

# A multisensory approach to spatial updating: the case of mental rotations

Manuel Vidal · Alexandre Lehmann ·  
Heinrich H. Bühlhoff

Received: 20 April 2009 / Accepted: 1 June 2009 / Published online: 21 June 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** Mental rotation is the capacity to predict the outcome of spatial relationships after a change in viewpoint. These changes arise either from the rotation of the test object array or from the rotation of the observer. Previous studies showed that the cognitive cost of mental rotations is reduced when viewpoint changes result from the observer's motion, which was explained by the spatial updating mechanism involved during self-motion. However, little is known about how various sensory cues available might contribute to the updating performance. We used a Virtual Reality setup in a series of experiments to investigate table-top mental rotations under different combinations of modalities among vision, body and audition. We found that mental rotation performance gradually improved when adding sensory cues to the moving observer (from *None* to *Body* or *Vision* and then to *Body & Audition* or *Body & Vision*), but that the processing time drops to the same level for any of the sensory contexts. These results are discussed in terms of an additive contribution when sensory modalities are co-activated to the spatial updating mechanism involved during self-motion. Interestingly, this multisensory approach can account for different findings reported in the literature.

**Keywords** Mental rotations · Spatial updating · Multisensory · Virtual reality

## Introduction

In everyday life, we often have to imagine ourselves from another perspective in order to correctly drive our behavior and actions within the environment. This process involves mental rotations (or perspective taking), which can be defined as the capacity to mentally update either spatial relationships of objects or structural features within an object, after orientation changes in the observer's reference frame. This dynamic process is considered as analogue to the actual physical rotation of the objects or observer, and provides a prediction of the outcome of these relationships after the rotation. The topic of the present research concerns the specific contributions and interactions of different senses to the spatial updating mechanism involved during the observer motion in order to perform mental rotations.

Attempting to identify an object from an unusual viewpoint requires a cognitive effort that has long been characterized. The recognition of an oriented object from a novel viewpoint is harder and reaction times are proportional to the angular difference between the two orientations presented simultaneously (Shepard and Metzler 1971), which suggests an on-going mental rotation process performed at constant speed. Other studies also found view dependency effects of spatial memory for object layouts presented successively (Rieser 1989; Diwadkar and McNamara 1997). Christou et al. (2003) investigated the effect of external cues concerning the change in viewpoint on the recognition of highly view-dependent stimuli. They found that both visual background and indication of the next viewpoint improved participants' performance, thus providing evidence for egocentric or view-based encoding of shapes.

When an observer is viewing an array of objects, the same relative change in viewpoint can arise either from object array rotation or from viewer rotation. Since in the

---

M. Vidal · H. H. Bühlhoff  
Max Planck Institute for Biological Cybernetics,  
Tübingen, Germany

M. Vidal (✉) · A. Lehmann  
LPPA, CNRS, Collège de France, Paris, France  
e-mail: manuel.vidal@college-de-france.fr

absence of external visual cues, these two transformations result in identical changes in the observer's retinal projection of the scene, traditional models of object recognition would predict a similar cognitive performance. Simons and Wang (1998) compared for the first time these two situations in a table-top mental rotation task. Participants learned an array of objects on a circular table. They were then tested on the same array either after a change in viewing position or a rotation of the table, both leading to the same change in relative orientation of the layout. They found that when the change in orientation resulted from the observer motion rather than from the table rotation, the recognition performance was much higher indicating an improved mental rotation mechanism. Surprisingly, when participants moved to a novel viewing position, performance was better if the layout had remained static than if it rotated in order to present the exact same view. This facilitation effect found for physically moving observers, often referred to as the viewer advantage, has also been reported in other types of mental rotation paradigms, namely with imagined rotations (Amorim and Stucchi 1997; Wraga et al. 1999, 2000), with virtual rotations (Christou and Bühlhoff 1999; Wraga et al. 2004), and finally for haptically learned object layouts (Pasqualotto et al. 2005; Newell et al. 2005). These findings strongly contradict earlier object recognition theories and suggest the existence of distinct mechanisms at work for static and moving observers. Until recently, the most widely accepted interpretation is that when the change in viewpoint arises from the observer locomotion, the latter benefits from the spatial updating mechanism (Simons and Wang 1998; Wang and Simons 1999; Amorim et al. 1997). This updating mechanism would allow integrating self-motion cues in order to dynamically update spatial relationships within the test layout, resulting in improved mental rotation performance. Nevertheless, recent findings cast doubts on the interpretation of the locomotion contribution. Indeed, this advantage was cancelled for larger rotations or when additional cues indicating the change in perspective are provided (Motes et al. 2006; Mou et al. 2009). Accordingly, a new interpretation of the role of locomotion was formulated in the later study: the updating involved during locomotion would simply allow keeping trace of reference directions between the study and test views. In order to ensure that differences arise exclusively from the spatial updating mechanism, in the present study we used a table texture with clearly visible wood stripes that provide a strong reference direction in every condition.

