RESEARCH ARTICLE

Perceptual learning of view-independence in visuo-haptic object representations

**Simon Lacey · Marisa Pappas · Alexandra Kreps ·
Kevin Lee · K. Sathian**

Abstract We previously showed that cross-modal recognition of unfamiliar objects is view-independent, in contrast to view-dependence within-modally, in both vision and haptics. Does the view-independent, bisensory representation underlying cross-modal recognition arise from integration of unisensory, view-dependent representations or intermediate, unisensory but view-independent representations? Two psychophysical experiments sought to distinguish between these alternative models. In both experiments, participants began from baseline, within-modal, view-dependence for object recognition in both vision and haptics. The first experiment induced within-modal view-independence by perceptual learning, which was completely and symmetrically transferred cross-modally: visual view-independence acquired through visual learning also resulted in haptic view-independence and vice versa. In the second experiment, both visual and haptic view-dependence were transformed to view-independence by either haptic-visual or visual-haptic cross-modal learning. We conclude that cross-modal view-independence fits with a model in which unisensory view-dependent representations are directly integrated into a bisensory, view-independent representation, rather than via intermediate, unisensory, view-independent representations.

Keywords Multisensory · Bisensory ·
Object recognition · Cross-modal · Vision · Touch

Introduction

We can recognize a vast range of objects using vision and touch, both within- and cross-modally: disentangling the complex of representations involved and their properties is a key problem in multisensory research. A particular issue is whether these representations are view-dependent (i.e., object recognition is degraded if the object is rotated away from the original view) or view-independent (i.e., objects are correctly identified despite being rotated to provide a different view). While much is known about visual representations in this respect, considerably less is known about haptic representations, or multisensory representations to which vision and haptics both contribute.

Within-modal visual object representations are view-dependent (reviewed in Peissig and Tarr 2007). The extent to which object rotation impairs visual recognition may depend on the axis of rotation (Gauthier et al. 2002; Lacey et al. 2007). Picture-plane rotations result in faster and more accurate performance than depth-plane rotations in both object recognition and mental rotation tasks, even though these tasks depend on different visual pathways—ventral and dorsal, respectively (Gauthier et al. 2002). Since the hands can simultaneously contact an object from different sides, it might be expected that haptic object representations are view-independent (Newell et al. 2001), particularly since following the contours of a three-dimensional object is

S. Lacey · M. Pappas · A. Kreps · K. Lee · K. Sathian (✉)
Department of Neurology, School of Medicine, Emory University,
WMB-6000, 101 Woodruff Circle, Atlanta, GA 30322, USA
e-mail: krish.sathian@emory.edu

K. Sathian
Department of Rehabilitation Medicine,
Emory University, Atlanta, GA, USA

K. Sathian
Department of Psychology, Emory University, Atlanta, GA, USA

K. Sathian
Rehabilitation R&D Center of Excellence,
Atlanta VAMC, Decatur, GA, USA

necessary for haptic object recognition (Lederman and Klatzky 1987). However, several studies have shown that haptic object representations are also view-dependent (Newell et al. 2001; Lacey et al. 2007; Lawson 2009). To some extent, this may be because the biomechanical constraints of the hands bias haptic exploration to the far (back) 'view' of an object, explored by the fingers while the thumbs stabilize rather than explore (Newell et al. 2001). However, even when objects are presented in an orientation that allows more comprehensive haptic exploration of multiple object surfaces, haptic recognition remains view-dependent (Lacey et al. 2007). Unlike vision, however, haptic object recognition is independent of the axis of rotation, suggesting that visual and haptic view-dependence may be qualitatively different (Lacey et al. 2007).

In contrast to visual and haptic within-modal object recognition, visuo-haptic cross-modal object recognition is view-independent (Lacey et al. 2007; Ueda and Saiki 2007). Rotating an object away from the original view between study and test did not degrade recognition, whether visual study was followed by haptic test or vice versa (Lacey et al. 2007; Ueda and Saiki 2007), although Lawson (2009) found view-independence only in the haptic study-visual test condition and not in the reverse condition. Cross-modal performance was also unaffected by the axis of rotation (Lacey et al. 2007). Thus, visuo-haptic cross-modal object recognition clearly relies on a different representation from that involved in the corresponding within-modal tasks (see also Newell et al. 2005).

