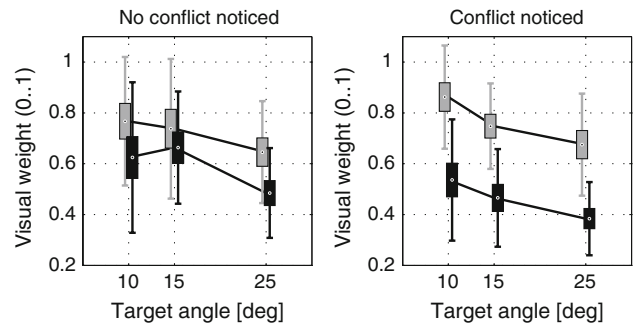


Fig. 6 Visual weights as a function of the detection of conflicts, target angle and attention condition, for 13 participants. *Gray* attend to visual rotation, *black* attend to inertial rotation. *Error bars* and *whiskers* show standard error and standard deviation over participants

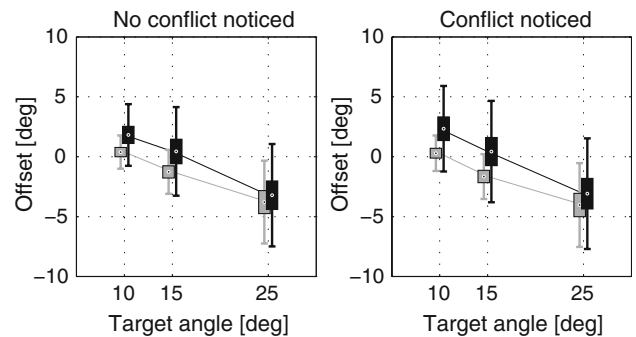


Fig. 7 Response offset as a function of detection of conflicts, target angle and attention condition, for 13 participants. *Gray* attend to visual rotation, *black* attend to inertial rotation. *Error bars* and *whiskers* show standard error and standard deviation over participants

Response variability in the different experimental conditions

If visual and inertial cues are integrated, this should result in a reduction of the variance in the participants’ responses, compared to the single-cue conditions VO and BO. Since each condition was tested eight times per subject, we can compute response variances for all participants and conditions. From the single-cue response variances σ_v^2 and σ_b^2 we can make a prediction for the variance of the responses in the combined-cue conditions under the assumption that the cues are integrated optimally (i.e., they follow a maximum-likelihood estimate; MLE):

$$\sigma_{v,b}^2 = \frac{\sigma_v^2 \times \sigma_b^2}{\sigma_v^2 + \sigma_b^2}$$

We can then compare this prediction with the response variances in the combined-cue conditions.

To compute the response variances, we first normalized the responses by dividing them by the target angle. In conditions VO and VA, the “target angle” was the passively presented visual rotation angle. In the BO and BA conditions it was the passively presented platform rotation angle. Then,

for each participant, target angle, and gain factor individually, we computed a response variance over the eight repetitions in that condition. We then took the arithmetic mean of these variances over the five gain factors. This resulted in 60 variance measures for the 20 participants \times 3 target angles. From these distributions we can calculate the mean, standard deviation, and standard error. Figure 8 shows the results for the different experimental blocks and the MLE prediction of the variance in the combined-cue conditions as black error bars with whiskers.

For the gray and white plots in the combined-cue conditions, the same procedure was followed, but the initial variances were calculated only from those trials which fell into the category “conflict detected” (white) or “conflict not detected” (gray). The variance for participant, target angle, and gain factor was only included if two or more data points fell into that category (one cannot compute a variance from just one, or no, data point). The distributions for VO, BO, and MLE are computed from 60 variances each, and each of these variances is computed over 40 trial repetitions. The black bar plots above VA and BA consist of 300 variances each, which are each computed over eight trial repetitions. The distributions in the “conflict detected” and “conflict not detected” conditions (gray and white bar plots) are computed from 200 to 283 variance measures, and each variance is computed from 4.0 to 5.4 trial repetitions on average.

