

Cross-modal visuo-haptic mental rotation: comparing objects between senses

Robert Volcic · Maarten W. A. Wijntjes ·
Erik C. Kool · Astrid M. L. Kappers

Received: 10 February 2010 / Accepted: 9 April 2010 / Published online: 1 May 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract The simple experience of a coherent percept while looking and touching an object conceals an intriguing issue: different senses encode and compare information in different modality-specific reference frames. We addressed this problem in a cross-modal visuo-haptic mental rotation task. Two objects in various orientations were presented at the same spatial location, one visually and one haptically. Participants had to identify the objects as same or different. The relative angle between viewing direction and hand orientation was manipulated (Aligned versus Orthogonal). In an additional condition (Delay), a temporal delay was introduced between haptic and visual explorations while the viewing direction and the hand orientation were orthogonal to each other. Whereas the phase shift of the response time function was close to 0° in the Aligned condition, we observed a consistent phase shift in the hand's direction in the Orthogonal condition. A phase shift, although reduced, was also found in the Delay condition. Counterintuitively, these results mean that seen and touched objects do not need to be physically aligned for optimal performance to occur. The present results suggest that the information about an object is acquired in separate visual and hand-centered reference frames, which

directly influence each other and which combine in a time-dependent manner.

Keywords Cross-modal perception · Touch · Vision · Frames of reference · Mental rotation · Hand

Introduction

The integration of multi-modal information forms our internal representation of the sensory world. Whenever we handle an object, we can effortlessly achieve a coherent percept based on the different visual and haptic sources. We see the object we are touching, and we touch the object we are looking at. This seemingly simple act of perceiving an object conceals, however, some intriguing issues. Both visual and haptic modalities are capable of encoding the coarse information about an object, e.g. its orientation, size and gross shape; however, each modality performs the encoding in its own reference frame first, vision retinotopically and haptics somatotopically. This information needs to be shared by way of translation or comparison between modalities, but only at a later stage. The issue of how the information about objects is shared across modalities is the topic of the current paper.

Humans can effectively compare the shape of 3D objects across the modalities of vision and touch, although cross-modal performance is usually poorer than unimodal performance (Gibson 1962, 1963; Norman et al. 2004, 2008; Phillips et al. 2009). These studies suggest that the two modalities either share a common representation or have independent objects' representations with similar formats for effective comparisons to take place. Several studies have investigated the effect of orientation on the cross-modal identification of 3D objects. Recognition

R. Volcic (✉)
Psychologisches Institut II, Westfälische Wilhelms-Universität
Münster, Fliednerstr. 21, 48149 Münster, Germany
e-mail: volcic@uni-muenster.de

M. W. A. Wijntjes
Faculty of Industrial Design Engineering,
Delft University of Technology, Delft, The Netherlands

E. C. Kool · A. M. L. Kappers
Physics of Man, Helmholtz Institute, Utrecht University,
Utrecht, The Netherlands

performance across modalities was assessed by means of different experimental methods. In an old/new recognition task and in a forced-choice object recognition task, objects were learned in one modality and then recognition was tested in the other modality (Newell et al. 2001; Lacey et al. 2007, 2009; Ernst et al. 2007). On the other hand, in a sequential matching task, an object presented in one modality was shortly afterwards compared with a test object through the other modality (Lawson 2009). In all these methods, the test objects were presented either in the same or in a different orientation. In some cases and independently from the experimental method, the recognition of objects across modalities had additional costs on the performance (Newell et al. 2001; Ernst et al. 2007; Lawson 2009), whereas in other cases the performance was unaffected irrespective of the change in orientation (Lacey et al. 2007, 2009; Lawson 2009). These studies support the idea that the achievement of object constancy, i.e. the recognition of objects despite changes in size, position and orientation, can thus be fast and accurate in both within-modal and cross-modal object recognition, although there is often an additional cost when objects' representations are compared across modalities. The additional cost might be attributed to the fact that object recognition supposedly relies on features represented in viewpoint- and modality-specific frames of reference.