This updating mechanism, which has been proposed as a fundamental capacity driving animal navigation behaviors (Wang and Spelke 2002), can be fed by a variety of sensory information. On one hand, the vestibular system provides self-motion-specific receptors which in conjunction with

the somatosensory inputs, allow humans and mammals to keep track of their position in their environment even in complete darkness, by continuously integrating these cues (Mittelstaedt and Mittelstaedt 1980; Loomis et al. 1999). On the other hand, since the 1950s, many studies have emphasized how processing optic flow provides humans with very efficient mechanisms to evaluate and guide self-motion in a stable environment (Gibson 1950). Klatzky et al. (1998) studied how visual, vestibular and somatosensory cues combine in order to update the starting position when participants are exposed to a two-segment path with a turn. Systematic pointing errors were observed when vestibular information was absent, which suggest an on/off discrete contribution of the vestibular input, rather than a continuous integration with other cues such as vision. In our view, a reduction of sensory information should reduce the global reliability of motion cues and therefore increase errors, but not in such systematic fashion. Recent disorienting maneuvers showed that updating object placed around participants does not involve allocentric information (Wang and Spelke 2000), suggesting that even if participants were given time to learn their locations, object-to-object relationships are not encoded. A recent study extended these results showing that if participants learned the object layout from an external point of view, which is the case of table-top mental rotation tasks, object-to-object properties are encoded and participants can point towards the objects even after disorientation (Mou et al. 2006). The authors explain that when objects are placed around participants, allocentric properties such as intrinsic axis cannot be drawn easily, and only the updating of egocentric representations can be used to perform the task, which is impaired by disorientation.

Simons and Wang (1998) investigated whether background visual cues around the test layout played a role in the viewer advantage using phosphorescent objects in a dark room. Moving observers still yielded significantly better mental rotation performance than static, although the advantage was reduced compared to conditions in which environmental cues were available. This reduction led us to believe that in this task, different sources of information could contribute to the viewer advantage, namely visual cues. Nonetheless, in follow-up experiments these authors reported that extra-retinal cues were responsible for this advantage and claim that visual cues do not contribute at all (Simons et al. 2002). In this study, very poor visual information was available compared to previous work: static snapshots before and after the change in viewpoint with a rather uniform background were used. Participants experienced only a limited field of view and were provided with no dynamic visual information about their rotation. In the present work, the visual contribution to the updating mechanism will be assessed independently from other modalities. We believe that richer visual cues can in fact improve

mental rotations, if not alone, at least contribute to the overall effect.

To summarize, several studies have manipulated modalities (visual, vestibular, somatosensory) that could be involved in the mental rotation process. Nevertheless, it is not yet clear what are the exact contribution of these sensory modalities and how they might interact together. In particular, no previous study has really assessed the role of audition, although relevant acoustic cues could also improve mental rotations. The aim of the present research was to use the exact same setup in order to compare mental rotation performance of moving observers, when the spatial updating mechanism is provided with unimodal information or with different bimodal combinations. A multimodal virtual reality platform was designed so as to systematically measure the performance in a table-top mental rotation task under different sensory contexts. We believe that the richer the sensory context available for the updating mechanism is, the better mental rotations will be. This new multisensory approach of mental rotations will allow explaining a large range of findings reported previously.

## General methods

### Participants

Twelve university students (4 females and 8 males, mean age  $24.4 \pm 3.3$ ) participated in experiment A and another 12 (5 females and 7 males, mean age  $24.5 \pm 3.9$ ) participated in experiment B. All were right-handed except one in experiment A and two in experiment B. All had normal or corrected-to-normal vision. None of the participants knew of the hypotheses being tested.

### Apparatus

The experiments were conducted using an interactive Virtual Reality setup that immersed participants in a partial virtual environment. The setup consisted of a cabin mounted on the top of a 6 degree of freedom Stewart platform (see Fig. 1, left panel). Inside of this completely enclosed cabin was a seat and a round physical table having a diameter of 40 cm that was placed between the seat and a large projection screen. The seat position was adjusted in order to have a constant viewing position across participants (50 cm away from the table central axis, 38.5 cm above the table surface, and 138 cm away from the front projection screen, subtending  $61^\circ$  of horizontal FOV). A real-time application was developed using Virtools™, a behavioral and rendering engine, in order to synchronously control the platform motion, the visual background and the table layout with the response interface. An infrared camera



**Fig. 1** *Left* A sketch of the experimental setup. Participants sat inside a closed cabin mounted on a motion platform that contained a front projection screen displaying the virtual scene, a table placed in the middle of the cabin with a screen and touch screen embedded displaying the test object layout and recording the participant answers. The different motion cues available during the viewpoint changes were achieved with a combination of the following manipulations: *P* the platform rotation, *R* the room rotation on the front screen, *T* the layout rotation on the table screen, and a speaker providing a stable external sound cue. *Right* The five objects used in the spatial layouts: a mobile phone, a shoe, an iron, a teddy bear and a roll of film

in the cabin of the setup allowed the experimenter to continuously monitor the participants. In the following sections, we will specify how viewpoint changes were done, how the test objects were displayed, and how the response interface was designed.

