

Storing upright turns: how visual and vestibular cues interact during the encoding and recalling process

Manuel Vidal · Heinrich H. Bühlhoff

Received: 21 February 2009 / Accepted: 7 July 2009 / Published online: 25 August 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Many previous studies have focused on how humans combine inputs provided by different modalities for the same physical property. However, it is not yet very clear how different senses providing information about our own movements combine in order to provide a motion percept. We designed an experiment to investigate how upright turns are stored, and particularly how vestibular and visual cues interact at the different stages of the memorization process (encoding/recalling). Subjects experienced passive yaw turns stimulated in the vestibular modality (whole-body rotations) and/or in the visual modality (limited lifetime star-field rotations), with the visual scene turning 1.5 times faster when combined (unnoticed conflict). Then they were asked to actively reproduce the rotation displacement in the opposite direction, with body cues only, visual cues only, or both cues with either the same or a different gain factor. First, we found that in none of the conditions did the reproduced motion dynamics follow that of the presentation phase (Gaussian angular velocity profiles). Second, the unimodal recalling of turns was largely uninfluenced by the other sensory cue that it could be combined with during the encoding. Therefore, turns in each modality, visual, and vestibular are stored independently. Third, when the intersensory gain was preserved, the bimodal reproduction was more precise (reduced variance) and lay between the two unimodal reproductions. This suggests that with both

visual and vestibular cues available, these combine in order to improve the reproduction. Fourth, when the intersensory gain was modified, the bimodal reproduction resulted in a substantially larger change for the body than for the visual scene rotations, which indicates that vision prevails for this rotation displacement task when a matching problem is introduced.

Keywords Self-motion · Yaw rotations · Spatial orientation · Vestibular · Multisensory integration

Introduction

Humans move around in a world where their different senses provide a set of cues about self-motion that allows them to have good estimations of changes in body position and heading over time. The spatial information contained in each of these cues is often derived from the same physical property of the environment, and therefore can be redundant. For instance, passive upright turns create a rotation of the visual and acoustic scene sensed by the visual and auditory system, respectively (external signals), together with a set of inertial-related cues sensed by the vestibular system and the somatosensory receptors (internal signals). In recent years, technological advances have allowed an increased number of studies on the way these signals are combined by the central nervous system in order to correctly perceive self-motion in space. However, only a few studies have assessed how sensory cues might interact during the memorization process. This paper focuses on the encoding and recalling stages of the self-motion memorization process using a reproduction task. The encoding stage consisted in the storage of an upright turn presented passively, and the recalling stage consisted

Electronic supplementary material The online version of this article (doi:10.1007/s00221-009-1980-5) contains supplementary material, which is available to authorized users.

M. Vidal (✉) · H. H. Bühlhoff
Max Planck Institute for Biological Cybernetics,
Tübingen, Germany
e-mail: manuel.vidal@college-de-france.fr

in remembering the stored turn to reproduce it actively. Specific manipulations were done in order to assess how distinct sensory cues available during the rotations might interact during the encoding stage, and combine to improve the reproduction during the recalling stage.

Self-motion perception with vision

Since the early 1950s, many studies have emphasized how processing optic flow alone provides humans with a very efficient mechanism to guide their movements in a stable environment. Although early work emphasized the importance of heading retrieval (Warren et al. 1991a, b), a general steering heuristic requiring far less complex processing of optic flow has been proposed (Wann and Land 2000). The discrimination of the traveled distance between two simulated visual translations can be performed with great accuracy, even when the velocity profile and duration vary (Bremmer and Lappe 1999; Frenz et al. 2003). Nevertheless, when testing a reproduction task, the same authors found significant overshoots for distances of up to 8 m and undershoots for larger distances (Bremmer and Lappe 1999; Frenz and Lappe 2005). In a recent study, using a passive velocity profile controlled by computer for the reproduction task, this range effect was largely cancelled when the visual stimulation was identical during the encoding and the recalling stages (Mossio et al. 2008). Furthermore, the large inter-subject differences reported earlier were explained in this work by highlighting the use of distinct strategies among the subjects. Indeed, some could cope with velocity profile manipulations exhibiting an integration process independent of the profile, while others could not reproduce accurately when the velocity profile was manipulated. In the present study, we will test the capacity to reproduce turns with an active control of the velocity profile. Therefore, we assumed that either subject would cope with the velocity profile they use in order to perform the task, or they would try to reproduce the same velocity profile as during the encoding. Finally, for more complex visually simulated movements in which heading is not tangent to the trajectory, combining an initial vestibular stimulation allows humans to accurately disambiguate the analysis of optic flow (Bertin et al. 2000; Bertin and Berthoz 2004). This result provides evidence for a multisensory advantage for the self-motion analysis.

Self-motion perception with the vestibular system

In complete darkness, the acceleration signals provided by the vestibular system and proprioception allow humans and other mammals to integrate self-motion and to keep track of one's position in the environment (Mittelstaedt and Mittelstaedt 1980; Loomis et al. 1999). Distances as well as

the temporal velocity profiles of passive linear translations performed in darkness can be reproduced with great accuracy (Berthoz et al. 1995; Israël et al. 1997). The production of joystick-controlled on-axis body rotations when subjects are verbally instructed of the total turn to perform was significantly undershot, in other words, the subjects felt that they had moved further than the actual physical movement (Israël et al. 1995). Similarly, the reproduction with a pointer of orientation change for on-axis rotations and curved paths was overestimated (Ivanenko et al. 1997). These findings suggest that humans tend to overestimate the rotation sensation given by the vestibular and somatosensory systems, although it is compensated when turns are reproduced and therefore matched within the same modality (Siegler et al. 2000). Curiously, in this study, the dynamics of the presented rotations, that is the temporal velocity profile, were not reproduced, as it has been reported for forward translations. Neither the same motion duration nor peak velocity was used for the reproductions. In a later study, Israël and collaborators investigated whether this could stem from the different processing of the signals provided by the otoliths and the canals. They tested on- and off-axis rotations with the same paradigm to find that the addition of otolithic cues lowered the performance and that the dynamics were closer to the presented one (Israël et al. 2005). The present study focused on pure whole-body on-axis yaw rotations stimulating mostly the semi-circular canals. We checked whether vision could also contribute to the storage of the rotation dynamics in order to reproduce velocity profiles.

Multisensory interactions for spatial orientation

How humans integrate cues provided by different senses for the same physical property have been studied for many decades. According to Welch and Warren's 'modality appropriateness hypothesis', the reliability of each signal should be taken into account, the relative modality dominance depending upon the precision of its signal in the given task context (Welch and Warren 1980). In a similar line of research, an increasingly popular model based on the maximum likelihood estimator (MLE) was recently proposed (Ernst and Banks 2002; Ernst and Bühlhoff 2004). This Bayesian framework for sensory integration puts forward the following principle: The available signals provided for the same physical property are combined into a percept defined by the weighted sum of the signals, with weights determined by the inverse of their respective variances (measured in unimodal conditions).

In recent years, a large number of studies that used direct perception tasks have reported that the MLE model very accurately predicts the observed performance when modalities are combined. Nevertheless, only a few have

assessed its validity for motion perception and spatial orientation in general. When dealing with general movements that include translations and tilting, the central nervous system is faced with an ambiguity in the interpretation of incoming signals. Indeed, all linear accelerometers measure gravito-inertial force (GIF), which is the sum of gravitational force (tilt) and inertial force due to linear acceleration (translation). A growing literature has studied and modeled how additional cues combine with the GIF measured by the otoliths to extract self-motion and orientation: from the semi-circular canals signal and graviceptors (Zupan et al. 2000; Merfeld et al. 2001), to visual input (MacNeilage et al. 2007), to top-down cognitive expectations (Wertheim et al. 2001).