In all of the foregoing object recognition experiments, relatively long temporal intervals occurred between the presentations of the first and the second stimulus (or set of stimuli). An alternative method for studying the sharing of information between the visual and haptic modalities that minimizes the aforementioned temporal interval is a matching task in which objects are compared simultaneously. A widely used task that satisfies these properties is the handedness recognition task employed in most mental rotation studies since its introduction by Shepard and Metzler (1971). Two objects of the same shape and in different orientations are compared, and the participant has to determine whether the two objects are mirrored versions of each other or identical except for their orientation. In the simplest case, response times are fastest when objects are physically aligned with each other and response times linearly increase as a function of the angular misalignment between objects. The physical alignment of objects does, however, not always induce the fastest response times: the shape and specifically the phase shift of the response time function depends on the reference frame in which the objects are encoded. In vision, retinal and gravitational encodings were contrasted by having the participants perceive the stimuli with the head either in the upright or tilted orientation (Corballis et al. 1976, 1978). The response time function shifted in accordance to the participants' head tilt. The phase shifts were, however, only partial. The stimuli

were encoded in a reference frame that was intermediate between a retinally defined egocentric reference frame and an allocentric reference frame.

The interactions between reference frames were explored also in the haptic domain. In unimanual mental rotation studies, participants were presented with a single letter in a normal or mirror-image form in various orientations (Carpenter and Eisenberg 1978; Prather and Sathian 2002). The orientation of the hand exploring the stimuli was varied and, as a consequence, the response time function partially shifted. In these studies, the haptic information was compared with an internal representation of the stimuli retrieved from memory. More recently, a haptic mental rotation study was conducted, in which two objects were separately but simultaneously explored by the two hands and, by this, haptic information was directly compared (Volcic et al. 2009). Not surprisingly, with hands aligned the fastest response times were measured when also the objects were physically aligned, thus the phase shift was equal to zero. However, when the hands were held in either a convergent or divergent orientation, the response time functions shifted in opposite directions. The condition-dependent directions and extents of the phase shifts suggested an interplay of multiple reference frames, in which the hand-centered reference frame plays the central role.

A natural step from the foregoing is to investigate the interaction of reference frames across the visual and haptic modality. In this context, fundamental questions arise. Does one modality take over from the other and by this provide a unique reference frame in which both visual and haptic information are compared? Or, alternatively, do multiple reference frames coexist simultaneously and interact with each other?

To address these questions, we used a visuo-haptic cross-modal mental rotation task. One of the objects was viewed and the other was haptically explored. Moreover, we designed the setup such that both objects were perceptually located in exactly the same spatial position (see Fig. 1b). The logic behind the present experiment was simple: varying the hand orientation while keeping the viewing direction constant allows the dissociation of the visual reference frame and the hand-centered reference frame (see Fig. 1a). In the Aligned condition, viewing direction and hand orientation were aligned. In the Orthogonal condition, the hand orientation was instead orthogonal to the viewing direction. With this experimental manipulation, the visual reference frame and the hand-centered reference frame were put in misalignment with each other. In an additional condition, the Delay condition, a temporal delay was introduced between exploration of the haptic object and display of the visual one. However, viewing direction and hand orientation were still orthogonal to each other. The latter condition was of interest