An important question is how this cross-modal view-independence arises. Visual view-independence could result from learning multiple different views of an object (Tarr and Bülthoff 1995; Booth and Rolls 1998; reviewed by Peissig and Tarr 2007). Typically, this is associated with familiar objects for which multiple views have been acquired over time. Alternatively, view-independence can arise when an object has distinct parts that are easily transformed, by mental rotation, to match a new view (Biederman 1987). However, while responses in dorsal visual areas are view-dependent in visual mental rotation (Gauthier et al. 2002), there is a lack of consensus on whether object recognition processes in ventral visual areas are view-dependent (Grill-Spector et al. 1999; Gauthier et al. 2002) or view-independent (James et al. 2002). Thus, the relationship between object representations employed in mental rotation processes and those used for object recognition remains unclear.

It is not clear whether the same principles apply to the acquisition of haptic within-modal view-independence. To our knowledge, the only relevant report is that of Ueda and Saiki (2007). In this study, participants haptically explored objects made from Lego™ bricks in five different views, rotated about the *y*-axis, before being tested with the objects rotated about the *x*-axis. Haptic recognition was view-dependent for participants who were told the test modality but view-independent for participants who were not. Since knowledge of the test modality might have been expected to confer an advantage, the reason for this discrepancy in performance is not known, especially since the textured surface of the Lego™ bricks provided a cue to rotation about the *x*-axis. This report does, however, provide a 'proof of concept' that haptics can integrate multiple views over time and arrive at view-independence.

In a previous study (Lacey et al. 2007), we found view-independence of cross-modal object recognition, using unfamiliar objects that lacked distinctive parts. This suggested that cross-modal view-independence does not require object familiarity or distinctive object parts, in contrast to the ideas advanced for visual view-independence (see above). How, then, might cross-modal view-independence have arisen? The most parsimonious model is a three-representation model of visuo-haptic object recognition (Fig. 1a) in which separate, unisensory, view-dependent representations in each modality feed directly into a higher-level, bisensory (or multisensory) view-independent representation, perhaps by integrating these lower-level, view-dependent representations [analogous to proposals for visual object representations (Riesenhuber and Poggio 1999)]. Although the results of our previous study (Lacey et al. 2007) are consistent with this model, we could not rule out a second explanation that cross-modal view-independence might depend on acquiring some degree of within-modal view-independence. This would suggest separate, unisensory view-independent 'gateways', one for
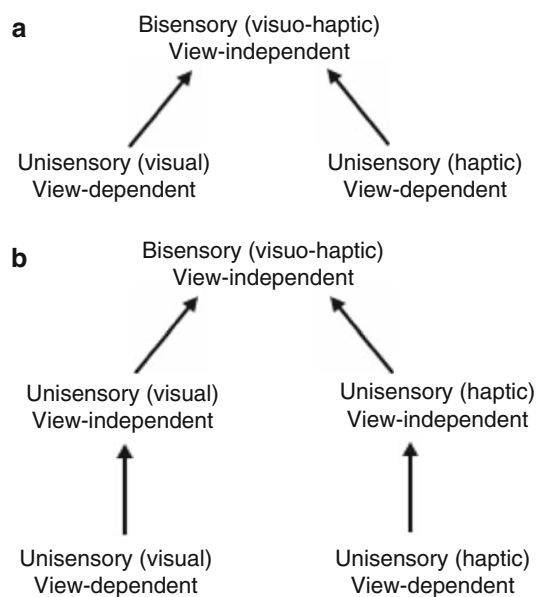
Fig. 1 Two models of bisensory view-independence in which this is either **a** derived directly from unisensory view-dependent representations or **b** gated by separate unisensory view-independent representations

Fig. 2 Example object shown (*left*) unrotated and (*right*) rotated 180° about the *y*-axis
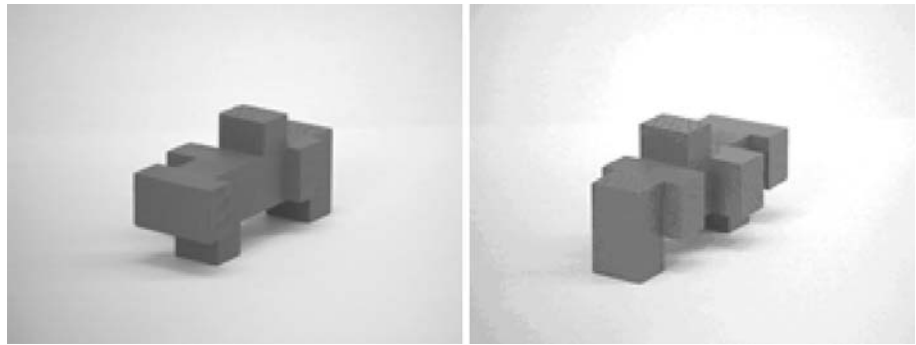
vision and one for haptics, which then feed into the highest level, multisensory view-independent representation, resulting in a five-representation model (Fig. 1b).