Finally, little research has focused on the visuo-vestibular interactions involved in everyday life navigation. Unlike in weightlessness or in a dense milieu such as water, in which navigation involves rotations about any axis and peculiar gravitational conditions (Vidal et al. 2003, 2004), natural navigation over planar surfaces includes only translations and upright yaw rotations. Recent electrophysiological studies on non-human primates found a neural structure—the dorsal medial superior temporal area (MSTd)—that provides a functional link between brain activity and heading perception based on inertial motion cues (Gu et al. 2007). In the very same structure, a sub-population of neurons with congruent heading preferences for visual and vestibular cues strongly correlated with monkeys' perceptual decisions when both modalities are present (Gu et al. 2008). Curiously, the processing of visual and vestibular cues in MSTd during self-rotations showed substantial qualitative differences compared to translations (Takahashi et al. 2007). Indeed, the large majority of neurons showed a maximal incongruence in the rotation direction preference for visual and vestibular cues, suggesting that the integration of these cues for robust perception cannot take place in MSTd. The role of vestibular signals in this area could then be restricted to disambiguate optic flow resulting from self-motion from that produced by eye, head and body rotations.

These neurophysiological results are consistent with the dissociation between visual and vestibular storage that was found in an older human behavioral study. After traveling a virtual path where physically turning the body was driving the visual orientation, when a multiplicative factor was introduced between body and visual rotations, subjects could reproduce each sensory turn separately depending on the task context (Lambrey et al. 2002). The observed locomotor reproduction of turns matched the ones performed during the presentation, whereas drawings of the paths matched the visual experience. These results concern rather high-level mechanisms, and one of the motivations of the present study was to test if this apparent task-

dependent behavior is to be related to the different primary modality involved in each task rather than the task itself.

Rationale

We designed an experiment to investigate how humans store upright turns, and particularly how vestibular and visual cues available during the rotations interact at the encoding and recalling stages of the memorization process. In this experiment, subjects experienced a turn and they were instructed to reproduce the turn amplitude backwards by controlling the velocity with a joystick. The sensory contexts of the encoding and recalling stages were manipulated, which could include visual and/or vestibular cues about the turn. Furthermore, a gain factor between these modalities was used during the presentation phase in order to dissociate the two modalities and to allow inferring on which sensory basis was the resulting reproduction done. On the one hand, we addressed the issue of the reproduction strategy: will the velocity profile of the encoding stage be reproduced for turns as it was reported for linear translations in darkness, and will the reproduced profiles differ according to the sensory context available? On the other hand, we addressed several issues about the sensory interactions in the memorization process. Concerning the encoding stage, will unimodal reproduction vary according to the encoding sensory context? Will each of the cues available in the stimulation be recovered accurately? Although visual and vestibular cues combine at a very early stage of processing in the brainstem, this integration concerns mostly low-level control of eye movements and does not imply that the initial unimodal information is lost. We, therefore, expect a large degree of independence in the storage of visual and vestibular cues, which should result in very little interaction during the encoding. Concerning the recalling stage, will the combination of modalities improve the reproduction of turns? Will visual cues prevail when a modality matching conflict is introduced during the reproduction? If an integration process occurs, it should result in better performance when both cues are available.

Materials and methods

Subjects

Twelve naïve subjects (nine male and three female) between 18 and 28 years participated in this experiment. Most of them were students and all but one were right handed. Subjects were paid for their participation time according to the Max Planck regulations and were informed that they could interrupt the experiment at any time.

Experimental set-up

A hexapod motion platform with a car seat and a projection screen was used to produce synchronized visual and body rotations (see Fig. 1).

Visual rotations were projected on a flat screen subtending 87° of horizontal field of view, centered on the subject's line of sight. The three-dimensional geometry of the visual stimulus was adjusted to the real perspective. A limited lifetime star field was projected on the surface of a virtual vertical cylinder with a radius of 6 m. The virtual observer was placed on the cylinder's axis, therefore at a simulated distance of 3 m from the dots. Visual rotations corresponded to the virtual rotation of the cylinder around its axis. The star field was composed of approximately 600 simultaneously visible dots, each subtending 0.2° . Dots were randomly positioned and displayed during 2 s, including a fade-in and fade-out period of 300 ms. After

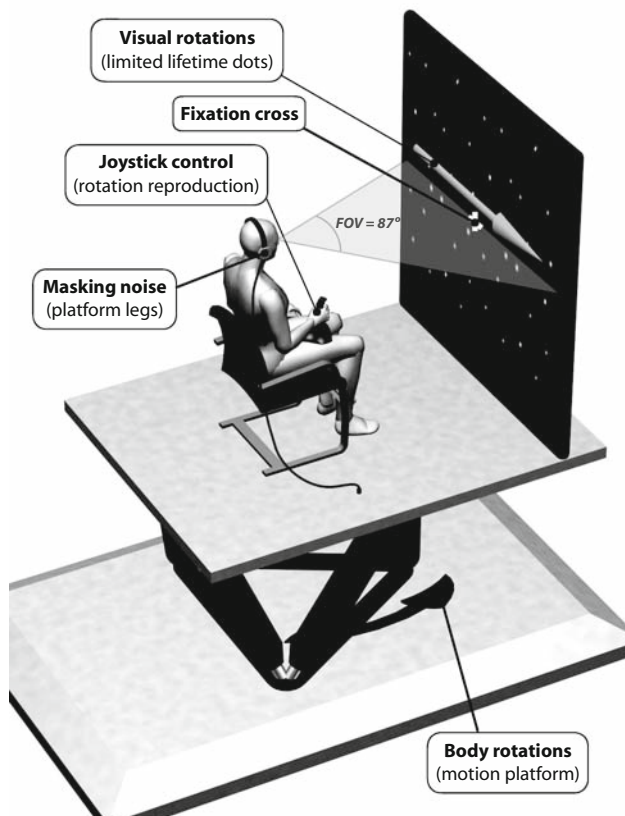


Fig. 1 The experimental set-up. Visual rotations were projected on a flat screen subtending 87° of field of view, and corresponded to the virtual rotation of a vertical cylinder with a limited lifetime star field on its surface. Body rotations were performed with a hexapod motion platform around the body vertical axis (yaw turns). Subjects could reproduce their rotation tilting a joystick leftwards or rightwards. During the rotations, subjects had to look at a central fixation cross, and were wearing a noise-cancellation headphone playing a masking noise recorded from the platform legs

extinction, each dot was replaced by another randomly positioned dot.

Body rotations were performed with a hexapod motion platform (Maxcue from MotionBase™) remote-controlled by computer. The seat was adjusted so that the yaw rotations performed with the platform were around the vertical axis that passed between the two ears of the subject.

The subject reproduced the turns with a joystick positioned in front of him. Tilting the handle leftward or rightward rotated the platform and/or the visual display. The rotation speed varied according to the tilt angle, e.g., the more the handle was tilted the faster the rotation was (with a visual or body maximum velocity of 40 or $26.7^\circ/s$, respectively). During both passive and reproduced turns, subjects were instructed to look at a fixation cross that was displayed in the center of the screen. A notifying sound was played through the headphone when the fixation appeared. Eye movements were recorded with an SMS iView eye tracker in order to verify subjects' fixations. During the whole experiment, subjects had to wear a headphone with noise cancellation that played a masking noise that was a recording of the sound made by the platform legs when moving. This suppressed the noise of the motion platform that could have been used as an additional cue for the motion speed.