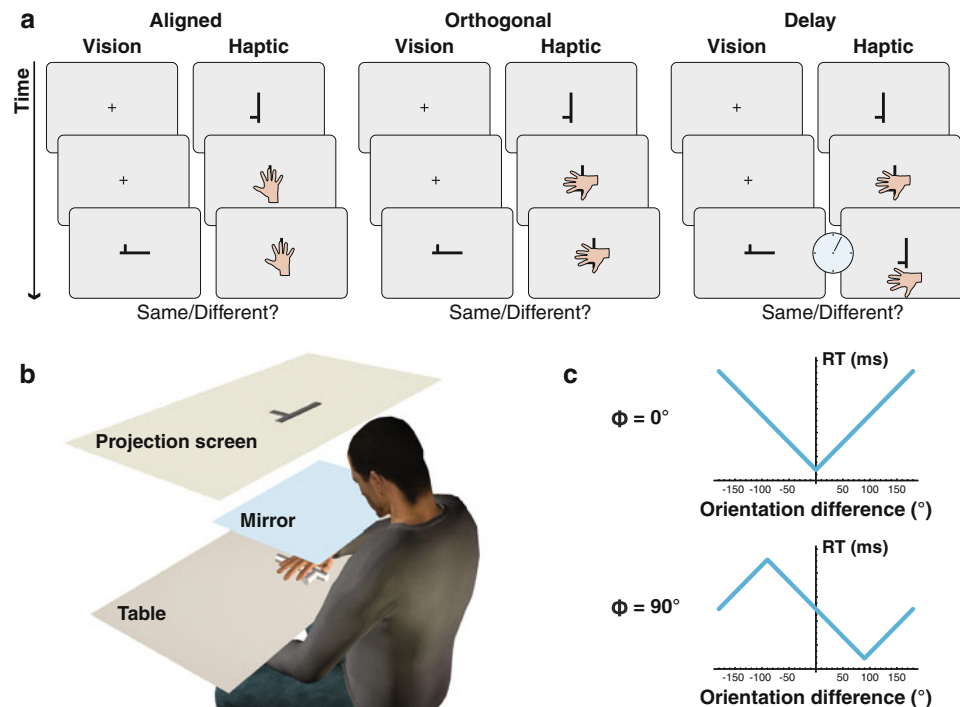


Fig. 1 **a** Experimental conditions of the cross-modal visuo-haptic mental rotation task. In the Aligned condition, viewing direction and hand orientation were aligned. In the Orthogonal condition, the hand was rotated counterclockwise by 90° . In the Delay condition, a 5 s temporal delay was introduced between the haptic and visual exploration of the stimuli. **b** Schematic view of the experimental setup. The participant looked in the mirror, which was positioned midway between the table and the projection screen. The visual stimulus displayed on the

projection screen was seen via the mirror as if it were located on the table exactly in the same location as the haptic stimulus. Both arms were occluded. The right hand explored the haptic stimulus, whereas the left hand controlled the keyboard below the table. **c** Forms of the triangle wave function when the phase shift is $\phi = 0^\circ$ or $\phi = 90^\circ$. Depending on the experimental condition, different hypotheses (see main text) predict that the function might take these forms or a form with an intermediate phase shift ($0^\circ < \phi < 90^\circ$)

because several studies suggest that different frames of reference may dominate at different time intervals (Bridgeman et al. 1997; Carrozzo et al. 2002; Milner et al. 1999; Rossetti et al. 1996; Zuidhoek et al. 2003). Typically, egocentric reference frames prevail at short time intervals and the allocentric frame of reference is strengthened at longer time intervals (in the range of 5–10 s). The temporal delay in the Delay condition should thus reduce the influence (if any) given by the misalignment of the hand-centered reference frame with respect to the visual reference frame.

Hypotheses about object encoding that are based on a single reference frame make straightforward predictions. If all the spatial information is encoded in a single visual reference frame, in a single haptic hand-centered reference frame or in an allocentric reference frame, then no deviation from the zero phase shift will occur in any of the conditions. The fastest responses would be expected to occur when the two objects have the same orientation with respect to the used reference frame, and in all conditions the triangle wave function would take the form depicted in Fig. 1c, $\phi = 0^\circ$. Interestingly, the same prediction is made by a hypothesis based on multiple reference frames, but

only if their interaction is optimal, i.e., the relative orientation of the viewing direction and the hand orientation is taken into account. On the contrary, an interaction of multiple reference frames that discards this proprioceptive information completely, predicts a phase shift in the direction and by the amount specified by the change in hand orientation (Fig. 1c, $\phi = 90^\circ$ in the Aligned condition, $\phi = 90^\circ$ in the Orthogonal condition). An intermediate phase shift ($0^\circ < \phi < 90^\circ$) in the Orthogonal condition will support the hypothesis of multiple interacting reference frames, in which, however, the proprioceptive information is only partially comprised. Any additional effect due to the temporal delay between haptic and visual exploration will be an indication of a time-dependent interaction of reference frames.

Materials and methods

Participants

Ten right-handed male participants took part in this experiment. Three of them are authors of the paper. All the

others were undergraduate students and were paid for their efforts. None of the participants (except the authors) had any prior knowledge of the experimental design and the task. The experiment was performed in accordance with the guidelines from the declaration of Helsinki.