Here, we aimed to distinguish between these possibilities, by addressing whether or not there are separate view-independent representations in vision and haptics. Two experiments were conducted, using objects and procedures similar to those used in our earlier study (Lacey et al. 2007). In the first experiment, participants, who were initially view-dependent in both vision and touch, were induced to acquire within-modal view-independence through perceptual learning in an object recognition task, either visually or haptically, by exposure to both rotated and unrotated views. If there were separate unisensory, view-independent representations, acquiring view-independence in one modality would not transfer to the other. Conversely, if such transfer occurred, it would imply a single, bisensory, view-independent representation. In a second experiment, we employed cross-modal perceptual learning, where the stimulus modality was switched between study and test, again with exposure to both rotated and unrotated views. This was expected to train object recognition based on the bisensory view-independent representation. The question was whether this would induce within-modal view-independence. If it did, it would be reasonable to conclude that there is only one view-independent object representation, one that is also modality-independent (Fig. 1a). On the other hand, if cross-modal training did not affect within-modal view-independence, this would favor the existence of distinct, view-independent representations in each modality, in addition to the modality-independent one (Fig. 1b).

Experiment 1: Does within-modal learning of view-independence transfer cross-modally?

Methods

Participants

A total of 32 people (16 male, 16 female, mean age ± SD 23 ± 4 years) took part and were remunerated for their
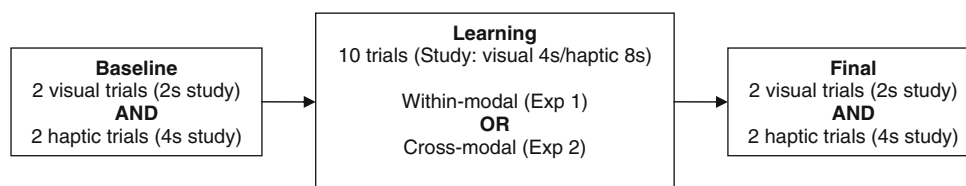
time. All gave informed written consent and all procedures were approved by the Emory University Institutional Review Board. Participants were randomly assigned to either a visual or a haptic learning group, 16 people (8 male, 8 female) in each.

Stimuli

We used 72 objects, each made from 6 smooth wooden blocks measuring $1.6 \times 3.6 \times 2.2$ cm (Fig. 2). The multipart objects were 9.5 cm high, and elongated along the *z*-axis, with the other dimensions free to vary according to the arrangement of the component blocks. Using smooth blocks avoided undesirable haptic cues to rotation about the *y*-axis, inherent in the textured surfaces of the Lego™ bricks used by Newell et al. (2001). The objects were painted medium gray to remove visual cues from variations in the natural color and grain. Each object had a small (<1 mm) gray pencil dot on one facet that cued the experimenter to present the object in the correct orientation. Debriefing showed that participants were never aware of these small dots. The 72 objects were divided into 18 sets of 4. As in our previous study (Lacey et al. 2007), we used difference matrices based on the number of differences in the position and orientation of component blocks to calculate the mean difference in object shape within each set of four. Paired *t* tests showed no significant differences between sets (all *p* values > 0.05) and the sets were therefore considered equally discriminable.

Task

All participants performed 18 trials of a four-alternative, forced-choice object recognition task (chance performance would therefore be 25%). The 18 trials were completed within a single session lasting 1–1.5 h. In each trial, participants studied four objects, identified by numbers 1–4. These objects were then presented twice each for recognition, once in the original orientation and once rotated 180° about the *y*-axis; participants were asked to identify each object by its number. Thus, each trial yielded four observations for unrotated, and four for rotated object recognition

| **Baseline** 2 visual trials (2s study) **AND** 2 haptic trials (4s study) | **Learning** 10 trials (Study: visual 4s/haptic 8s) Within-modal (Exp 1) **OR** Cross-modal (Exp 2) | **Final** 2 visual trials (2s study) **AND** 2 haptic trials (4s study) |

Fig. 3 A schematic timeline for the experiments: after recording within-modal baseline performance, participants performed 10 learning trials. In Experiment 1 these were either within-modal visual or haptic and in Experiment 2, either visual study-haptic test or vice versa. The effect of learning was assessed with final within-modal visual and haptic trials

following initial study of the four objects. Study and test in all trials in this experiment were within-modal, either visual or haptic, and different objects were used in each trial. Within each set of four objects, rotated or unrotated presentation of any one object during recognition was pseudorandom and the use of each set for baseline, learning or final phases was counterbalanced across subjects. Participants did not receive feedback at any stage.