Procedure

The experiment consisted of two experimental sessions, separated by a minimum of 10 days to prevent learning effects from transferring across sessions. All trials had the same structure, including a presentation phase where subjects had to memorize a turn (encoding stage) and a reproduction phase where they were asked to reproduce this turn backwards (recalling stage).

In the presentation phase of the first session, subjects passively experienced a whole-body upright rotation synchronized to a visual scene rotation with a gain factor of 1.5 (i.e., turning 1.5 times faster). This conflict between vision and vestibular rotations was used in order to dissociate the two modalities and to be able to infer on which sensory basis the reproduction is done. We used pilot subjects to determine a good trade-off between higher gain factors for increased dissociation and gain factors closer to 1 for unnoticed conflicts. Indeed, a sensory discrepancy remaining unnoticed allows improving the ecological validity and prevents subjects from developing unnatural specific strategies for solving the task (De Gelder and Bertelson 2003). This issue was verified for each subject in a debriefing questionnaire at the end of the experiment. After a 2-s delay started the reproduction phase. Subjects were instructed to reproduce the perceived amplitude of the rotation (i.e., rotation displacement) in the opposite

direction using the joystick in one of the four following conditions (see Fig. 2): with the visual scene rotation but no platform motion (*VB to V*), with the body rotation and only the fixation cross displayed on the screen (*VB to B*), with vision and body where the vision/body rotation used the same 1.5 gain (*VB to VBsame*) or a different 1.0 gain (*VB to VBdiff*) than during the presentation phase.

The second session was similar to the first except that this time stimulations were always unimodal. The presentation consisted of purely visual or purely vestibular turns and the reproduction phase was done in the exact same modality as the presentation phase, again in the opposite direction. In this experimental session, only two conditions were studied (see Fig. 2): *V to V* and *B to B* labeled according to the studied modality.

Subjects were never informed before the reproduction in which condition they would be tested on. Reproduced turns were validated by pressing the joystick’s trigger, then in a few seconds the platform was slowly repositioned (5°/s) in order to dispose of a maximal reproduction range for the following trial, which then started automatically. The delay between the presentation and the reproduction phase was voluntarily chosen to be small. A longer delay would have allowed the semi-circular canals to return to a stable activation state before the next stimulation, but this would have required about 30 s and

the memory decay would then have become problematic. Furthermore, another study found no difference between waiting or not for this stable state before reproducing body turns (Siegler et al. 2000).

Presented turns could take two directions (leftward or rightward), and three different amplitudes (45° and/or 30°, 60° and/or 40°, 75° and/or 50° for the visual and/or body rotations, respectively). These rotation angles were chosen so to cover a natural range of turns in everyday life locomotion. As mentioned earlier, the difference between visual and body rotations stem from the unnoticed conflict that was used in order to disentangle these modalities. The angular velocity of the rotations followed gaussian profiles, which are known to be a natural profile for head rotations (see Appendix A of Supplementary material). Furthermore, in order to avoid an easy correlation between the turn amplitude and either the total rotation duration or the peak velocity, both the peak velocity and the total rotation duration varied according to the final amplitudes. Rotation durations were 5.4, 6.1, and 6.6 s; visual (body) peak velocities were 19.5°/s (13.0°/s), 23.6°/s (15.7°/s) and 27.4°/s (18.2°/s) for the three rotation amplitudes, respectively. The velocity range was chosen to cover natural heading changes resulting from full body turns during navigation. The velocity profiles that were used for the presentations, e.g., gaussian profiles, are quite similar to

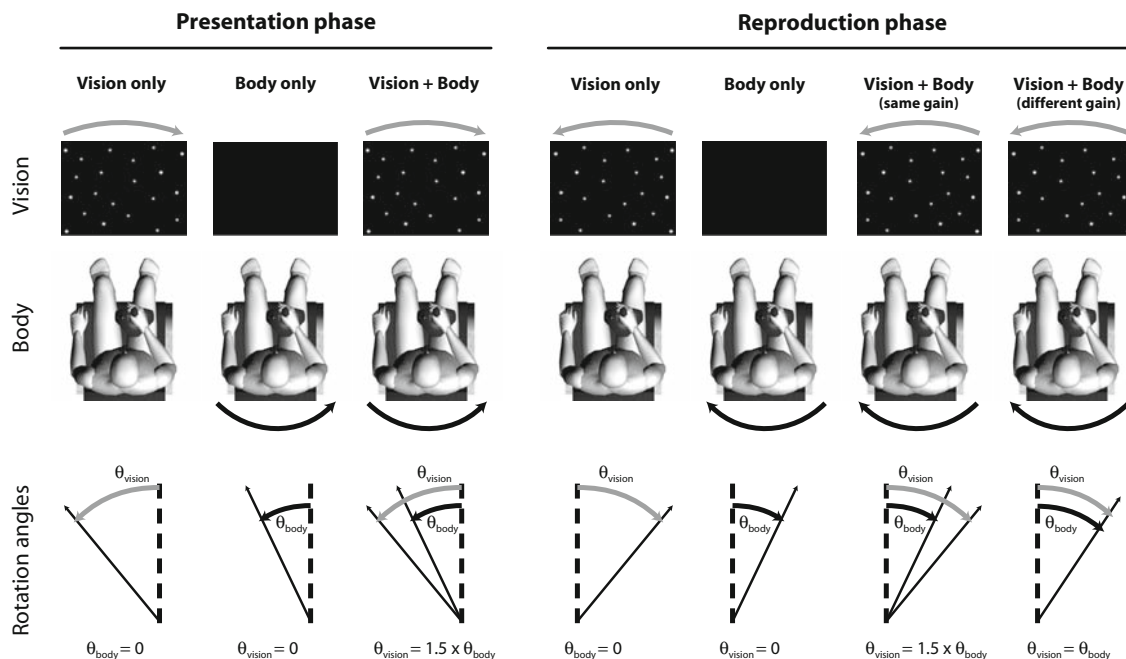


Fig. 2 Illustration of the different presentation and reproduction phases defining the six experimental conditions that were studied. The presentation phase was either a purely visual (*V to V*) or a purely vestibular (*B to B*) stimulation representing upright passive rotations, or a body rotation coupled with the corresponding visual rotation amplified by a gain factor of 1.5 (*VB to V*, *VB to B*, *VB to VBsame* and

VB to VBdiff). In the reproduction phase, subjects were asked to reproduce backwards the perceived rotation in one of the four following sensory contexts: with vision only (*V to V* and *VB to V*), body only (*B to B* and *VB to B*), or both modalities with the same 1.5 gain (*VB to VBsame*) or with a different 1.0 gain (*VB to VBdiff*) than during the presentation

single cycle sinusoids with frequencies ranging from 0.076 to 0.1 Hz. At these frequencies, the sensitivity thresholds range roughly between 1.75 and 2.25°/s (Grabherr et al. 2008), which is far below the velocities experienced. Finally, although these frequencies lie at the limits of the sensitivity plateau where the threshold is about 0.75°/s, it still falls within a range where the vestibular system is fully operational. Note that in the second experimental session, presentations were unimodal and reproductions were tested using the same modality. Therefore, there were no conflicting cues although the tested visual and body turns were not the same. This difference was preserved in order to allow for within subject comparisons with the first experimental session conditions. Note also that in all conditions, the gaze fixation was used in order to prevent subjects from using the vestibulo-ocular reflex (body rotations) or target pursuit (visual rotations) to perform the task.