Apparatus and stimuli

The setup consisted of a large horizontal table in the center of which an iron plate (30 × 30 cm) was positioned. The iron plate was covered with a plastic layer on which a protractor was printed. The center of the protractor was 20 cm from the long table edge. Participants were seated on a stool nearby the longest table edge. The 3D objects used as haptic stimuli were made of two cylindrical bars, with a diameter of 1 cm. The main bar had a length of 20 cm, and attached perpendicularly to this at 5 cm from the center was a smaller bar with a length of 5 cm (see Fig. 1b). One pair of objects had the smaller bar attached on the right side of the main bar, whereas the other pair had it attached on the left side. The main bar had an arrow-shaped end on one side that allowed the orientation to be read off with an accuracy of 0.5°. Small magnets were attached under the bar to prevent accidental rotations. Color photographs of the same objects were used as visual stimuli. Visual stimuli were presented as virtual images in the plane of the table. This was achieved by projecting the images of the objects with an LCD projector onto a horizontal rear projection screen suspended 51 cm above the table. A horizontal front-reflecting mirror was placed face up 25.5 cm above the table. Participants viewed the reflected image of the rear projection screen binocularly by looking down in the mirror (see Fig. 1b). By matching the screen-mirror distance to the mirror-table distance, all projected images appeared to be in the plane of the table. The center of the images of visual stimuli was perfectly aligned with the center of haptic stimuli, and the stimuli were matched in size. On a surface 13 cm below the table plane, a keyboard was placed which was used to collect participants' responses. For visual stimuli presentation and data collection, we used MATLAB with Psychtoolbox (Brainard 1997; Pelli 1997).

Stimuli were presented in pairs: one stimulus haptically and one stimulus visually. The haptic stimulus was oriented at 0°, 90°, 180° or 270°. An orientation of 0° is parallel to the long table edge; increasing orientation values signify a rotation in counterclockwise direction. The visual stimulus was presented at 18 different orientations, between 0° and 340°, in steps of 20°. We chose to use more incremental steps with the visual stimulus than with the haptic one, because the haptic stimulus had to be manually adjusted by the experimenter, whereas the

visual stimulus was automatically presented on the screen. Most importantly, it was the relative orientation between haptic and visual objects that was manipulated experimentally. Each stimulus was paired with either another identical stimulus (Same trial) or its mirror version (Different trial).

Stimuli were presented in three different experimental conditions (see Fig. 1a). In the Aligned condition, the main axis of the right hand exploring the haptic stimulus was aligned with the viewing direction. In the Orthogonal condition, the main axis of the right hand was rotated 90° counterclockwise and was thus orthogonal with respect to the viewing direction. In both conditions, haptic and visual stimuli were simultaneously explored. In the Delay condition, the relation between the exploring hand and the viewing direction was the same as in the Orthogonal condition, but it differed with respect to the timing of the visual stimulus presentation. The visual stimulus was presented with a delay of 5 s after participants stopped exploring the haptic stimulus.

In total, each participant completed 864 trials (2 objects × 4 orientations of the haptic object × 18 orientations of the visual object × 2 same/different pairs × 3 conditions). The order of trials in each experimental condition was random and different for each participant. The order of the experimental conditions was counterbalanced across participants.

Procedure

Participants had to perform a cross-modal visuo-haptic mental rotation task. The time-line of each condition is represented in Fig. 1a. Before the start of each trial, the experimenter set the haptic stimulus and gave a start signal to the participant. Participants had no direct view of their arm and hand because these were covered by the mirror and a curtain. They were instructed to position their hand in the orientation determined by the experimental condition and touch the haptic stimulus. The time it took to position the hand and identify the distinctive parts was approximately 1 s. By presenting the haptic stimulus before the visual one, we were able to exclude this task-irrelevant exploration time from the computation of the response time. As soon as they had identified the distinctive parts of the stimulus, they pressed a key with their left hand. In the Aligned and in the Orthogonal conditions, the key press made the projector display the visual stimulus. During the presentation of the visual stimulus, the right hand was kept in contact with the haptic stimulus. In the Delay condition, the visual stimulus was displayed 5 s after the key press. In this period, participants lifted their hand from the haptic stimulus and repositioned it on the table, but kept it in the same orientation. The timing of the participants' response

started with the presentation of the visual stimulus, that is, simultaneously with the key press in the Aligned and Orthogonal conditions, and 5 s after the key press in the Delay condition. They had then to respond as fast as possible whether the two stimuli were the same or different. The visual stimulus stayed on until participants' response, which terminated the trial. These responses were collected via key presses. It was stressed that the answer should be correct. Participants received feedback on their responses and when an incorrect response was given, the trial was repeated at the end of the experimental condition. Each experimental condition was preceded by practice trials. Experimental sessions ended after one hour to prevent fatigue and participants took on average 3 h to complete all conditions.

In the present study, we performed only the cross-modal conditions, because the main purpose was to allow the participants to explore the haptic and visual objects simultaneously. Objects should have been inevitably presented sequentially in within-modal conditions that would require a change in the viewing direction or hand orientation making the cross-modal and within-modal conditions difficult to compare.