A schematic timeline for the experiment is shown in Fig. 3: the detailed procedure was as follows. For the visual learning group, baseline performance was assessed over two visual trials followed by two haptic trials. In the visual baseline trials, the objects were studied for 2 s each, and in the haptic baseline trials, for 4 s each. This was followed by the learning phase in which participants performed ten visual trials: for these 10 trials, the study time for each object was doubled to 4 s. There was a short break halfway through these learning trials. Final performance was assessed over two further visual trials followed by two haptic trials: in these final trials, as in the baseline trials, participants had 2 s for visual study and 4 s for haptic study. Response times for the test trials were always unrestricted. The procedure for the haptic learning group was the same except that baseline and final performance were assessed over two haptic trials followed by two visual trials, and the ten learning trials were haptic trials in which the study time was 8 s. It is worth noting that the study times used here were much shorter than those used in our previous study (Lacey et al. 2007: visual 15 s and haptic 30 s). Pilot studies showed that these shorter durations were necessary to allow a sufficient range for perceptual learning effects. However, a 2:1 haptic:visual exploration time ratio was maintained, consistent with prior studies (Newell et al. 2001; Lacey and Campbell 2006; Lacey et al. 2007; and see Freides 1974), i.e., baseline and final trials: haptic 4 s, visual 2 s; learning trials: haptic 8 s, visual 4 s.

Participants sat facing the experimenter at a table on which objects were placed for both visual and haptic exploration. The table was 86 cm high so that, for visual exploration, the viewing distance was 30–40 cm and the viewing angle as the participants looked down at the objects was approximately 35°–45°. For visual presentations, the objects were placed on the table oriented along their z-axis as in Fig. 2. Participants were free to move their head and eyes when looking at the objects but were not allowed to get up and move around them. For haptic exploration, participants felt the objects behind an opaque cloth screen. Each object was placed into the participant's hands, oriented along its elongated z-axis in order to allow comprehensive haptic exploration of multiple surfaces (Lacey et al. 2007). Participants were free to move their hands over the object but were not allowed to rotate, manipulate, or otherwise move it out of its original orientation.

Results

There was no significant effect of gender on recognition accuracy (% correct responses) within either group, as shown by a preliminary ANOVA [visual: $F(1,14) = 0.07$, $p = 0.8$; haptic: $F(1,14) = 4.3$, $p = 0.06$]. Accuracy data for baseline and final performance were separately analyzed for each learning group with a three-way, repeated-measures analysis of variance (RM-ANOVA) with factors of modality (visual, haptic), time (baseline, final), and rotation (unrotated, rotated).