At the beginning of each experimental session, a set of practice trials including two trials per condition was conducted in order to ensure that the task was clearly understood. During these trials, subjects were instructed to play around with the joystick and feel the effect it has on their motion in the various sensory contexts studied. The data from these trials were not collected. Six repetitions of each of the turn \times condition were performed, corresponding to a total number of 144 trials and 72 trials in the first and second session, respectively. The order was randomized for each subject and trials were blocked in groups of 16. At the end of each block, they could rest a while without leaving the platform; and after 3 blocks, they could rest for up to 10 min outside of the experimental room.

Data analysis

The trajectory of the reproduction was recorded for each trial, e.g., the instantaneous visual and/or body orientation sampled at a frame rate of 60 Hz. The final orientation determined the reproduced rotation amplitude and the amplitude ratio (reproduced/presented rotation amplitude). Matlab scripts were used to process these files in order to analyze the reproduction dynamics. Consequently, we achieved an optimal visualization of the angular velocity profiles of all trials according to each condition, and the best Gaussian fit for each of these profiles was computed providing the root mean square error (RMSE) as an indicator for fitness. The motion duration and maximum angular velocity (peak velocity) were also extracted.

Several repeated-measures ANOVA designs were used to analyze the statistical effects and significance when comparing the different experimental conditions. Student's *t* tests were also used in order to compare observations with single reference values. Post hoc analyses were performed with Tukey's HSD test.

Results

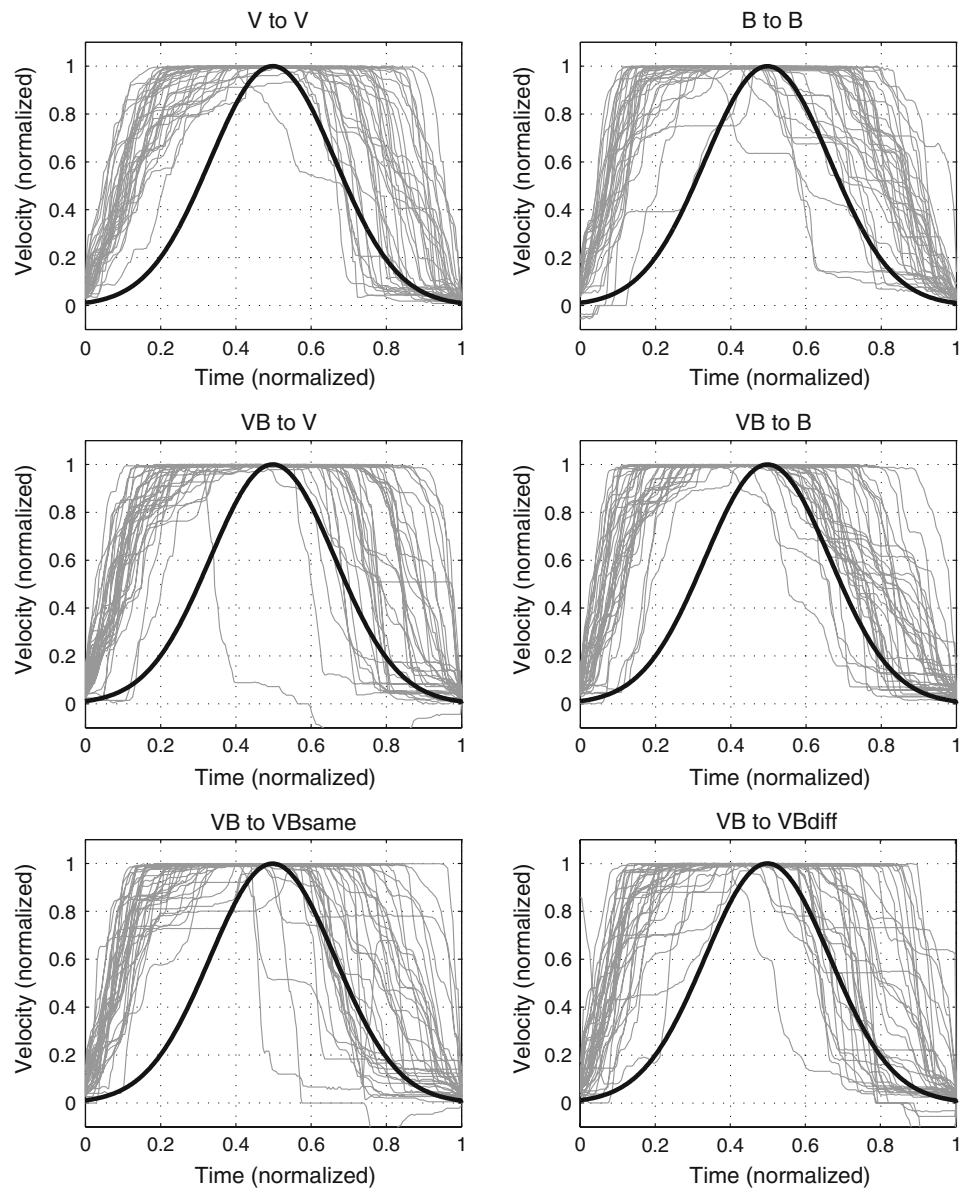
Turn direction had no significant effect in any of the statistical tests performed, therefore we disregarded the direction factor and leftward/rightward data were pooled together for all the analyses presented hereafter.

Reproduction dynamics

Figure 3 shows all the reproduction velocity profiles of a standard subject (gray lines) together with the Gaussian velocity profile of the presentation phase (black thick line) for each of the six conditions studied. The subject was selected as the closest to median values of the RMSE (described above) obtained for each condition. The time axis was normalized with the motion duration of each reproduction plotted, cutting the start and end tails where subjects did not move with the joystick. The velocity axis was normalized with the peak velocity of each reproduction. Applying the same normalizations to the presented velocity profile leads to a unique Gaussian curve shown in black in each condition plot. A quick qualitative comparison when reading these plots allows stating that the overall shape of the reproduced profiles does not match the Gaussian presented profiles. Rather, the strategy adopted by subjects seems to rely on the use of well-controlled trapezoidal rotational velocity profile, showing an initial linear speed increase in order to reach the speed of the plateau that is then held constant for most of the reproduction duration before decreasing to stop the motion. The decreasing period being subjected to some corrections, it results in the larger spread observed in the plots as compared to the increasing period.

The average motion duration ratio and peak velocity ratio (reproduced/presented) are shown for each condition in Fig. 4a and b, respectively. There was no significant difference among the conditions for the motion duration ratio, which was for all conditions significantly below 1 (Student test: $t(11) > 4.5$; $p < 0.001$). On the other hand, the condition had a main effect on the peak velocity ratio ($F(5, 55) = 7.3$; $p < 0.001$). Post hoc tests showed that the peak velocity ratio with only visual cues available during the reproduction (*V to V* or *VB to V*) was significantly higher than with only body cues available (*B to B* or *VB to B*) ($p < 0.005$ for all comparisons). The peak velocity ratio of bimodal reproduction (*VB to VB_{same}*) was not significantly different from any of the unimodal reproduction. Finally, all but the *VB to V* condition were significantly below 1 (Student test: $t(11) = 2.3$; $p < 0.05$ for *V to V* and $t(11) > 5.6$; $p < 0.001$ for the remaining conditions). The RMSE of the best Gaussian fit of each reproduced velocity profile, normalized with the presented peak velocity and averaged for each condition, is showed in Fig. 4c.