### Changes in observer position

In every trial the participants were rotated passively around an off-vertical axis centered on the tabletop using different combinations of motion cueing such as, body motion, visual scene motion, and acoustic scene motion. Note that this rotation corresponds to a rotation in the simulated environment, therefore in the case of purely visual rotations, the participant is not physically rotated. In each of the modalities, the rotation corresponded to the same smooth off-axis yaw rotation performed around the table's vertical axis. The rotation amplitude was always  $50^\circ$  to the right and the duration was 5 s using a raised-cosine velocity profile. The body rotations around the table were performed via a rotation of the platform that stimulated the vestibular system (canals and otoliths) and to a lesser extent proprioception from the inertial forces applied to the body. The visual scene rotation

corresponded to the rotation of the viewpoint in the virtual environment. This environment was displayed on the back screen and consisted of a detailed model of a rectangular room (a 2 m wide  $\times$  3 m long indoor space with furniture). The acoustic scene rotation corresponded to the rotation of a church bell ringing during the entire rotation. The bell was outside the virtual room, and the sound was played using a loudspeaker that was placed 30° to the right of participants in the initial position at a height of 1.05 m. Note that with this setup, since the auditory stimulation was static, it was necessarily combined with the body rotation and its contribution was not assessed independently. At the end of the trials involving body motion, the repositioning of the motion platform to the starting position was performed using a trapezoidal velocity profile with a maximum instant velocity of 10°/s.

#### *The table and the objects*

For the mental rotation task, the same five objects were used to create the spatial layouts (mobile phone, shoe, iron, teddy bear, roll of film, see Fig. 1, right panel). The size and familiarity of these objects were matched. The size was adjusted so as to have equivalent projected surfaces on the horizontal plane and limited discrepancy in height across objects. As for the familiarity, we chose objects from everyday life that could potentially be found on a table. These layouts were displayed on a 21 in. TFT-screen that was embedded in the cylindrical table physically placed in the middle of the cabin and virtually located in the center of the room (see Fig. 1, left picture). Only a disc section of this screen was visible (29 cm diameter). The object configurations were generated automatically according to specific rules to avoid overlapping objects. Both the virtual room and the objects on the table were displayed using a passive stereoscopic vision technique based on anaglyphs. Therefore, during the entire experiment, participants wore a pair of red/cyan spectacles in order to correctly filter each image that was to be displayed for each eye. Independently of the participants' rotations, the table and object layout could be virtually rotated while hidden.

#### *The response interface*

The mental rotation task required participants to select the object from the layout that was moved. In order to make the responding as natural and intuitive as possible, we used a touch screen mounted on the top of the table screen allowing participants to pick objects simply by touching the desired object with the index finger of their dominant hand. The name of the selected object was displayed on the table screen at the end of the response phase, in order to provide a possible control for errors in

the automatic detection process. Reaction times (RTs) were recorded.

#### *Procedure*

In the present work, we will compare different combinations of sensory cues available during the changes in position of the observer, resulting from two separate experiments. The results will then be compared to those of a previous experiment (Lehmann et al. 2008), in which for some conditions the observer remained in the same position. The latter validated the experimental setup by replicating as close as possible the first experiment of (Wang and Simons 1999). Nonetheless, all conditions will be detailed below and the associated experiment will always be mentioned.