In the haptic learning group (Fig. 4a), visual recognition was more accurate than haptic recognition [$F(1,15) = 21.2$, $p < 0.001$], final recognition was more accurate than baseline recognition [$F(1,15) = 47.6$, $p < 0.001$], and recognition of rotated objects was less accurate than recognition of unrotated objects [$F(1,15) = 115.2$, $p < 0.001$]. In addition to these main effects, there was a significant Time × Rotation interaction [$F(1,15) = 85.5$, $p < 0.001$]: separate two-way (Modality, Rotation) RM-ANOVAs conducted on baseline and final performance showed that the Time × Rotation interaction arose because object rotation reduced baseline recognition [$F(1,15) = 166.9$, $p < 0.001$] but not final recognition [$F(1,15) = 0.4$, $p = 0.53$] (the main modality effect was significant at each time point). Thus, initial view-dependence was transformed after the perceptual learning trials into view-independence. Although training was confined to the haptic modality in these participants, the resulting view-independence was not: it was present in vision as well as haptics.
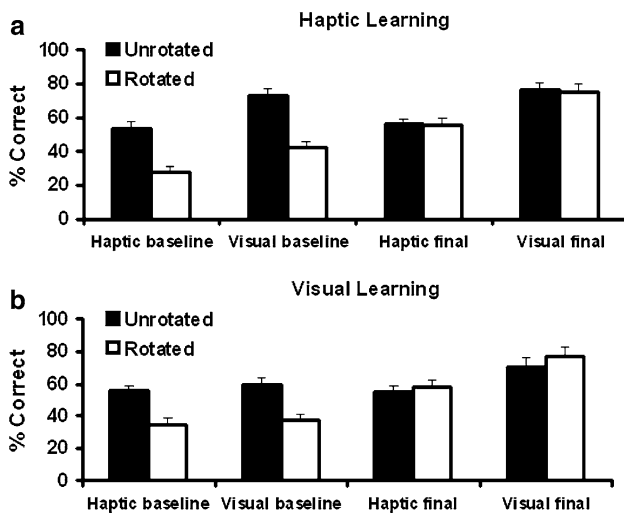
Fig. 4 Mean visual and haptic within-modal recognition accuracy for unrotated and rotated objects before and after **a** within-modal haptic learning and **b** within-modal visual learning. *Error bars* SEM

In the visual learning group (Fig. 4b), visual recognition was more accurate than haptic recognition [$F(1,15) = 8.3$, $p = 0.01$], final recognition was more accurate than baseline recognition [$F(1,15) = 12.0$, $p = 0.003$], and recognition of rotated objects was less accurate than recognition of unrotated objects [$F(1,15) = 18.4$, $p = 0.001$]. In addition to these main effects, there were significant Time × Rotation [$F(1,15) = 27.6$, $p < 0.001$] and Time × Modality [$F(1,15) = 5.5$, $p = 0.03$] interactions. Two-way (Modality, Rotation) RM-ANOVAs conducted separately on baseline and final performance showed that the Time × Rotation interaction arose from reduced recognition due to object rotation in the baseline sets [$F(1,15) = 44.3$, $p < 0.001$] but not in the final sets [$F(1,15) = 2.37$, $p = 0.15$]; while the Time × Modality interaction was explained by visual recognition being better than haptic recognition in the final sets [$F(1,15) = 16.7$, $p < 0.002$] but not in the baseline sets [$F(1,15) = 0.4$, $p = 0.54$]. For this group as well, then, perceptual learning transformed initial view-dependence into view-independence. Again, although training in these participants was purely visual, the acquired view-independence was found in both haptics and vision.

Discussion

Prior to the learning phase, visual and haptic within-modal object recognition were both view-dependent: in both learning groups, rotating the object away from the studied view significantly reduced recognition accuracy. This replicates earlier research with both unfamiliar (Newell et al. 2001; Lacey et al. 2007) and familiar objects (Lawson 2009). After visual learning, visual recognition became view-independent: rotated and unrotated objects were

recognized equally well. The same was true for haptic object recognition following haptic learning. Thus, the paradigm was effective at inducing within-modal view-independence. Most importantly, view-independence acquired in one modality transferred to the other: haptic within-modal view-independence was acquired after exclusively visual learning, and visual view-independence after exclusively haptic learning. This complete, symmetric cross-modal transfer suggests that vision and haptics share a single view-independent representation, as suggested by the three-representation model (Fig. 1a), arguing against the five-representation model in which cross-modal view-independence is processed via separate, unisensory, view-independent representations (Fig. 1b).

The paradigm of our Experiment 1 differs from cross-modal perceptual learning where each trial comprises visual study of a single object followed by haptic test with a single object, or vice versa (Norman et al. 2008): perceptual learning in this paradigm could simply reflect more efficient transfer of information between unisensory representations. Moreover, in the study of Norman et al. (2008), the view was not manipulated. We sought to provide converging evidence for the three-representation model by conducting a second experiment in which we modified the cross-modal perceptual learning paradigm of Norman et al. (2008). In the learning phase of this second experiment, participants studied four objects visually or haptically, as in Experiment 1, before being tested in the opposite modality using both unrotated and rotated views. We reasoned that this would reduce the repeated switching between modalities involved in the cross-modal paradigm of Norman et al. (2008), forcing participants to rely on a stored representation, and that this would tend to enhance participants' abilities to use the modality-independent representation. Because this representation is also view-independent (Lacey et al. 2007), the three-representation model predicts that this would result in enhanced within-modal view-independence, even though this was not specifically trained. On the other hand, the five-representation model allows the possibility that improved cross-modal recognition could stem from facilitating bisensory integration at the highest level, without necessarily improving within-modal view-dependence.