Fig. 3 Plots of all the reproduced velocity profiles of a standard subject according to each of the six studied conditions. The Gaussian velocity profile of the presentation phase is plotted in *black thick line*. Both the velocity and the time were normalized in order to focus on the profile shape only and to be able to compare them across conditions (see the text for more details about these normalizations)



Similarly, there was no significant difference across conditions, and for each condition these errors were significantly larger than 0 ($t(12) > 9.1$; $p < 0.001$ for all conditions). To summarize, there was no difference between the conditions in any of the motion dynamics parameters of the reproduction, except for the peak velocity of purely visual reproduction, which was significantly higher. Moreover, the motion dynamics of the reproduction was nearly always significantly different from that of the presented rotations, indicating that subjects did not reproduce the velocity profiles.

Modality interactions during the encoding stage

In order to see whether the two sensory cues interact at the encoding stage, we looked at the recalling of turns in each

sensory context (visual or body) and compared the performance for unimodal and bimodal turn presentations. Figure 5 shows the reproduced turn amplitude for visual reproduction and body reproduction as a function of the presented turn angle. For visual reproduction of turns, there was no significant main effect of the encoding sensory context (V to V vs. VB to V , $F(1,11) = 0.79$; $p = 0.39$) and interaction with the turn angle ($F(2,22) = 0.25$; $p = 0.78$). Similarly, for bodily reproduction of turns, there was no significant main effect of the encoding sensory context (B to B vs. VB to B , $F(1,11) = 0.67$; $p = 0.43$) and interaction with the turn angle ($F(2,22) = 1.41$; $p = 0.26$). The reproduction dynamics did not differ according to the presentation sensory context: there was no significant difference at all between bimodal and unimodal encoding for both maximum angular velocities and motion durations (see Fig. 4).

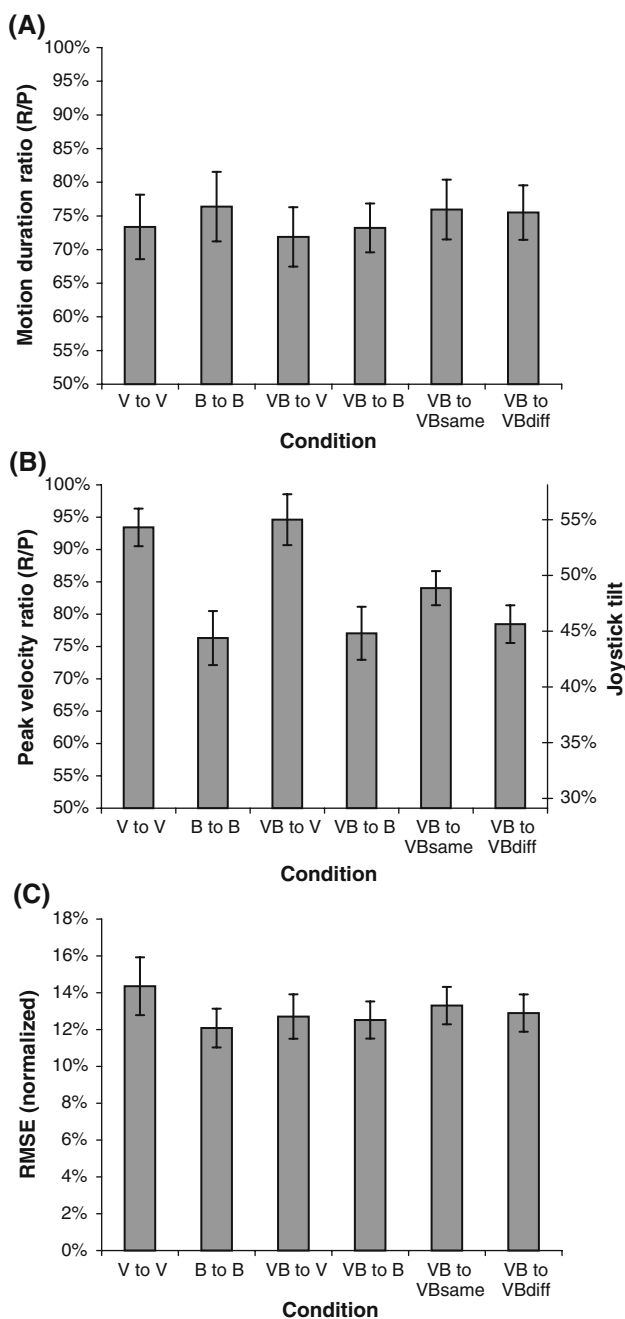


Fig. 4 The reproduced/presented motion duration ratio (a), peak velocity ratio with the corresponding joystick tilt (b), and the normalized RMSE of each reproduced velocity profile's best Gaussian fit (c), averaged for each experimental condition. *Error bars* show the standard errors across subjects

Bimodal reproduction

Figure 6a shows the amplitude ratio (reproduced/presented) when subjects were presented a bimodal turn which they had to reproduce with visual cues alone (*VB to V*), vestibular cues alone (*VB to B*), or both with the same gain (*VB to VBsame*). Reproduction with vision alone seems

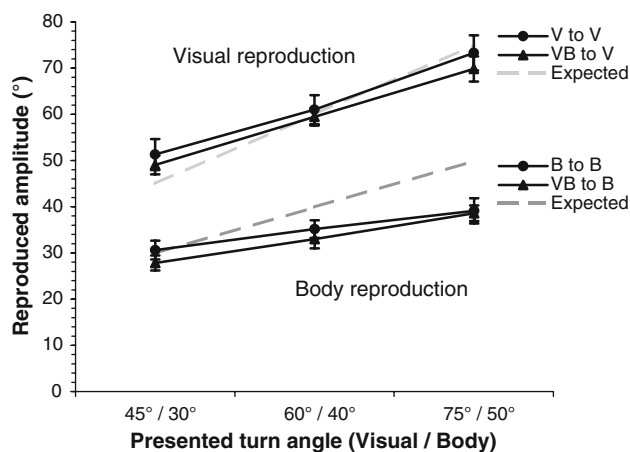


Fig. 5 The reproduced amplitudes for visual only reproduction (*V to V* and *VB to V* conditions) and body only reproduction (*B to B* and *VB to B* conditions), as a function of the presented turn angle. *Dashed lines* show the expected correct visual (*light gray*) and body (*dark gray*) reproductions. *Error bars* show the standard errors across subjects

rather accurate whereas with body alone there is an overall underestimation of about 15%. When both modalities are available for reproduction, bimodal performance appears a little underestimated as compared to that of vision alone, and a little overestimated as compared to that of body alone. The reproduction condition has a significant effect on the amplitude ratio ($F(2,22) = 5.79$; $p < 0.01$). Post hoc tests showed that the bimodal reproduction was significantly smaller than for the visual reproduction of 60° ($p < 0.05$) and 75° ($p < 0.02$) visual turns, and was significantly larger than for the body reproduction of 30° ($p < 0.001$), 45° ($p < 0.001$) and 60° ($p < 0.001$) body turns. Therefore, when both modalities are available during reproduction, the average of the response lies in between the reproduction with vision or body alone.