Data analysis

Data analysis was focused on the response times of the Same trials. The analysis of the Different trials does not convey any information since the angle through which the different objects must be rotated to achieve congruence is not defined. The error rates of participants' responses were low (below 10%) and were not further analyzed.

Fitting procedure

Response times on Same trials were grouped separately for each participant, for each condition, and each orientation difference. For each orientation difference, we took the median of the response times. For each participant and each condition, a triangle wave function was then fitted through the data to extract the amplitude, the phase shift and the vertical shift from the response time data (see Volcic et al. 2009). The fit of the triangle wave function was performed by minimizing the sum of squares between the median response times and the model. The triangle wave function is a periodic function with a fixed wave period of 360°. We define it as:

$$T(x, A, \phi, \mu) = 2A \left| \text{Int} \left(\frac{x - \phi}{360^\circ} \right) - \frac{x - \phi}{360^\circ} \right| + \mu - \frac{A}{2} \quad (1)$$

where A is the amplitude, ϕ is the phase shift and μ is the vertical shift. The function $\text{Int}(x)$ gives the integer closest to x .

Results

Figure 2a represents the response times averaged over participants in the Aligned, Orthogonal and Delay conditions. The fitted lines correspond to the triangle wave function. As it is clear from these graphs, the response time functions are very similar in the three conditions except for their phase shifts. The phase shift is associated with the orientation difference between the haptic and the visual object at which the response times are fastest. From that point, response times linearly increase in both positive and negative directions.

To analyze the differences between conditions, we ran separate repeated measures ANOVA on the phase shifts, vertical shifts and amplitudes with the experimental conditions as a factor. The parameters were computed for each participant individually. We also ran the same analyses without including the authors' data, and we obtained the same results (except for a t -test, see below).

The average phase shifts were 5.5°, 37° and 14.4° in the Aligned, Orthogonal and Delay conditions, respectively, and are represented in Fig. 2b. We found a significant effect of condition, $F(2, 18) = 10.507$, $P < 0.001$. Subsequent pair-wise comparisons with Bonferroni corrections showed a significant difference between the Aligned and Orthogonal conditions, $P < 0.005$, and a significant difference between the Orthogonal and Delay conditions, $P < 0.05$. The comparison between the Aligned and Delay conditions was not significant, $P = 0.762$. In addition, we ran three simple t -tests to check which response time functions actually shifted with respect to the reference point defined by zero degree orientation difference. The phase shifts in the Aligned condition were not significantly different from zero ($t(9) = 1.349$, $P = 0.21$). However, the phase shifts did differ significantly from zero in both Orthogonal ($t(9) = 5.787$, $P < 0.001$) and Delay conditions ($t(9) = 2.861$, $P < 0.05$). By excluding the authors' data from the analyses, the phase shifts in the Delay condition did not differ from zero ($t(6) = 1.851$, $P = 0.11$). All the other analyses yielded the same results.

The vertical shifts were 866 ms, 884 ms and 776 ms in the Aligned, Orthogonal and Delay conditions, respectively. No significant effect of condition was found, $F(2, 18) = 1.386$, $P = 0.275$. The amplitudes were 490 ms, 510 ms and 304 ms in the Aligned, Orthogonal and Delay conditions, respectively. The effect of condition was marginally significant, $F(2, 18) = 3.523$, $P = 0.051$. However, the subsequent pair-wise comparisons did not reach significance.

Discussion

There is general agreement that spatial information in different sensory modalities can be encoded in multiple