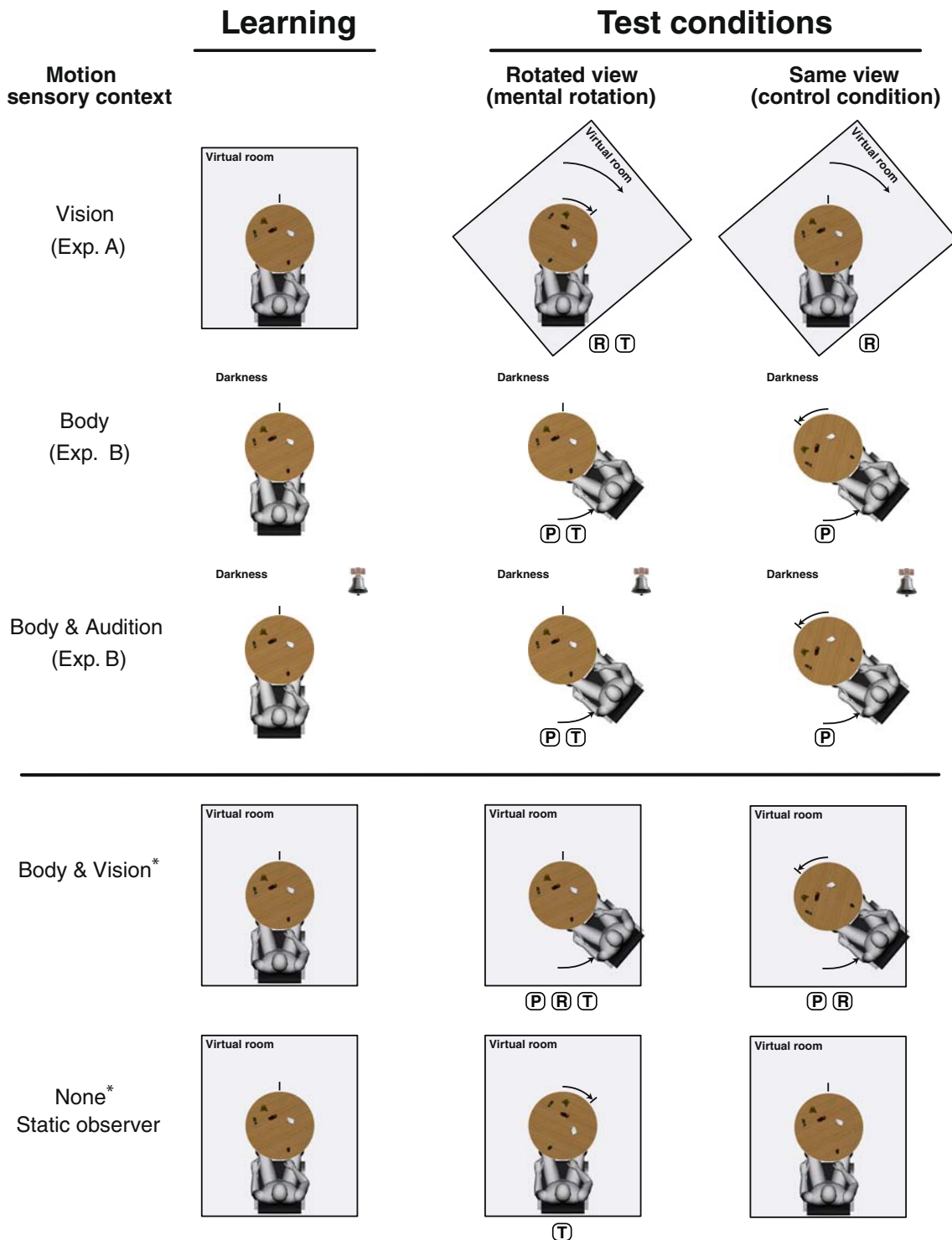
#### *Time-course of a trial*

On each trial, participants viewed a new layout of the five objects on the table for 3 s (learning phase). Then the objects and the table disappeared for 7 s (hidden phase). During this period, participants and the table could rotate independently in the virtual room, and systematically one of the five objects was translated 4 cm in a random direction, insuring that the movement avoided collisions with the other four. The objects and the table were then displayed again, and participants were asked to pick the object they thought had moved.

#### *Experimental conditions*

As illustrated in Fig. 2, the experimental conditions are defined by a combination of two factors: the view of the layout (*same* or *rotated*) and the sensory context available (*Vision*, *Body*, *Body & Audition*, *Body & Vision* and *None*). The changes in position were done during the hidden phase where participants were passively rotated counter-clockwise around the table by 50° to the second viewpoint, with a given sensory context. The description of the conditions will be done in pairs, corresponding to the two possible values of the test view of the layout. Note that this factor is defined according to an egocentric reference frame. In trials with the *same* view, participants were tested with the same view of the table as in the learning phase; therefore, the table was rotated so as to compensate for the observer rotation. In trials with *rotated* view, they were tested with a clockwise 50° orientation change resulting from the rotation of the observer around the table.

Experiment A addressed the role of visual cues (conditions in the *Vision* row of Fig. 2). There was a pair of experimental conditions in which participants were provided with visual information about their rotation by means of the



**Fig. 2** Illustration of the experimental conditions according to different simulated self-motion sensory contexts (consistent manipulations of body physical position, visual orientation in the virtual room and external sound source). *P*, *R* and *T* indicate the technical manipulations detailed in Fig. 1 involved in each condition. The first column shows

the learning context while the other two show the corresponding test conditions with an egocentric rotation of the layout's view (5 mental rotation conditions) or not (5 control conditions). The two asterisked sensory contexts in the bottom were studied in the validation experiment published elsewhere (Lehmann et al. 2008)

projected room rotation in the front screen, and no physical motion of the platform. Experiment B addressed the role of auditory and vestibular cues (conditions in the *Body* and

*Body & Audition* rows). The two pairs of conditions were either a pure vestibular rotation, or the vestibular rotation coupled with an acoustic cue, always in the dark. The

acoustic cue was a static external sound source, a church bell played through a loudspeaker (see Fig. 1), thus rotating in the egocentric reference frame.

The validation experiment published elsewhere addressed on the one hand the role of visio-vestibular cues (conditions in the *Body & Vision* row) when participant's position was changed, and on the other hand the baseline performance when the participant remained in the same position and only the table rotated (conditions in the *None* row). In the latter condition, there was no sensory cue about the participant rotation, though they were informed about the table rotation. In the four conditions involved, the visual environment was visible and could rotate accordingly with the platform in the case of a change in position.

### Experimental design

In all the experiments, each condition was tested 20 times. In order to avoid both the difficulties in switching from one condition to another, and possible order effects, trials were partially blocked within conditions and their order was counterbalanced both within and across participants using a nested Latin-Square design (as in Wang and Simons 1999). The rank of the Latin-Square corresponded to the number of experimental conditions that a given participant was tested on (2 or 4 for experiments A or B, respectively). Trials were arranged into blocks (10 or 20 trials) where all conditions were tested for five successive trials. The orders of the conditions within these blocks were created using a Latin-Squares design, and each participant experienced all of these blocks. Finally, the order of these blocks was counterbalanced across participants, also using a Latin-Squares design. At the beginning of each block, participants were informed of the condition with a text message displayed on the front projection screen. This defined whether and how the view of the layout would change ("The table will rotate" or "You will rotate around the table") or remain the same ("Nothing will rotate" or "You and the table will rotate"). Experiments A and B lasted approximately 45 and 75 min, respectively.