The data shows a reduction of the variance in the combined-cue conditions compared to both single-cue conditions only in the VA condition, and only if the analysis is restricted to those trials in which participants did not notice a cue conflict (gray bar above “VA” in Fig. 8). Two-sample *t* tests show that the distribution of variances is significantly lower than the response variances in VO ($P < 0.05$) and BO ($P < 0.001$) conditions, and is not statistically distinguishable from the MLE prediction ($P > 0.05$). The overall variance distribution in VA (black

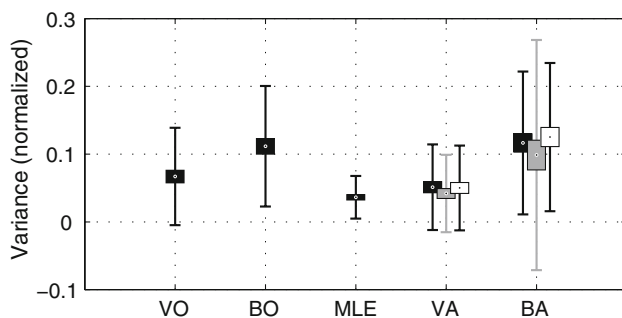


Fig. 8 Normalized response variances in single-cue conditions VO and BO, the MLE-prediction of the combined-cue variances from the single-cue variances, and measured variances in the combined-cue conditions VA and BA, overall (black), when no cue conflicts are noticed (gray) and when cue conflicts are noticed (white). Error bars and whiskers represent standard error and standard deviation, respectively

bar above “VA”) and the variance distribution restricted to “conflict noticed” trials (white bar above “VA”) are not significantly different from the distribution in VO, nor are they different from the MLE prediction.

Discussion

In this experiment we investigated the effect of task-induced attention to a particular modality on the integration of visual and inertial cues during actively reproduced turns. These effects were compared for conditions in which cue conflicts were perceived and when they were not perceived.

We found that asking participants to use one of the two modalities while ignoring the other significantly affected relative cue weighting. However, the influence of the ignored cue could not be completely suppressed. The cross-modal bias of vision on inertial cues was stronger than the bias of inertial cues on the visual rotation.

We also found that the amount of influence that the task had on the relative cue weights in the responses correlated with an awareness of the cue conflict. In trials in which no cue conflict was noticed, the effect of the task on the responses was not significant. If, however, conflicts were noticed, attending to one modality increased its weight in the combined estimate. In the following, we discuss the different experimental findings in more detail.

Over- and under-rotations in the turn reproduction task

We found that participants did not reproduce a large enough angle during large rotation trials (25°) in all four conditions BO, VO, BA, and VA. Under-rotations for actively returning a passively presented whole-body rotation have also been observed in a previous study by Israël et al. (1996), though for much larger rotation angles (180°). Such an under-rotation is somewhat surprising if both the passively presented and actively reproduced turns are misestimated similarly. Specifically, any under- or overestimations of the turn angles should cancel out in the turn reproduction task. The difference we find must thus come from a difference in misestimating the two rotations in a trial. The effect could be explained by a stronger underestimation of passive rather than active turns. Alternatively, it could be explained by a stronger underestimation of the first compared to the second turn; for example, if the memory of the first rotation decays over time. A third possibility is that since active returns were under the control of the participants, their motion profiles differed from a raised cosine, which could influence the perceived size of the rotation.

Jürgens and Becker (2006) proposed a model for human rotation perception, based on experimental evidence, in which a “default velocity” represents a top-down prior

which is integrated with bottom-up sensory cues in a Bayesian framework. This default velocity could be adapted to the average velocity during an experiment, and draw responses towards an average value. Jürgens and Becker found that this “tendency towards the mean” is reduced when more sensory cues are available, which is consistent with the idea that the tendency towards the mean is caused by a top-down Bayesian prior.

A similar prior could be the reason for the underestimation in our experiment, which would represent a very small or even zero rotation (0°). The Bayesian framework predicts that the less reliable the sensory estimate, the stronger the effect of this prior would be. Since the first turn has to be kept in memory, a decrease in reliability of the memorized first turn would increase the influence of a zero-rotation-prior on that turn, reducing its effective size, so that participants would turn less in their active response.