Figure 6b shows the individual variances within a given experimental condition of the amplitude ratio, averaged across subjects, together with the variance predicted by the MLE for the bimodal reproduction. The classic equation linking the unimodal variances with the expected bimodal variance was used to compute these predictions (Ernst and Banks 2002), further details on how this model was applied to our experimental data can be found in Appendix B of Supplementary material. The reproduction condition has a significant effect on these variances ($F(2,22) = 6.29$; $p < 0.01$). Post hoc tests showed that the bimodal reproduction variance was significantly lower than that of both unimodal reproductions ($p < 0.05$ for *VB to V* and $p < 0.01$ for *VB to B*). The measured bimodal reproduction variance was significantly higher than that predicted by the MLE for the measured unimodal variances (2.8% instead of 1.7% predicted, Student test: $t(11) = 3.5$; $p < 0.005$). These results indicate that variability reduces when

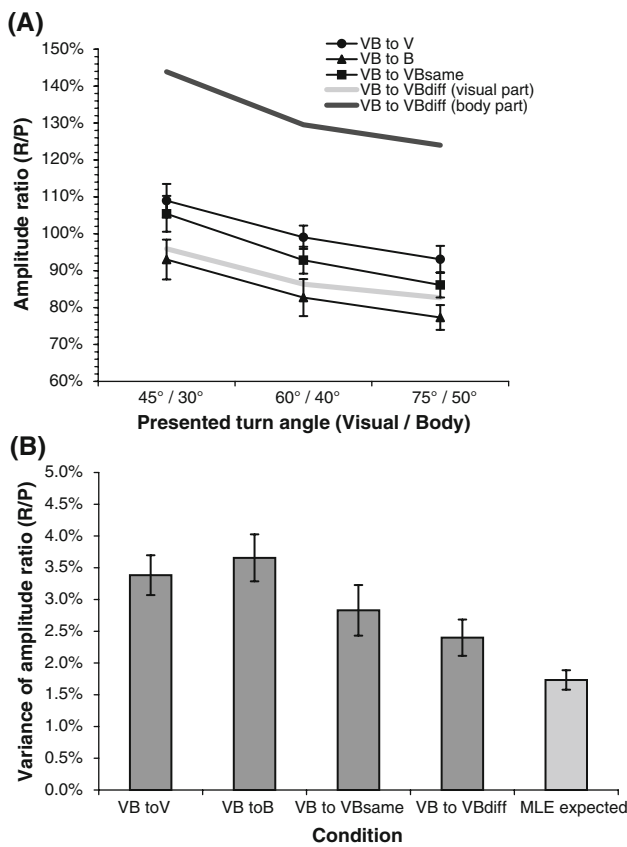


Fig. 6 **a** Amplitude ratio (reproduced/presented) plotted as a function of the presented turn angle for bimodal (*VB to VBsame*) and unimodal reproductions (*VB to V* and *VB to B*). The *light* and *dark gray* lines show the *VB to VBdiff* ratios for the visual and body parts of the reproduction, respectively. **b** Individual variance averages measured for each condition together with the variance predicted by the maximum likelihood estimator (MLE). *Error bars* show the standard errors across subjects

bimodal information is available as compared to unimodal information, but this improvement is not optimal as defined by this Bayesian framework for sensory fusion.

Reproduction with a modified sensory gain

When the vision/body gain factor was altered for the reproduction, using a gain of 1 instead of 1.5, we introduced a matching problem in the sense that it becomes impossible to simultaneously match both the visual and body rotations during the reproduction. We calculated the relative variation of the visual or body reproduction as follows:

$$\Delta_{\text{Vision}} = \frac{r_{\text{Different gain}}^{\text{Visual}} - r_{\text{Same gain}}^{\text{Visual}}}{r_{\text{Different gain}}^{\text{Visual}} - r_{\text{Different gain}}^{\text{Body}}} \quad \text{and}$$

$$\Delta_{\text{Body}} = \frac{r_{\text{Different gain}}^{\text{Body}} - r_{\text{Same gain}}^{\text{Body}}}{r_{\text{Different gain}}^{\text{Visual}} - r_{\text{Different gain}}^{\text{Body}}}$$

where $r_{\text{Different or Same gain}}^{\text{Visual or Body}}$ is the visual or body amplitude ratio in the same or different gain conditions (light and dark gray lines shown in Fig. 6a). The reproduced body rotation increases by 84.5% of the variation (see Fig. 7), whereas the reproduced visual rotation decreases by only 15.5% (significant difference, Student test: $t(11) = 11.45$; $p < 0.001$). In other words, the visual matching remains rather unchanged whereas the body matching becomes totally inappropriate, as compared to the baseline performance yielded with the *VB to VBsame*.

Discussion

Turn reproduction strategy

Ruling out the motor strategy

In the conditions where presentation was bimodal, the experienced turns were identical whether the reproduction sensory context used the same or a different gain. Since the reactivity of the joystick was adjusted so to have the same rotation speed for the visual stimulation and the platform relatively to the expected angle, subjects could then have used a motor strategy in order to reproduce turns. For all conditions except the one with a modified gain (*VB to VBdiff*), reproducing the same dynamic tilt profile with the

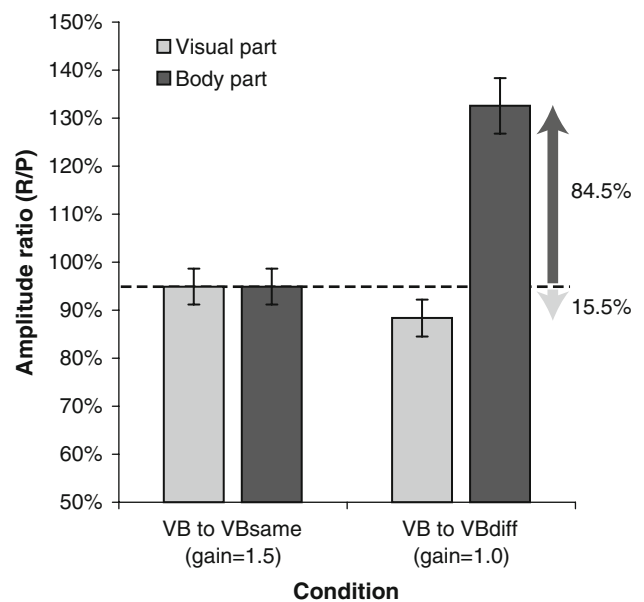


Fig. 7 Distribution of the amplitude ratio variation due to the matching error when the vision to body gain is changed from 1.5 (same as presented) to 1.0 (different than presented). 84.5% goes in the increase of the body rotation whereas only 15.5% in the decrease of the visual rotation. In other words, subjects chose implicitly to match the visual turn correctly rather than the body turn

joystick would lead to the same answers across conditions. Accordingly, subjects could only estimate the “size” of the turn presented, and then generate a constant motor profile independently of the condition assuming that these were constant across conditions. The analysis of the reproduction’s peak velocity shows that this was not the case: visual reproductions (*VB to V* or *V to V*) were significantly larger than vestibular reproductions (*VB to B* or *B to B*) of about 23% in average.

It is important to notice that the joystick range did not limit the peak velocities of the reproduction profile’s plateaus: in all conditions the tilt angle was way below the maximum (see Fig. 4b), with a maximum of about 55% tilt for visual reproduction. In fact, when controlling the platform rotation with the joystick, subjects felt more comfortable not turning too fast, whereas with visual turns they could turn as fast as they wished without feeling shaken by the physical motion. This provides an additional argument against the possible use of a motor strategy allowing turn reproductions regardless of the sensory context available.