### Data analysis

The statistical analyses of the accuracy and RT were done using two kinds of repeated measures ANOVA designs, according to whether the comparison across sensory contexts were within or between participants from the three experiments (A, B and validation). The independent variables were the view of the layout (same or different) and the sensory context, the latter being either a within participants factor or a group factor (between participants). For each participant, accuracy and RT costs of the mental rotations in a given motion sensory context were computed as the

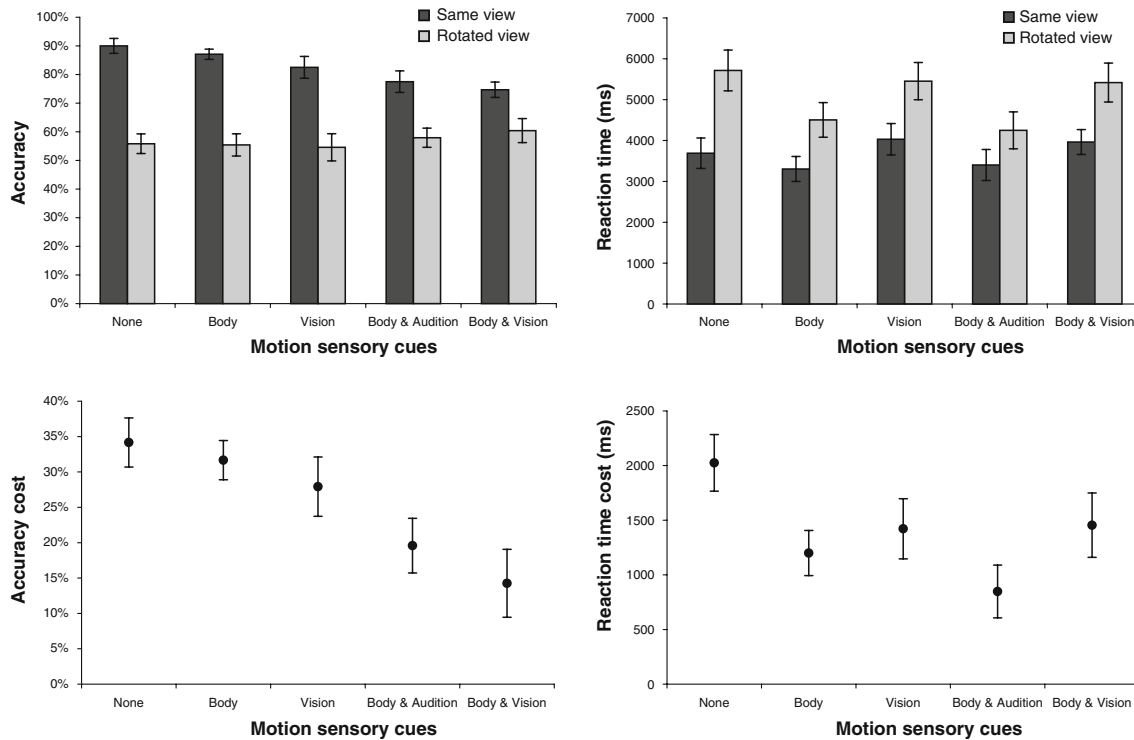
difference between trials where the view of the layout was the same and trials where it was rotated. Costs were introduced to facilitate the presentation of the results, as comparing them across sensory context is statistically equivalent to the analysis of the interaction between the two independent factors (view of the layout  $\times$  sensory context). Note that for RTs, the cost signs were inverted for consistency.

### Results

Performance in this task has consequences on two observable variables: the accuracy in the moved object detection and the RT. Accuracy is more related to the efficacy of the mechanisms involved, whereas RT provides an indication of the cognitive processing time. The accuracy chance level for this task was 20%, therefore even in the worst condition, participants performed well above. The accuracy and RT measured in experiments A and B, as well as baseline data from the previous validation experiment, are presented in Fig. 3 (top plots) together with the associated sensory context costs (bottom plots).

The accuracy cost when only the table rotated without motion cues (*None*: 34%) was not statistically different than with vestibular cues [*Body*: 32%,  $F(1,22) = 0.32$ ;  $P = 0.6$ ] or with visual cues [*Vision*: 28%,  $F(1,22) = 1.32$ ;  $P = 0.26$ ]. Nevertheless, the associated RT costs were significantly shorter of 830 ms for *Body* [ $F(1,22) = 6.2$ ;  $P < 0.02$ ] and marginally shorter of 600 ms for *Vision* [ $F(1,22) = 2.6$ ;  $P = 0.12$ ]. When two modalities were available, the accuracy cost was statistically lower than without motion cues, for both *Body & Audition* [20%,  $F(1,22) = 7.9$ ;  $P < 0.01$ ] and *Body & Vision* [14%,  $F(1,11) = 12.5$ ;  $P < 0.005$ ], and the RT cost was significantly shorter for both *Body & Audition* of 1,180 ms [ $F(1,22) = 11.1$ ;  $P < 0.005$ ] and *Body & Vision* of 570 ms [ $F(1,11) = 6.7$ ;  $P < 0.03$ ]. Planned comparisons showed that the significant improvement in the performance detailed above stem from statistical differences when the layout view is the same (control conditions) for the accuracy costs of *Body & Audition* ( $P < 0.02$ ) and *Body & Vision* ( $P < 0.001$ ), whereas they stem from statistical differences when the layout view is rotated (mental rotations) for the RT costs of *Body & Audition* ( $P < 0.04$ ), and marginally for *Body* ( $P = 0.078$ ) and *Body & Vision* ( $P = 0.083$ ). These results indicate that when participants were provided with unimodal sensory contexts, spatial updating allowed for shorter processing times although the mechanism was not more efficient. In turn, both accuracy and RTs improved when participants were stimulated by a combination of two modalities.