Experiment 2: Does cross-modal perceptual learning lead to within-modal view-independence?

Method

Participants

A total of 24 people (11 male, 13 female; mean age ± SD 22 ± 3 years) took part and were remunerated for their

time. All gave informed written consent and all procedures were approved by the Emory University Institutional Review Board. Participants were randomly assigned to either a visual-haptic (V-H) or a haptic-visual (H-V) learning group, 12 people in each (V-H: 6 male, 6 female; H-V: 5 male, 7 female).

Stimuli and task

The same objects were used as in Experiment 1. The paradigm was the same except that the ten within-modal learning trials were replaced by cross-modal learning trials. In the V-H group, the objects were studied visually and tested haptically; the reverse occurred in the H-V group. Baseline and final trials were identical to those in Experiment 1.

Results

A preliminary ANOVA showed no effect of gender on recognition accuracy (% correct responses) within either group [V-H: $F(1,10) = 0.3$, $p = 0.61$; H-V: $F(1,10) = 2.1$, $p = 0.18$]. Accuracy data for baseline and final performance for each group were separately analyzed with a three-way [Modality (visual, haptic), Time (baseline, final), Rotation (unrotated, rotated)] RM-ANOVA.

In the V-H learning group (Fig. 5a), visual recognition was more accurate than haptic recognition [$F(1,11) = 10.3$, $p < 0.01$], final recognition was more accurate than baseline recognition [$F(1,11) = 36.8$, $p < 0.001$], and recognition of rotated objects was less accurate than recognition of unrotated objects [$F(1,11) = 30.3$, $p < 0.001$]. In addition to these main effects, there was a significant Time × Rotation interaction [$F(1,11) = 51.2$, $p < 0.001$]. Separate two-way (Modality, Rotation) RM-ANOVAs showed that object rotation reduced baseline recognition [$F(1,11) = 75.2$, $p < 0.001$] but not final recognition [$F(1,11) = 0.3$, $p = 0.6$]. There was no main effect of modality in these separate analyses. Thus, cross-modal training induced within-modal view-independence in both vision and haptics.

In the H-V learning group (Fig. 5b), visual recognition was more accurate than haptic recognition [$F(1,11) = 11.0$, $p = 0.007$], final recognition was more accurate than baseline recognition [$F(1,11) = 29.9$, $p < 0.001$], and recognition of rotated objects was less accurate than recognition of unrotated objects [$F(1,11) = 86.7$, $p < 0.001$]. In addition to these main effects, there was a significant Time × Rotation [$F(1,11) = 43.7$, $p < 0.001$] interaction. Separate two-way (Modality, Rotation) RM-ANOVAs showed that object rotation reduced recognition in the baseline sets [$F(1,11) = 89.7$, $p < 0.001$] but not in the final sets [$F(1,11) = 0.2$, $p = 0.64$]; the main modality effect was only significant for the final sets. Thus, once again, view-independence was achieved in each modality by cross-modal training.
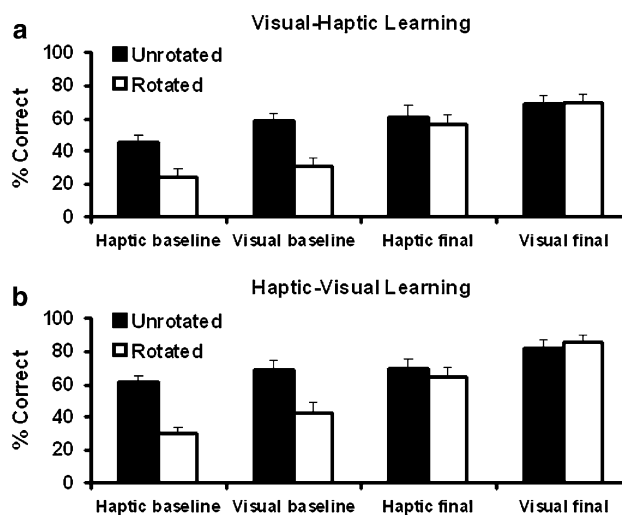
Fig. 5 Mean visual and haptic within-modal recognition accuracy for unrotated and rotated objects before and after **a** cross-modal visual-haptic learning and **b** cross-modal haptic-visual learning. *Error bars* SEM

Discussion

The results of Experiment 2 were clear: visual and haptic within-modal object recognition were both view-dependent before the cross-modal learning task but view-independent afterward. This was true for both the V-H and H-V learning groups. Thus, cross-modal training induces within-modal view-independence. As in Experiment 1, these results favor the three-representation model (Fig. 1a), rather than the five-representation model (Fig. 1b).