Velocity profiles are not reproduced

During the encoding stage, subjects could have memorized the entire velocity profile instead of just storing its temporal integration corresponding to the total turn amplitude. Then, in order to reproduce backwards the turn amplitude as they were instructed to, they could just rollback the velocity profile converting it into a motor command to tilt the joystick. This strategy has been reported for vestibular reproduction of linear translations (Berthoz et al., 1995), and was interpreted by the authors as an indication to the nature of how self-motion might be encoded.

In the current experiment, we wanted to check whether the whole velocity profile would also have been stored for on-axis turns using different sensory contexts, namely visual, vestibular or visuo-vestibular motions. To that end, we left on purpose this velocity profile strategy available: not only the velocity was linearly controlled by the tilt angle of the joystick (velocity command), but also the reactivity of the joystick was kept constant in a congruent fashion across the experimental conditions. Despite these efforts, in our experiment subjects used velocity profiles for their reproduction of turns that were particularly distinct in many aspects from the presented Gaussian velocity profile. In all the sensory contexts that we have studied, neither the motion duration, nor the peak velocity, nor the best Gaussian fit for each single reproduction, were compelling with an attempt to reproduce the velocity profile of the turn presented (see Fig. 4a, b). The dynamics of the reproduction observed in the present experiment rather follows a standard trapezoidal velocity profile, with some small

corrections in the end suggesting a temporal integration process undergoing, which is consistent with what was found for varying velocity profiles for visual translations (Mossio et al. 2008). Both the peak velocity maintained during the plateaus were, as discussed before, highly dependent on the sensory context of the reproduction.

This observation leads to the conclusion that the velocity profile might not be stored and that for turn reproduction tasks, a temporal integration is performed during the encoding stage, which is then matched to the temporal integration of the produced turn. Accordingly, the findings reported in Berthoz et al. (1995) could result from an experimental bias related to the chosen velocity profiles: trapezoidal, triangular and rectangular. These profiles all fall into the trapezoidal category, the last two being just singular cases (trapezoidal without plateau and trapezoidal with steep acceleration and deceleration slopes), which is precisely the profile that subjects tended to use in our experiment. It seems then that using such profiles would naturally lead to a reproduction with a similar profile, e.g., a trapezoidal profile.

Encoding interactions: sensory-independent storage

In order to determine the possible interactions between visual and vestibular senses during the encoding of turns, we designed two sets of unimodal reproduction conditions: one with unimodal presentations (visual or vestibular), and the other with bimodal presentations using a visuo-vestibular gain factor of 1.5.

On one hand, the free reproduction of pure visual turns was very accurate whereas a significant underestimation of about 15% was observed for pure vestibular turns. As reported in the dynamic analysis, this could stem from the distinct motor behaviors observed: a reduced peak velocity was used when rotating the platform as compared to rotating the visual field, whereas the motion duration itself was similar. These performances provide the baseline characteristics observed for totally independent unimodal reproductions.

On the other hand, when simultaneous visual and vestibular turns are presented during the encoding stage, one could expect that these senses interact during the encoding such that each unimodal reproduction would be biased towards that of the other presented modality. Since the visual field rotations turned faster than the corresponding body rotations, such interaction would lead to an overestimation when reproducing body turns and an underestimation when reproducing visual turns. Interestingly, these biases were not observed. There was no difference whatsoever in visual reproduction whether the encoding was purely visual or visuo-vestibular, which indicates that body turns did not interfere with the storage of visual turns.

Similarly, there was no significant difference in the vestibular reproduction. Surprisingly, there is a very small tendency to underestimate the reproduction of small body turns when vision was also available during the encoding, which goes in the opposite direction to the expected interaction. Nevertheless, to a large extent, visual turns did not interfere with the storage of vestibular turns.

Taken together, these results allow us to conclude that subjects could accurately extract each modality from the encoding of the presented visuo-vestibular turns in order to reproduce it. Therefore, whether or not a combination of cues occurs, the unimodal visual and vestibular information is not lost during the encoding process, which shows that each modality is stored independently. In fact, a merged visuo-vestibular information could have also been stored, but this third encoding would be largely independent of the unimodal encoding. This is consistent with recent findings that showed a dissociation between the processing of visual and vestibular cues in MSTd (Takahashi et al. 2007), suggesting that for self-rotations, the integration of these cues when both are present should occur in another cerebral structure. This storage independence confirms using the same low-level reproduction task, what had been reported previously using distinct tasks (Lambrey et al. 2002), and stays in-line with what was reported for independent sensory modalities (Hillis et al. 2002).

Recalling interactions: bimodal reproduction

Two reproduction conditions using different gain factors between the visual scene and the body rotations were studied in order to assess how the modalities interact at the recalling of visuo-vestibular turns. In a first condition, we kept this gain factor at 1.5 as during the presentation, which allowed to test with a classical multisensory approach how the availability of these two modalities might have improved the reproduction. In a second condition, we deliberately changed this gain factor so to make the matching of both modalities impossible. This last condition was designed to evaluate where the trade-off between the two modality matchings would be placed.

Multisensory integration

Performance observed when subjects had to reproduce the turns in the same sensory context (controlling both the body and visual scene rotations with the same discrepancy as during the presentation) shared some of the characteristics of an optimal signal combination provided by each of the senses. On the one hand, the average position of the bimodal reproduction of turns lay in between that of visual or vestibular alone. In other words, the vestibular part of the bimodal reproduction was pulled towards that of vision

alone, and simultaneously the visual part was pulled towards that of body alone. Consequently, since the vestibular reproduction was globally undershot compared to the visual reproduction, the vestibular part and the visual part of the bimodal reproduction were increased and decreased, respectively. In a similar fashion, the peak velocity of the reproduction observed for the bimodal condition was influenced by both intervening sensory cues, corresponding to a significantly faster reproduction than with the body only, but slower than with vision only. On the other hand, the variance of the reproduced turn amplitudes with bimodal stimulation was significantly smaller than the one observed with each of the unimodal stimulation, whether vestibular or visual.

These findings show that there is an advantage of matching both modalities over matching a single modality, showing that a sensory integration process is actually undergoing. The multimodal contribution for spatial mechanisms has also been elicited in an earlier study for mental rotations (Vidal et al. 2009). Nevertheless, this integration was not optimal as defined in the Bayesian MLE framework, the variance being still significantly higher than what the model predicts. We believe that the sensory integration optimality is not achieved because of the indirect nature of the tested perception. All previous studies that found a combination of cues consistent with the MLE model were direct perceptual tasks in the sense that the physical property to be evaluated did not rely on cues that were to be integrated over time: the cues could be directly sensed. In contrast, for turn perception, the estimated measure, e.g., the rotation amplitude, is not available at once. A temporal integration of visual and/or vestibular cues is required in order to estimate the total amplitude. Furthermore, since in our experiment, the velocity was not constant, one can expect the sensory weighting in the integration to evolve: slower rotations could favor the visual input, whereas faster rotations could favor the vestibular input. For these reasons, we believe that the sensory integration in such indirect perceptual task is not straightforward, which is why it did not follow the MLE predictions.