A different reason is suggested by the results of a study by Lappe et al. (2007). They showed that whether traveled distances are under- or overestimated depends on the task given to the participant. The difference is explained by a leaky integration during the movement, of either the distance to the starting point or the remaining distance to a target. If participants had to travel to a previously indicated target, they did not travel far enough (they thought they had traveled *farther* than they actually did); when they had to indicate the starting position after moving, they set the target too close (they thought that they moved *less* than they actually did).

In our experiment, both cases applied. Specifically, during the passively presented rotation, participants had to keep track of the starting position and during the active return, they had to update the distance to the target. According to Lappe et al. (2007), participants would first assume that they traveled less than they actually did in the passive rotation and traveled farther than they actually did in the active return. Consequently, both effects would add up and cause an under-rotation in the response, which is consistent with our findings.

Effect of turn size on cue weights

We found a significant effect of the target angle on the cue weights, with a higher visual weight for small rotations (10° target angle) and a lower visual weight for large rotations (25° target angle), while the opposite was true for inertial cue weights.

In this experiment, all passively presented rotations were 3 s long, independent of the rotation angle. This means that larger rotations also had higher accelerations and velocities than smaller rotations. Therefore, we could not determine whether the differences in cue weighting depended on rotation angle, velocity, or acceleration.

If the MLE framework is correct, cue weights should follow the respective reliabilities of the individual cues. Therefore, in our experiment it would be predicted that visual rotations would be relatively more reliable for small angles and inertial rotations would be more reliable for larger rotations. The smallest of our rotations ($3.3^\circ/s$ average) were quite close to vestibular thresholds compared to the largest rotations ($8.3^\circ/s$ average). The threshold for such cosine yaw rotations of about 3 s duration is around $1.5^\circ/s$ in darkness and $0.55^\circ/s$ in the presence of a visual target (Benson et al. 1989; Benson and Brown 1989), but the variability across participants is high. The threshold for visual motion is lower, in the range of $0.3^\circ/s$ (Mergner et al. 1995).

Also, for larger visual rotations, the visual target region went off the projection screen, which may have reduced the reliability of the estimate of the size of the visual turn. It is therefore reasonable to assume that the reliability of the inertial cue increases with respect to the reliability of the visual cue as the turn size increases.

Effects of attending to a modality and noticing cue conflicts

Participants responded differently when they were instructed to use visual cues than when they were instructed to use inertial cues for their response. This suggests that the cues are not always mandatorily fused into a singular “combined percept” of the rotation. Instead, our results indicate that for their response, participants put more weight on the modality which was to be attended according to the task instructions. Further analyses showed that the effect of task-defined attention on the response is significant only in trials in which participants responded that they noticed a difference between the visual and the inertial rotation, but is not significant if they reported that they did not notice a difference. This suggests that participants could use task-defined attention to reduce the influence of a to-be-ignored cue on the response, particularly when they notice that it is conflicting with the to-be-attended cue. In our results, even when participants noticed cue conflicts, the ignored modality still had an effect on the response. Thus, participants did not have access to pure sensory signals from the individual modalities during cue conflict conditions and could not simply use one cue for the response while completely ignoring the other cue. This is consistent with findings from a study by Bertelson and Radeau (1981) on visual-auditory cue integration, where it was found that a crossmodal influence on stimulus localization can still occur even when cue conflicts are noticed.

In another experiment investigating the influence of a concurrent visual stimulus on auditory localization,

Wallace et al. (2004) asked participants after each trial whether they had perceived the signals in the two modalities as coming from the same or from different sources. Comparable to the results in our study, they found that the reported auditory stimulus location was strongly drawn towards the location of the visual stimulus only when participants reported that they had perceived visual and auditory stimuli as a unified event, but not when they were perceived as separate events. Also, the response variability was lower when visual and auditory events were perceived as unified. This is consistent with our results, as we also found lower response variances when participants did not perceive a conflict between visual and inertial cues. We found an indication of true integration, as evidenced by a reduction of the response variance in comparison to the single-cue conditions, only in the VA condition when the visual cue was to be attended, and only if no cue conflicts were reported. In that condition, the variance of the responses was not statistically distinguishable from what would be expected if visual and inertial cues were optimally integrated in a maximum-likelihood fashion.