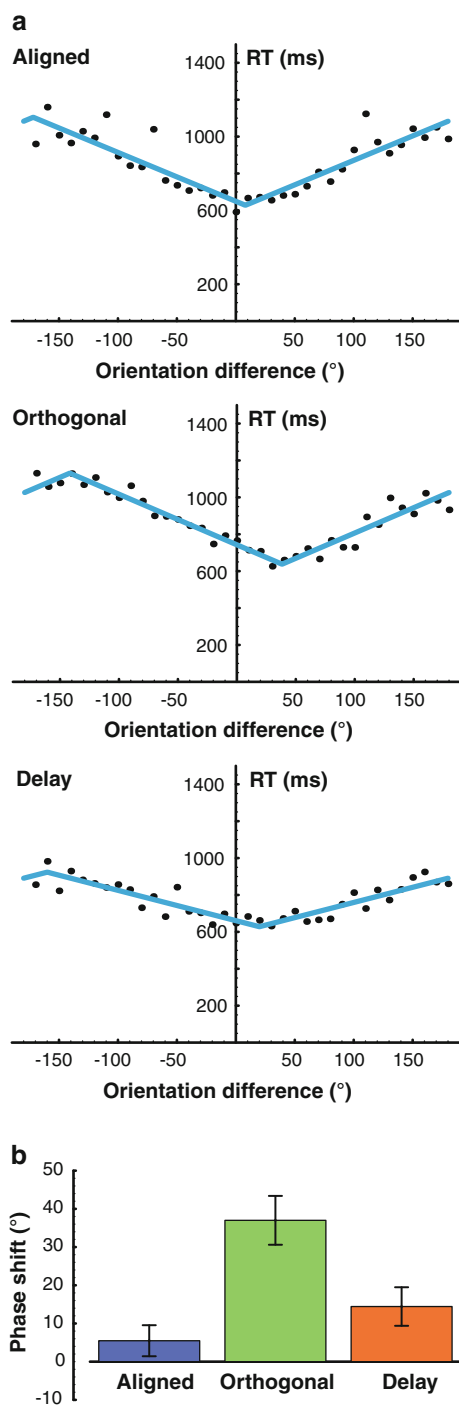


Fig. 2 **a** Response times as a function of the orientation difference averaged over all participants for the Aligned, Orthogonal and Delay conditions. Data are fitted by the triangle wave function. **b** Phase shifts (ϕ) averaged over all participants for the Aligned, Orthogonal and Delay conditions. Error bars indicate the 95% confidence interval of the standard error of the mean

reference frames. A less well-understood problem concerns the interplay of reference frames across modalities. Here, we shed light on how spatial information is encoded in visual and haptic reference frames, and on how these

reference frames interact with each other. In the present cross-modal mental rotation experiment, we found that the response time function shifted in the direction of the misalignment between the viewing direction and the orientation of the exploring hand. However, the phase shift was only partial and was reduced with a longer temporal delay between haptic and visual explorations. These phase shifts indicate that, contrary to common sense, the haptic and visual objects do not need to be physically aligned with each other to be quickly identified as being the same.

The hypotheses involving a single reference frame in the encoding of the objects can be discarded on the basis of the present results, because they did not predict any phase shift. If both visual and haptic spatial information were encoded in a unique reference frame and the common reference frame was, for instance, visual, then the response time functions would have been invariant to the orientation of the hand. The same holds for the haptic hand-centered reference frame and for the allocentric reference frame.

The alternative hypothesis postulates the interaction of multiple reference frames. However, this interaction could take different forms. One alternative would suggest a translation process either from one modality into the other or from both modalities into a multi-modal format. Another alternative would presuppose a direct comparison of the encoded information across modalities. Unfortunately, we cannot distinguish between any of these possibilities, because they do not strictly exclude each other, and they are most likely closely intertwined. Nevertheless, it is clear that each modality encodes the spatial representation in its own frame of reference and it is the interplay of these frames of reference that gives rise to the phase shifts observed in the present study.

An optimal spatial mapping between the visual retinotopic information and the tactile somatotopic information should also comprise the proprioceptive information about the current hand posture. This was clearly not the case. The hand's misalignment with respect to the viewing direction actually induced the phase shift of the response time function. It evidently follows that the proprioceptive information was largely ignored. Interestingly, since the effect on the phase shift was reduced after the temporal delay, we might presuppose that the proprioceptive information was partially comprised at a later stage only. The spatial information in the hand-centered reference frame in combination with the proprioceptive information about one's hand posture and one's body position in space constitutes the necessary information for the construction of an allocentric spatial representation. The temporal delay might have induced the recoding of egocentric spatial information into an allocentric reference frame, which led to the decrement in the phase shift. It should be noted here that, given that the phase shift was still biased by the

orthogonal orientation of the hand, it is necessary to interpret the effect as the result of an interaction of different reference frames. This outcome is in line with previously reported results (Bridgeman et al. 1997; Carrozzo et al. 2002; Milner et al. 1999; Rossetti et al. 1996; Zuidhoek et al. 2003).