Matching problem: vision prevails

If we consider the task as a matching within modality between the presented stimulus stored in memory and the one being reproduced, we could wonder what happens when both cannot be simultaneously matched anymore. Modifying the visuo-vestibular gain in the reproduction phase created this situation. Changing from a gain of 1.5 between modalities to a gain of 1.0 introduced a compulsory mismatch of 50% that subjects had to distribute among the visual and the vestibular reproduction. One could expect two extreme strategies adopted to face this

problem. In one extreme, subjects would evenly distribute this mismatch between the two modalities, which would happen if the importance given to each of these modalities was equivalent for this task. In the other extreme, subjects would try to minimize the mismatch introduced for one particular modality, which would be the prevailing modality for this specific task. The distribution of the mismatch measured in the different gain conditions was closer to the second extreme. Indeed, 84.5% of the mismatch was attributed to the vestibular reproduction whereas only 15.5% to the visual reproduction, which shows that subjects prefer to adjust correctly the visual rotation disregarding the much larger vestibular rotation that was associated to it. This finding is valid within the studied velocity range, which covers natural heading changes resulting from full body turns. For higher velocities such as those experienced during head rotations, the visual system might not be precise any longer and the vestibular contribution could become predominant.

Conclusion

The purpose of our work was to reinforce the connection between self-motion perception and memorization with the multisensory framework. We focused on the general issue of on which sensory basis can humans memorize visuo-vestibular turns, and how these senses interact during the encoding and recalling stages.

First, recalling a memorized turn does not rely on the reproduction of the velocity profile, which suggests that for this task the velocity profile might not be stored. Instead a temporal integration is performed during the encoding stage, which would then be matched to the temporal integration of the produced turn, regardless of the velocity profile used. Second, the unimodal recalling of turns, either visual or vestibular, is independent of the other sensory cue that it might be combined with during the encoding stage. Therefore, turns in each modality (visual and vestibular) are stored independently. Third, when the intersensory gain of the recalling was preserved, the visuo-vestibular bimodal reproduction of turns was more precise (reduced variance) and lay between each unimodal reproductions. This suggests that with both visual and vestibular cues available, these combine in order to improve the reproduction (trade-off between modalities). As discussed before, the predictions of the MLE model did not apply possibly because of the fundamental difference between the instantaneous sensing of a physical property (direct perception) and, like in our experiment, a perception that requires integration over time (indirect perception). Fourth, modifying the reproduction gain resulted in a substantially larger change for the body than for the visual scene rotations reproduced.

Therefore, within the studied dynamic range, which corresponds to a natural range of whole-body rotation speeds involved during navigation, vision prevails when a visual/ vestibular matching problem is introduced.

Acknowledgments This research was supported by the Max Planck Society. The authors wish to thank Daniel Berger for his support with the experimental set-up, and for his helpful comments on earlier versions of the manuscript.

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

- Benson AJ, Hutt EC, Brown SF (1989) Thresholds for the perception of whole body angular movement about a vertical axis. *Aviat Space Environ Med* 60:205–213
- Berthoz A, Israël I, Georges-François P, Grasso R, Tsuzuku T (1995) Spatial memory of body linear displacement: what is being stored? *Science* 269:95–98
- Bertin RJ, Berthoz A (2004) Visuo-vestibular interaction in the reconstruction of travelled trajectories. *Exp Brain Res* 154:11–21
- Bertin RJ, Israël I, Lappe M (2000) Perception of two-dimensional, simulated ego-motion trajectories from optic flow. *Vis Res* 40:2951–2971
- Bremmer F, Lappe M (1999) The use of optical velocities for distance discrimination and reproduction during visually simulated self motion. *Exp Brain Res* 127:33–42
- De Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. *Trends Cogn Sci* 7:460–467
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433
- Ernst MO, Bühlhoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8:162–169
- Frenz H, Lappe M (2005) Absolute travel distance from optic flow. *Vis Res* 45:1679–1692
- Frenz H, Bremmer F, Lappe M (2003) Discrimination of travel distances from ‘situated’ optic flow. *Vis Res* 43:2173–2183
- Grabherr L, Nicoucar K, Mast FW, Merfeld DM (2008) Vestibular thresholds for yaw rotation about an earth-vertical axis as a function of frequency. *Exp Brain Res* 186:677–681
- Gu Y, Deangelis GC, Angelaki DE (2007) A functional link between area MSTd and heading perception based on vestibular signals. *Nat Neurosci* 10:1038–1047
- Gu Y, Angelaki DE, Deangelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci* 11:1201–1210
- Hillis JM, Ernst MO, Banks MS, Landy MS (2002) Combining sensory information: mandatory fusion within, but not between, senses. *Science* 298:1627–1630
- Israël I, Sievering D, Koenig E (1995) Self-rotation estimate about the vertical axis. *Acta Otolaryngol* 115:3–8
- Israël I, Grasso R, Georges-François P, Tsuzuku T, Berthoz A (1997) Spatial memory and path integration studied by self-driven passive linear displacement. I. Basic properties. *J Neurophysiol* 77:3180–3192
- Israël I, Crockett M, Zupan L, Merfeld D (2005) Reproduction of ON-center and OFF-center self-rotations. *Exp Brain Res* 163:540–546

- Ivanenko Y, Grasso R, Israël I, Berthoz A (1997) Spatial orientation in humans: perception of angular whole-body displacements in two-dimensional trajectories. *Exp Brain Res* 117:419–427
- Lambrey S, Viaud-Delmon I, Berthoz A (2002) Influence of a sensorimotor conflict on the memorization of a path traveled in virtual reality. *Brain Res Cogn Brain Res* 14:177–186
- Loomis JM, Klatzky RL, Golledge RG, Philbeck JW (1999) Human navigation by path integration. In: *Wayfinding behavior*. The John Hopkins University Press, Baltimore and London, pp 125–151
- MacNeilage PR, Banks MS, Berger DR, Bulthoff HH (2007) A Bayesian model of the disambiguation of gravito-inertial force by visual cues. *Exp Brain Res* 179:263–290
- Merfeld DM, Zupan LH, Gifford CA (2001) Neural processing of gravito-inertial cues in humans. II. Influence of the semicircular canals during eccentric rotation. *J Neurophysiol* 85:1648–1660
- Mittelstaedt ML, Mittelstaedt H (1980) Homing by path integration in a mammal. *Naturwissenschaften* 67:566–567
- Mossio M, Vidal M, Berthoz A (2008) Traveled distances: new insights into the role of optic flow. *Vis Res* 48:289–303
- 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* 27:9742–9756
- Vidal M, Lipshits M, McIntyre J, Berthoz A (2003) Gravity and spatial orientation in virtual 3D-mazes. *J Vestib Res Equilib Orientat* 13:273–286
- Vidal M, Amorim MA, Berthoz A (2004) Navigating in a virtual three-dimensional maze: how do egocentric and allocentric reference frames interact? *Cogn Brain Res* 19:244–258
- Vidal M, Lehmann A, Bühlhoff HH (2009) A multisensory approach to spatial updating: the case of mental rotations. *Exp Brain Res* 197:59–68
- Wann J, Land M (2000) Steering with or without the flow: is the retrieval of heading necessary? *Trends Cogn Sci* 4:319–324
- Warren WH Jr, Blackwell AW, Kurtz KJ, Hatsopoulos NG, Kalish ML (1991a) On the sufficiency of the velocity field for perception of heading. *Biol Cybern* 65:311–320
- Warren WH Jr, Mestre DR, Blackwell AW, Morris MW (1991b) Perception of circular heading from optical flow. *J Exp Psychol Hum Percept Perform* 17:28–43
- Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88:638–667
- Wertheim AH, Mesland BS, Bles W (2001) Cognitive suppression of tilt sensations during linear horizontal self-motion in the dark. *Perception* 30:733–741
- Zupan LH, Peterka RJ, Merfeld DM (2000) Neural processing of gravito-inertial cues in humans. I. Influence of the semicircular canals following post-rotatory tilt. *J Neurophysiol* 84:2001–2015