Lambrey and Berthoz (2003) also investigated the effect of awareness of conflicts between visual and body yaw rotations on cue weights, but in a different way. They did not instruct the participants to use one of the modalities and ignore the other, and they also did not tell them about the cue conflicts. During the experiment the participants were interrogated repeatedly to find out at what point during the experiment they became aware of the conflict. The experimenters then compared the weights of the cues before and after the participants became aware of the conflict. They found that about half of their participants had a bias towards using visual cues, and the other half inertial cues. After the participants became aware of the conflict, the bias towards the preferred cue was increased. Our results agree with those findings, and additionally show that task-induced attention can select the preferred cue. Our results also show that the strength of the attentional bias can change on a trial-by-trial basis as a function of the current awareness of a conflict when participants are not naïve about the possible occurrence of cue conflicts.

Helbig and Ernst (2008) performed a visual-haptic cue integration experiment to investigate the effect of attention on modalities on multimodal integration. They manipulated the amount of available resources for the processing of signals in visual and haptic modalities differentially by introducing a secondary task. They did not find an effect of the secondary task on cue weights, showing that visual-haptic cue integration was immune to their manipulation of attentional resources. Since they kept cue conflicts so small that they were not noticed by the participants, these results are also consistent with the present study.

Some caveats

There are a few caveats when interpreting the results of this study. Firstly, since participants always rotated back actively, we could not control the duration and movement profile of the second turn in each trial. Many participants, but not all, tried to imitate the raised-cosine rotation profile of the passive rotations. The variability in the motion profile of their active turns might add to the response variability we find within and across participants. Also, the rather short delay between the passive rotation and the active return might cause a contamination of our results by rotation aftereffects. Such aftereffects can, however, not explain any response differences as a function of task instruction or conflict awareness. A study by Siegler et al. (2000), where much larger and longer rotations were shown to the participants, did not find any difference in the response accuracy whether or not a yaw turn was reproduced immediately after presentation or after the end of post-rotatory sensations.

We found a correlation between “becoming aware of a cue conflict” and the “strength of the task-defined attentional bias” towards one of the modalities. However, we could not determine from this study whether this is a causal relationship and if so, what the causal direction is. Further experiments will be needed to evaluate whether noticing a cue conflict enforces top-down attentional influences, or whether participants become aware of cue conflicts more often when they are more attentive.

Neural basis of self-motion perception

The neural basis of the multimodal perception of self-motion in humans is still obscure. Most imaging methods do not allow movement of the participant’s head, and imaging methods that would be feasible during self-motion, e.g., EEG and near-infrared spectroscopy (NIRS), have a rather low spatial resolution. Some studies have investigated brain activity in response to large-field optic flow stimuli (Brandt et al. 1998; Beer et al. 2002; Kleinschmidt et al. 2002; Deutschländer et al. 2004; Wall and Smith 2008) and interactions of visual and vestibular self-motion signals by using caloric stimulation (Deutschländer et al. 2002). Since the stimuli used are not true self-motion stimuli, it remains unclear whether the same results would be obtained with actual self-motion. Some of the observed effects might be attributable to discrepancies of vestibular and visual stimulation.

Animal studies have shown that vestibular and visual signals already interact at the level of the vestibular nuclei (Henn et al. 1974) and that several separate but interconnected regions in the cortex process vestibular information. Most importantly these include, the posterior insular region

termed the “posterior insular vestibular cortex” (PIVC) as well as regions in parietal, somatosensory, cingulate, and premotor cortices (Guldin and Grüsser 1998)