Previous unimodal mental rotation studies reported a substantial reference frame influence on the way spatial information is encoded and compared, both in vision and in haptics (Carpenter and Eisenberg 1978; Corballis et al. 1976, 1978; Prather and Sathian 2002; Volcic et al. 2009). Here, we present the novel finding that spatial orientational information is encoded within modality-specific reference frames and that performance in a visuo-haptic cross-modal mental rotation task is bound to the relative alignments of reference frames and their interactions. In addition, the intervening temporal delay is presumed to have affected the integration of the proprioceptive information. Although mental rotation and recognition of rotated objects show behaviorally similar effects, they rely on different processes (e.g., Gauthier et al. 2002). However, the present findings can make an important contribution to the general discussion about how the information about objects is shared across modalities. In this respect, we might tentatively hypothesize that the view-dependence/independence effects in cross-modal object recognition could depend on the solution of a conflict between modality-specific reference frames.

Acknowledgments This research was supported by a grant from the Netherlands Organisation for Scientific Research (NWO) and a grant from the EU (FP7-ICT-217077-Eyeshots).

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

- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436
- Bridgeman B, Peery S, Anand S (1997) Interaction of cognitive and sensorimotor maps of visual space. *Percept Psychophys* 59:456–469
- Carpenter PA, Eisenberg P (1978) Mental rotation and the frame of reference in blind and sighted individuals. *Percept Psychophys* 23:117–124
- Carrozzo M, Stratta F, McIntyre J, Lacquaniti F (2002) Cognitive allocentric representations of visual space shape pointing errors. *Exp Brain Res* 147:426–436
- Corballis MC, Zbrodoff J, Roldan C (1976) What's up in mental rotation? *Percept Psychophys* 19:525–530
- Corballis MC, Nagourney BA, Shetzer LI, Stefanatos G (1978) Mental rotation under head tilt: factors influencing the location of the subjective reference frame. *Percept Psychophys* 24:263–273
- Ernst MO, Lange C, Newell FN (2007) Multisensory recognition of actively explored objects. *Can J Exp Psychol* 61:242–253
- Gauthier I, Hayward WG, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (2002) Bold activity during mental rotation and viewpoint-dependent object recognition. *Neuron* 34:161–171
- Gibson JJ (1962) Observations on active touch. *Psychol Rev* 69:477–491
- Gibson JJ (1963) The useful dimensions of sensitivity. *Am Psychol* 18:1–15
- Lacey S, Peters A, Sathian K (2007) Cross-modal object recognition is viewpoint-independent. *PLoS ONE* 2:e890
- Lacey S, Pappas M, Kreps A, Lee K, Sathian K (2009) Perceptual learning of view-independence in visuo-haptic object representations. *Exp Brain Res* 198:329–337
- Lawson R (2009) A comparison of the effects of depth rotation on visual and haptic three-dimensional object recognition. *J Exp Psychol Human* 35:911–930
- Milner AD, Paulignan Y, Dijkerman HC, Michel F, Jeannerod M (1999) A paradoxical improvement of misreaching in optic ataxia: new evidence for two separate neural systems for visual localization. *Proc Biol Sci* 266:2225–2229
- Newell FN, Ernst MO, Tjan BS, Bühlhoff HH (2001) Viewpoint dependence in visual and haptic object recognition. *Psychol Sci* 12:37–42
- Norman JF, Norman HF, Clayton AM, Lianekhammy J, Zielke G (2004) The visual and haptic perception of natural object shape. *Percept Psychophys* 66:342–351
- Norman JF, Clayton AM, Norman HF, Crabtree CE (2008) Learning to perceive differences in solid shape through vision and touch. *Perception* 37:185–196
- Pelli DG (1997) The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* 10:437–442
- Phillips F, Egan EJJ, Perry BN (2009) Perceptual equivalence between vision and touch is complexity dependent. *Acta Psychol* 132:259–266
- Prather SC, Sathian K (2002) Mental rotation of tactile stimuli. *Cognitive Brain Res* 14:91–98
- Rossetti Y, Gaunet F, Thinus-Blanc C (1996) Early visual experience affects memorization and spatial representation of proprioceptive targets. *Neuroreport* 7:1219–1223
- Shepard RN, Metzler J (1971) Mental rotation of three-dimensional objects. *Science* 171:701–703
- Volcic R, Wijntjes MWA, Kappers AML (2009) Haptic mental rotation revisited: multiple reference frame dependence. *Acta Psychol* 130:251–259
- Zuidhoek S, Kappers AML, van der Lubbe RHJ, Postma A (2003) Delay improves performance on a haptic spatial matching task. *Exp Brain Res* 149:320–330