The accuracy costs of the studied bimodal sensory contexts were always statistically different than each associated



**Fig. 3** The mental rotation task performance plotted together with the results from the previous experiment. The average accuracy and reaction times as a function of the change in layout view (*top plots*), and the corresponding mental rotation costs (*bottom plots*), for the various sen-

sory combination contexts: *Body*, *Vision* and *Body & Audition* (from the current experiments), *None* and *Body & Vision* (from the previous validation experiment). The *error bars* correspond to the inter-individual standard error

unimodal contexts: *Body & Audition* versus *Body* [ $F(1,11) = 13.1$ ;  $P < 0.005$ ]; *Body & Vision* versus *Body* [ $F(1,22) = 9.8$ ;  $P < 0.005$ ]; *Body & Vision* versus *Vision* [ $F(1,22) = 4.6$ ;  $P < 0.05$ ]. Here, again these differences stemmed rather from the conditions in which participants were tested with the same layout (with  $P < 0.04$ ,  $P < 0.001$  and  $P = 0.11$ , respectively). None of these effects was found in the analyses of the RT.

## General discussion

To summarize the results, first the performance improved for unimodal sensory contexts as indicated by the shorter processing times (RT), as compared to when the participants remained in the same position and only the table was rotated. Second, with bimodal sensory contexts, the performance improved in efficacy (accuracy) as compared to unimodal as well as static situations. Third, the processing times were not different between unimodal and bimodal situations, but were shorter than when participants were not moved. Fourth, although in terms of efficacy, the performance improve stemmed from the conditions in which participants were tested with the same view of the layout, the shorter processing times resulted from the conditions in which participants were tested with rotated views.

## Interpreting our results in a sensory-based framework

Taken together, these findings allow concluding that there is a gradual efficacy increase of the spatial updating mechanism underlying mental rotations when adding sensory cues to the moving observer, but that the processing time drops to the same level for any of the sensory contexts. This is consistent with the idea that the updating is an ongoing process, which takes the same amount of time independently of the modalities available. This suggests that it is always the same spatial updating mechanism at stake during self-motion, which is sensory-independent but with additive effects when different sensory modalities are consistently co-activated. Note that this improvement of spatial updating should not be considered as quantitatively dependent on the number of sensory modalities stimulated, but rather on the global richness of the cues involved. This predicts a minimal cost in an ecological environment that would be maximally rich in terms of relevant sensory cues available, which is in line with the previous results obtained with real setups that showed a negative cost (Simons and Wang 1998). The fact that our experimental platform uses mixed elements of a physical and virtual setup allowed for innovative multisensory stimulations, despite the limitations in terms of ecological validity. Indeed, while previous apparatus have offered limited (real setups) to null (imagined)

assessment of the modality contributions, we could manipulate independently a large set of sensory combinations involved when an observer is moving.

The findings reported here provide strong evidence that spatial updating of a memorized object layout requires a certain amount of continuous motion-related information in order to bind cognitively the layout with the change in viewpoint. We showed that this binding could be performed efficiently if a minimal sensory context provides the observer with information about his own movements. Indeed, visual information alone was sufficient to exhibit an advantage for moving observers, and we believe that with maximally rich visual information (full field of view and a natural environment) this improvement could have been observed in the accuracy also. Moreover, combining the visual modality with vestibular information also improves the performance, which again indicates that vision plays a role in the mental rotation task when resulting from changes in observer position. These findings clearly contradict the claims of (Simons et al. 2002). Similarly, combining an auditory cue with the vestibular stimulation also resulted in a significant improvement of mental rotations, showing the possibility to use stable acoustic landmarks for the spatial updating mechanism.

#### Inhibition of the automatic spatial updating process

Finally, we should comment why, in terms of spatial updating efficiency only, the accuracy cost reduction observed for each sensory context seems to be produced by the decrease in performance of the control conditions (same view), and to a lesser extent by the improvement of the mental rotation conditions (rotated view). Note that it was the contrary for the drop in processing times. In fact, the spatial updating mechanism being automatic, in four of the control conditions it must be inhibited in order to perform the task with the same view of the layout. Otherwise the expected view would not match the tested view, even though it is precisely the learned view. Still considering the sensory-based framework, it is easy to understand that the richer the sensory cues about the rotation of the observer are, the harder will be preventing the automatic spatial updating mechanism from working. In fact, the inhibition of the updating is only effective when little or no sensory cues about self-motion are present. This is exactly what we observed in our experiment. It explains why the control conditions get impaired, and the one with no motion cue at all gets the best performance. Conversely, the better the motion cueing is, the best spatial updating performance will be, resulting in an improved mental rotation performance. As commented before, in our setup the visual and the body stimulations could have been richer, which lead to not as improved mental rotation performance as one could expect.

Nevertheless, what really matters is the global effect of the sensory context on the cost of the mental rotation, which is characterized by the difference between the mental rotation condition and the control condition within the *same* sensory context.