General discussion

In these experiments, visual and haptic object recognition was both initially view-dependent. The finding of haptic view-dependence replicated earlier research with both unfamiliar (Newell et al. 2001; Lacey et al. 2007) and familiar objects (Lawson 2009)—this point is important because it is somewhat counter-intuitive: when the hands can contact an object from different sides and move all over an object, one might expect haptic object recognition to be view-independent. Despite the fact that the objects were unfamiliar and lacked distinctive parts, within-modal recognition in each modality became view-independent following a short period of within-modal learning. This is consistent with learning studies of visual view-independence in a number of experimental paradigms (e.g., Jolicoeur 1985; Kraebel and Gerhardstein 2006; Perry et al. 2006; Liu 2007; and see Wallis and Bülthoff 1999, for a review). It is also consistent with a previous haptic study that showed haptic view-independence when participants were allowed

to explore multiple views (Ueda and Saiki 2007). The present study, however, is the first to show that haptic view-independence can be acquired in a learning paradigm. View-independence was achieved in a relatively short time (the experiment lasted approximately an hour) for these novel and closely similar objects. In pilot studies, we observed that it was necessary to double the study times in the learning trials to consistently achieve view-independence over the session. Clearly, the amount of familiarization required to achieve view-independence will vary depending on such factors as object complexity. However, we note that, at the end of the experiment, these objects could still not be said to approach the familiarity of everyday objects.

A striking finding of the present study was the complete, symmetric cross-modal transfer of within-modal view-independence (Experiment 1): visual view-independence acquired following exclusively visual learning also resulted in haptic view-independence, and vice versa. This suggests that both visual and haptic view-independence rely on a single, shared representation, supporting the notion that the cross-modal view-independence established previously (Lacey et al. 2007) is not dependent on separate, unisensory, view-independent representations (Fig. 1a). An equally striking finding was that cross-modal learning, both visual-haptic and haptic-visual, changed the status of both visual and haptic within-modal recognition from initial view-dependence to final view-independence (Experiment 2). Again, this fits with the idea that within-modal and cross-modal view-independence rely on the same shared representation. Thus, the two experiments of the present study, and its predecessor (Lacey et al. 2007) converge in support of the three-representation model of view-independence, in which separate view-dependent, unisensory representations feed directly into a view-independent, bisensory representation. One way in which this could be achieved is by integrating multiple low-level, view-dependent representations into a higher-order representation that is view-independent. Such a model has been proposed for vision (Riesenhuber and Poggio 1999).

The two main theories of view-independent visual object recognition suggest that this is accomplished either by structural descriptions of distinctive parts that can be transformed by rotation to match novel views (Biederman 1987) or by acquiring multiple views of an object and matching to these or interpolating between them (Bülthoff and Edelman 1992; Edelman and Bülthoff 1992). Since the objects used here lacked distinctive parts and were highly similar as a set, the present results fit better with the 'multiple views' hypothesis, particularly in respect of the view-independent, multisensory representation that supports visuo-haptic cross-modal recognition. Theories of visual view-independence can be used to drive theories of

haptic view-independence; for example, several studies have examined whether visual view-independence derives from temporal or spatial coherence in acquiring these multiple views. Liu (2007) concluded that temporal coherence was sufficient because object recognition after viewing a spatially disordered but temporally coherent sequence of views of an object was just as accurate as after viewing an orderly sequence of views of an object as though it were being rotated in the real world (both spatial and temporal coherence). However, Perry et al. (2006) showed that viewing ordered, interleaved sequences of views of different objects resulted in view-independence. Because the temporal link between one view of an object and the next spatially continuous view was broken by this interleaving, Perry et al. concluded that spatial coherence was sufficient and that temporal coherence merely facilitated view-independence. Similar questions could be framed for haptic view-independence: whether an orderly series of views around one axis or a random series of views around different axes would lead to view-independence, and whether this ordered exploration is a natural haptic exploratory procedure (Lederman and Klatzky 1987).