Studies on cortical processing of self-motion stimuli using single-cell recordings in animals have focused on only a few regions; in particular areas MSTd (Froehler and Duffy 2002; Gu et al. 2007; Takahashi et al. 2007; Morgan et al. 2008; Britten 2008) and the ventral intra-parietal sulcus VIP (Bremmer et al. 2002a, b; Britten 2008). It has been suggested that cortical area MSTd, which contains cells that are sensitive to large fields of visual motion and also to vestibular stimulation, could be a central cortical area for integration of visual and vestibular signals of self-motion. A subpopulation of the neurons in MSTd shows responses to visual and vestibular translations which are consistent with a Bayesian cue integration model (Gu et al. 2007; Morgan et al. 2008). For rotations, however, such cells were not found (Takahashi et al. 2007). Virtually all cells in MSTd which are sensitive to both visual and vestibular rotations prefer opposite rotation directions for the two modalities. This suggests that instead of integrating visual and vestibular stimuli, cells in MSTd actually remove head rotations from the visual motion signal so that the resulting responses represent movement of objects in space while discounting self-rotation. Therefore, MSTd is most likely not the brain region in which visual and vestibular signals of self-rotations are integrated.

For yaw rotations, like those investigated in this experiment, it has been shown in rats that there is a special system of interconnected subcortical and cortical regions which maintains and updates a heading signal of the animal based on visual, vestibular, and somatosensory cues; the so-called “head-direction cell system” (Taube et al. 1990a, b; Taube and Bassett 2003; Zugaro et al. 2000). Although mostly studied in rats, such cells have also been found in primates (Robertson et al. 1999), which suggests that they might also be present in humans and might play a larger role in the cortical processing of yaw self-rotations than currently recognized.

Conclusions

We find that the transition from integration to a task-defined bias towards one of the cues coincides with becoming aware of the conflict between the cues. Also, we found a reduction in the response variability compared to single-cue conditions (which is an indicator of actual cue integration), only when the visual stimulus was attended to and no cue conflicts were noticed. This suggests that participants do not integrate visual and inertial cues in all situations, but that they can choose one cue for the response when they notice cue conflicts. Further,

participants can, to some extent, ignore the conflicting cue, but not completely. We could not find significant evidence for an effect of task-defined attention on cue weights when no cue conflicts were noticed.

Taken together, our results corroborate findings from studies on the integration of other sensory modalities (Warren et al. 1981; Bertelson and Radeau 1981; Wallace et al. 2004) and suggest that the integration of visual and inertial cues during earth-vertical rotations breaks down if participants notice cue conflicts.

These results are consistent with a “robust integration” model of multisensory integration, in which task-defined top-down influences, like the instruction to attend to one stimulus and ignore the other, can be used by participants to resolve conflicting sensory stimuli in accordance with the current task.

As has been proposed by previous studies (Warren et al. 1981; Talsma et al. 2007; Mozolic et al. 2008), the task to attend to one cue and ignore the other could reduce the integration of cues from different modalities, whereas attending to both cues at the same time could promote their integration. In our study, we found that the integration broke down particularly when participants noticed conflicts between the cues, and only then did the task instructions have a significant effect. Thus, in cases when participants assume that both cues give compatible information about the measured value, it might be more difficult for them to follow the task instructions and ignore one of the cues than when they become aware that the information in the cues is conflicting.

Acknowledgments This work was supported by the Max Planck Society and the Sonderforschungsbereich 550 of the Deutsche Forschungsgemeinschaft (DFG). We would like to thank Massimiliano Di Luca, John Butler, Jean-Pierre Bresciani and Jennifer Campos for helpful discussions.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alsius A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual integration of speech falters under high attention demands. *Curr Biol* 15:839–843
- Alsius A, Navarra J, Soto-Faraco S (2007) Attention to touch weakens audiovisual speech integration. *Exp Brain Res* 183:399–404
- Andersen TS, Tiippana K, Sams M (2004) Factors influencing audiovisual fission and fusion illusions. *Cogn Brain Res* 21:301–308
- Beer J, Blakemore C, Previc FH, Liotti M (2002) Areas of the human brain activated by ambient visual motion, indicating three kinds of self-movement. *Exp Brain Res* 143:78–88