#### Accounting for previous findings

Interestingly, our sensory-based framework also accounts for the wide range of cost reduction that has been previously reported in the literature. Indeed, in a real setup where all cues are naturally available, this cost can even become negative (Wang and Simons 1999), which means that for moving observers, it is easier to perform a mental rotation than to have to keep the same layout view in memory. This result can be explained by the great difficulty of preventing the automatic updating process when the observer is moving, which is necessary if the table also rotates in order to preserve the same view. As discussed in the previous section, the inhibition of the updating is only effective when little or no sensory cues about self-motion are present, again showing the link with the sensory richness. Nevertheless, in the previous experiments the mental rotation cost for moving observers was also reduced when the sensory context was not maximal, such as passively moved observers (Wang and Simons 1999) or provided with no visual background (Simons and Wang 1998). According to our framework, the amount of mental cost reduction depends on the richness of cues provided by the diverse experimental setups and conditions. It is thus even possible to extend it to imagined rotation studies; the scenario given to the participants and their vividness of imagination modulated the strength of the effect (Amorim and Stucchi 1997; Wraga et al. 1999, 2000).

When it comes to the effect of the manipulation of the visual background, the interpretations provided by different authors has not always been consistent. Some claim that the updating of object structures for moving observers is mediated by extra-retinal cues alone, and that visual cues provided by the background do not contribute (Simons et al. 2002). Our findings contradicted to some extent this conclusion. Although we could not provide evidence that visual cues alone can lead to a significant cost reduction—for reasons discussed previously—we found that together with the vestibular and somatosensory cues, the facilitation becomes significant. The fact that individual cues alone are not sufficient to properly enhance mental rotations, shows that visual information contributes to the quality of the updating mechanism. As we explained before, we believe that the manipulation of the visual cues in the study of Simons et al. was not convincing. Indeed, only visual snapshots of before and after the rotation were provided, with no dynamic information about the observer's change in



viewpoint. According to our sensory interpretation of the viewer advantage effect, it is not surprising that they could not find significant improvement in the mental rotation task with such a limited “amount” of sensory information and continuity about the rotation.

#### Allocentric encoding versus additional sensory cue

In a recent study using a similar paradigm, Burgess et al. (2004) introduced a phosphorescent cue card providing a landmark external to the array. This cue card could be congruent with an egocentric reference frame (i.e., move with the observer) or with an allocentric reference frame (i.e., remain static in the room). They independently manipulated three factors (subject viewpoint S, table T and cue card C) in order to test three different types of representations of object locations based either on visual snapshots, or egocentric representations that can be updated by self-motion, or allocentric representations that are relative to the external cue. They found that the consistency with the cue card also allowed for improved performance, although less markedly than for static snapshot recognition or for the updating of an egocentric representation. They conclude that part of the effect attributed to egocentric updating by Wang and Simons can be explained by the use of allocentric spatial representations provided by the room.

In our view, there is a confound between the level of spatial knowledge representation and the level of spatial information processing. Indeed, speaking of visual snapshots defines one way of storing spatial relationships using mental imagery, therefore stands at the representational level. Recognition tasks relying on this type of storage are more efficient when the view is not changed, as detailed in the introduction, showing its intrinsic egocentric nature. In contrast, speaking of representations updated by self-motion involves both an egocentric storage (representational level), which might well be visual snapshots after all, and the spatial updating mechanism (processing level), which allows making predictions about the outcome after a change in viewpoint, provided that continuous spatial information is available. Finally, we believe that for these table-top mental rotation tasks, the use of extrinsic allocentric cues enabling the encoding of object-to-environment spatial relationships may not be very efficient given the short presentation delays and the distance between the table and the possible external cues. Instead, intrinsic object-to-object relationships within the layout are probably encoded. Indeed, learning intrinsic directions improves recognition (Mou et al. 2008a), and the movement of an object within the layout is more likely detected when the other objects are stationary than when they move (Mou et al. 2008b). Finally, the updating of such spatial structure during locomotion would rely on an egocentric process in which the

global structure orientation is updated in a self-to-object manner. In other words, locomotion would allow keeping track of the layout’s intrinsic “reference direction”, as described by Mou et al. (2009). Nonetheless, these reference directions are not necessarily allocentric (e.g. related to the environment).

The multisensory approach introduced in this paper brings another explanation to the cue card effects reported and the so-called use of allocentric representations. One can distinguish two types of self-motion cues that can be processed in order to update spatial features, those provided by internal mechanisms (idiothetic information, as defined by Mittelstaedt and Mittelstaedt 1980), and those coming from external signals (allothetic information). On one hand, since idiothetic information includes signals from the vestibular and somatosensory systems and efferent copies of motor commands, they are intrinsically egocentric. On the other hand, allothetic information provided for instance by optic or acoustic flow must rely on stable sources of the environment in order to be efficient. These sources are then allocentric in nature. Indeed, without a stable world, a moving observer could not use static sources in order to correctly integrate their egocentric optic flow. Therefore, the cue card introduced in the experiment of Burgess et al. (2004) can be considered as nothing more than a stable visual landmark that contributes to the perception of self-motion and provides an additional sensory cue to the updating mechanism. Accordingly, let us compare the resulting mental rotation costs when adding this visual cue to the sensory context of the observer’s motion around the table. We find a considerable improvement when adding this visual cue ( $ST - S \approx -12\%$ , stable cue card) as compared to the body cues alone ( $STC - SC \approx -2\%$ , moving cue card).