- Benson AJ, Brown SF (1989) Visual display lowers detection threshold of angular, but not linear, whole-body motion stimuli. *Aviat Space Environ Med* 60(7):629–633
- Benson AJ, Hutt ECB, Brown SF (1989) Thresholds for the perception of whole body angular movement about a vertical axis. *Aviat Space Environ Med* 60:205–213
- Bertelson P, Radeau M (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys* 29(6):578–584
- Bertelson P, Vroomen J, de Gelder B, Driver J (2000) The ventriloquist effect does not depend on the direction of deliberate attention. *Percept Psychophys* 62(2):321–332
- Brandt T, Bartenstein P, Janek A, Dieterich M (1998) Reciprocal inhibitory visual-vestibular interaction—visual motion stimulation deactivates the parieto-insular vestibular cortex. *Brain* 121:1749–1758
- Bremmer F, Duhamel J-R, Hamed SB, Graf W (2002a) Heading encoding in the macaque ventral intraparietal area (VIP). *Eur J Neurosci* 16(8):1554–1568
- Bremmer F, Klam F, Duhamel J-R, Hamed SB, Graf W (2002b) Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur J Neurosci* 16(8):1569–1586
- Britten KH (2008) Mechanisms of self-motion perception. *Annu Rev Neurosci* 31:389–410
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215
- Deuschländer A, Bense S, Stephan T, Schwaiger M, Brandt T, Dieterich M (2002) Sensory system interactions during simultaneous vestibular and visual stimulation in PET. *Hum Brain Mapp* 16:92–103
- Deuschländer A, Bense S, Stephan T, Schwaiger M, Dieterich M, Brandt T (2004) Rollvection versus linearvection: comparison of brain activations in PET. *Hum Brain Mapp* 21:143–153
- Driver J (1996) Enhancement of selective listening by illusory mislocation of speech sound due to lip-reading. *Nature* 381:66–68
- Ernst MO (2006) A Bayesian view on multimodal cue integration. In: Knoblich G, Thornton IM, Grosjean M, Shiffrar M (eds) *Human body perception from the inside out*. Oxford University Press, New York, pp 105–131
- Ernst MO (2007) Learning to integrate arbitrary signals from vision and touch. *J Vis* 7(5):1–14
- Ernst MO, Bühlhoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8(4):162–169
- Froehler MT, Duffy CJ (2002) Cortical neurons encoding path and place: where you go is where you are. *Science* 295:2462–2465
- Gu Y, DeAngelis GC, Angelaki DE (2007) A functional link between area MSTd and heading perception based on vestibular signals. *Nat Neurosci* 10(8):1038–1047
- Guldin WO, Grüsser OJ (1998) Is there a vestibular cortex? *Trends Neurosci* 21(6):225–271
- Harrar V, Harris LR (2007) Multimodal ternus: visual, tactile, and visuo-tactile grouping in apparent motion. *Perception* 36:1455–1464
- Helbig HB, Ernst MO (2008) Visual-haptic cue weighting is independent of modality-specific attention. *J Vis* 8(1):1–16
- Henn V, Young LR, Finley C (1974) Vestibular nucleus units in alert monkeys are also influenced by moving visual fields. *Brain Res* 71:144–149
- Israël I, Bronstein AM, Kanayama R, Faldon M, Gresty MA (1996) Visual and vestibular factors influencing vestibular ‘navigation’. *Exp Brain Res* 112:411–419
- Jaekl PM, Jenkin MR, Harris LR (2005) Perceiving a stable world during active rotational and translational head movements. *Exp Brain Res* 163:388–399
- Jürgens R, Becker W (2006) Perception of angular displacement without landmarks: evidence for Bayesian fusion of vestibular, optokinetic, podokinesthetic, and cognitive information. *Exp Brain Res* 174:528–543
- Kitazaki M, Sato T (2003) Attentional modulation of self-motion perception. *Perception* 32:475–484
- Kleinschmidt A, Thilo KV, Büchel C, Gresty MA, Bronstein AM, Frackowiak RSJ (2002) Neural correlates of visual-motion perception as object- or self-motion. *Neuroimage* 16:873–882
- Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. *PLoS ONE* 2(9):e943
- Lambrey S, Berthoz A (2003) Combination of conflicting visual and non-visual information for estimating actively performed body turns in virtual reality. *Int J Psychophysiol* 50:101–115
- Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res* 35(3):389–412
- Lappe M, Jenkin M, Harris LR (2007) Travel distance estimation from visual motion by leaky path integration. *Exp Brain Res* 180:35–48
- Li Q, Kochiyama T, Wu J (2007) Effects of modality-specific attention on audiovisual early integration. In: *Proc 2007 IEEE int conf complex medical engineering*, pp 1467–1470
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748
- Meng M, Tong F (2004) Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *J Vis* 4:539–551
- Mergner T, Schweigart G, Kolev O, Hlavacka F, Becker W (1995) Visual-vestibular interaction for human ego-motion perception. In: Mergner T, Hlavacka F (eds) *Multisensory control of posture*. Plenum Press, New York, pp 157–167
- Morgan ML, DeAngelis GC, Angelaki DE (2008) Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59:662–673
- Mozolic JL, Hugenschmidt CE, Peiffer AM, Laurienti PJ (2008) Modality-specific selective attention attenuates multisensory integration. *Exp Brain Res* 184:39–52
- Nakamura S, Shimojo S (1999) Critical role of foreground stimuli in perceiving visually induced self-motion (vection). *Perception* 28:893–902
- Paffen CLE, Alais D, Verstraten FAJ (2006) Attention speeds binocular rivalry. *Psychol Sci* 17(9):752–756
- Palmisano S (2002) Consistent stereoscopic information increases the perceived speed of vection in depth. *Perception* 31:463–480
- Pouget A, Deneve S, Duhamel J-R (2004) A computational neural theory of multisensory spatial representations. In: Spence C, Driver J (eds) *Crossmodal space and crossmodal attention*. Oxford University Press, Oxford, pp 123–140
- Robertson RG, Rolls ET, Georges-Francois P, Panzeri S (1999) Head direction cells in the primate pre-subiculum. *Hippocampus* 9:206–219
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31(8):401–409
- Shams L, Kamitani Y, Shimojo S (2000) What you see is what you hear. *Nature* 408:788
- Siegler I, Viaud-Delmon I, Israël I, Berthoz A (2000) Self-motion perception during a sequence of whole-body rotations in darkness. *Exp Brain Res* 134:66–73
- Takahashi K, Gu Y, May PJ, Newlands SD, DeAngelis GC, Angelaki DE (2007) Multimodal coding of three-dimensional rotation and translation in area MSTd: Comparison of visual and vestibular selectivity. *J Neurosci* 27(36):9742–9756

- Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J Cogn Neurosci* 17(7):1098–1114
- Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex* 17:679–690
- Taube JS, Bassett JP (2003) Persistent neural activity in head direction cells. *Cereb Cortex* 13:1162–1172
- Taube JS, Muller RU, Ranck JB (1990a) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J Neurosci* 10(2):420–435
- Taube JS, Muller RU, Ranck JB (1990b) Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J Neurosci* 10(2):436–447
- Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cogn Psychol* 12:97–136
- van Ee R, van Dam LCJ, Erkelens CJ (2002) Bi-stability in perceived slant when binocular disparity and monocular perspective specify different slants. *J Vis* 2:597–607
- van Ee R, Adams WJ, Mamassian P (2003) Bayesian modeling of cue interaction: bistability in stereoscopic slant perception. *J Opt Soc Am A* 20(7):1398–1406
- Vatakis A, Spence C (2007) Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli. *Percept Psychophys* 69(5):744–756
- Wall MB, Smith AT (2008) The representation of egomotion in the human brain. *Curr Biol* 18:191–194
- Wallace MT (2004) The development of multisensory processes. *Cogn Process* 5:69–83
- Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004) Unifying multisensory signals across time and space. *Exp Brain Res* 158:252–258
- Warren DH, Welch RB, McCarthy TJ (1981) The role of visual-auditory “compellingness” in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept Psychophys* 30(6):557–564
- Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88(3):638–667
- Yuille AL, Bülthoff HH (1996) Bayesian decision theory and psychophysics. In: Knill DC, Richards W (eds) *Perception as Bayesian inference*. Cambridge University Press, London, pp 123–161
- Zugaro MB, Tabuchi E, Wiener SI (2000) Influence of conflicting visual, inertial and substratal cues on head direction cell activity. *Exp Brain Res* 133:198–208