#### Conclusion

The original contribution of this work is to have directly assessed the influence of sensory cue richness on the quality of the spatial updating mechanism involved when predicting the outcome of a spatial layout after the observer’s position change. We found that the sensory contribution to the egocentric updating mechanism in table-top mental rotation tasks is additive: the richer the sensory context during observer motion is, the better mental rotations performance will be. Finally, we showed that this multisensory approach can account for most of the findings reported previously using similar tasks. In particular, it provides an alternate interpretation to the allocentric contribution introduced by Burgess et al. (2004). Indeed, there is a more ecological explanation based on the multisensory redundancy processed by our brains, in order to efficiently update information while moving.

**Acknowledgments** This work was presented at the IMRF 2008 conference. Manuel Vidal received a post-doctoral scholarship from the Max Planck Society and Alexandre Lehmann received a doctoral scholarship from the Centre Nationale pour la Recherche Scientifique. We are grateful to the workshop of the Max Planck Institute for the construction of the table set-up.

## References

- Amorim MA, Stucchi N (1997) Viewer- and object-centered mental explorations of an imagined environment are not equivalent. *Brain Res Cogn Brain Res* 5:229–239
- Amorim MA, Glasauer S, Corpinot K, Berthoz A (1997) Updating an object's orientation and location during nonvisual navigation: a comparison between two processing modes. *Percept Psychophys* 59:404–418
- Burgess N, Spiers HJ, Paleologou E (2004) Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences on spatial memory. *Cognition* 94:149–166
- Christou C, Bühlhoff HH (1999) The perception of spatial layout in a virtual world, vol 75. Max Planck Institute Technical Report, Tübingen, Germany
- Christou CG, Tjan BS, Bühlhoff HH (2003) Extrinsic cues aid shape recognition from novel viewpoints. *J Vis* 3:183–198
- Diwadkar VA, McNamara TP (1997) Viewpoint dependence in scene recognition. *Psychol Sci* 8:302–307
- Gibson JJ (1950) *The perception of the visual world*. Houghton Mifflin, Boston
- Klatzky RL, Loomis JM, Beall AC, Chance SS, Golledge RG (1998) Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychol Sci* 9:293–298
- Lehmann A, Vidal M, Bühlhoff HH (2008) A high-end virtual reality setup for the study of mental rotations. *Presence Teleoperators Virtual Environ* 17:365–375
- Loomis JM, Klatzky RL, Golledge RG, Philbeck JW (1999) Human navigation by path integration. In: *Wayfinding behavior*. The John Hopkins University Press, Baltimore, pp 125–151
- Mittelstaedt ML, Mittelstaedt H (1980) Homing by path integration in a mammal. *Naturwissenschaften* 67:566–567
- Motes MA, Finlay CA, Kozhevnikov M (2006) Scene recognition following locomotion around a scene. *Perception* 35:1507–1520
- Mou W, McNamara TP, Rump B, Xiao C (2006) Roles of egocentric and allocentric spatial representations in locomotion and reorientation. *J Exp Psychol Learn Mem Cogn* 32:1274–1290
- Mou W, Fan Y, McNamara TP, Owen CB (2008a) Intrinsic frames of reference and egocentric viewpoints in scene recognition. *Cognition* 106:750–769
- Mou W, Xiao C, McNamara TP (2008b) Reference directions and reference objects in spatial memory of a briefly viewed layout. *Cognition* 108:136–154
- Mou W, Zhang H, McNamara TP (2009) Novel-view scene recognition relies on identifying spatial reference directions. *Cognition* 111:175–186
- Newell FN, Woods AT, Mernagh M, Bühlhoff HH (2005) Visual, haptic and crossmodal recognition of scenes. *Exp Brain Res* 161:233–242
- Pasqualotto A, Finucane CM, Newell FN (2005) Visual and haptic representations of scenes are updated with observer movement. *Exp Brain Res* 166:481–488
- Rieser JJ (1989) Access to knowledge of spatial structure at novel points of observation. *J Exp Psychol Learn Mem Cogn* 6:1157–1165
- Shepard RN, Metzler J (1971) Mental rotation of three-dimensional objects. *Science* 171:701–703
- Simons DJ, Wang RF (1998) Perceiving real-world viewpoint changes. *Psychol Sci* 9:315–320
- Simons DJ, Wang RXF, Roddenberry D (2002) Object recognition is mediated by extraretinal information. *Percept Psychophys* 64:521–530
- Wang RXF, Simons DJ (1999) Active and passive scene recognition across views. *Cognition* 70:191–210
- Wang RF, Spelke ES (2000) Updating egocentric representations in human navigation. *Cognition* 77:215–250
- Wang RF, Spelke ES (2002) Human spatial representation: insights from animals. *Trends Cogn Sci* 6:376–382
- Wraga M, Creem SH, Proffitt DR (1999) The influence of spatial reference frames on imagined object- and viewer rotations. *Acta Psychol (Amst)* 102:247–264
- Wraga M, Creem SH, Proffitt DR (2000) Updating displays after imagined object and viewer rotations. *J Exp Psychol Learn Mem Cogn* 26:151–168
- Wraga M, Creem-Regehr SH, Proffitt DR (2004) Spatial updating of virtual displays during self- and display rotation. *Mem Cognit* 32:399–415