The cerebral cortical localization of the modality-independent, view-independent object representation is not known. Responses in the intraparietal sulcus appear to be view-dependent (James et al. 2002), although this is a well-known convergence region for visual and haptic shape processing (Amedi et al. 2001; Zhang et al. 2004; Stilla and Sathian 2008). The lateral occipital complex is also a convergence site for multisensory shape processing (Amedi et al. 2001; Zhang et al. 2004; Stilla and Sathian 2008) but it is as yet unclear whether representations in this area are view-dependent (Grill-Spector et al. 1999; Gauthier et al. 2002) or view-independent (James et al. 2002). Further work is required to define the locus and nature of the modality-independent, view-independent object representation.

Few studies have examined viewpoint effects in the blind in order to assess the influence of visual experience. However, Heller and colleagues (Heller et al. 2002, 2006) conducted a series of experiments in which visually impaired and blindfolded, sighted participants matched physical objects to raised-line tangible pictures. While visual experience did not appear to be necessary for understanding linear perspective, some views provided more information and facilitated recognition more than others (Heller et al. 2002; see also Woods et al. 2008). However, when the orientation of the physical object was mismatched with the drawn orientation, early blind participants performed significantly less accurately than late blind, very low vision or blindfolded, sighted participants, with none of the latter three groups differing significantly (Heller et al. 2006) suggesting that there may be a critical period for view-independence

(see also Sathian 2005). Further studies are required that explicitly test the effect of object rotation in people with varying degrees of visual experience.

Finally, we found no consistent effect of gender. Although there was a nearly significant effect for the haptic learning group in Experiment 1, it was nowhere near significance in the visual learning group in this experiment or for either group in Experiment 2. In addition, examination of performance by gender did not reveal an interpretable trend even in Experiment 1. Although gender differences are well documented in visuo-spatial tasks, less is known about the haptic domain and still less about the interface between the two modalities. Males appear to be more accurate in haptic parallelity tasks (Kaas and van Mier 2006; Zuidhoek et al. 2007) and in haptic estimation of width (Gupta and Gupta 1997) while no significant gender differences were found for haptic perception of the horizontal (Robert et al. 1994; Heller et al. 1999) or the vertical (Heller et al. 1999) or for tactile-visual matching of 3D shape (Chellew and Persinger 1994). A related issue is the use of reference frames. It has been suggested that, while males are better than females in using allocentric frames of reference in vision (where egocentric frames are also available), haptics almost invariably relies on egocentric frames and that this removes the gender difference in some haptic spatial tasks (Zuidhoek et al. 2007). Further research should address the issue of reference frames in cross-modal contexts.

We conclude that vision and haptics share a single, multisensory, view-independent representation that enables both within- and cross-modal view-independent object recognition. An important task for future research will be the elaboration of the multisensory aspects of view-independence. This is obviously significant for understanding visuo-haptic interactions. Further, we propose that adopting a multisensory perspective could illuminate models of view-independence that hitherto have been exclusively concerned with vision.

References

Amedi A, Malach R, Hendler T, Peled S, Zohary E (2001) Visuo-haptic object-related activation in the ventral pathway. Nat Neurosci 4:324–330

Biederman I (1987) Recognition by components: A theory of human image understanding. Psychol Rev 94:115–147

Booth MCA, Rolls ET (1998) View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb Cortex 8:510–523

Bülthoff HH, Edelman S (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proc Natl Acad Sci USA 89:60–64

Chellew G, Persinger MA (1994) Women but not men exhibit a positive correlation between complex partial epileptic-like signs and tactile-visual cross-modal matching: Implications for hemispheric intercalation. Percept Mot Skills 78:1312–1314

Edelman S, Bülthoff HH (1992) Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vis Res 32:2385–2400

Freides D (1974) Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. Psychol Bull 8:284–310

Gauthier I, Hayward WG, Tarr MJ, Anderson AW, Skudlarski P et al (2002) BOLD activity during mental rotation and view-dependent object recognition. Neuron 34:161–171

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24:187–203

Gupta U, Gupta BS (1997) Sex differences in haptic perceptual judgment. Psychol Stud 42:72–74

Heller MA, Calcaterra JA, Green SL, Barnette SL (1999) Perception of the horizontal and vertical in tangible displays. Perception 28:387–394

Heller MA, Brackett DD, Scroggs E, Steffen H, Heatherly K, Salik S (2002) Tangible pictures: viewpoint effects and linear perspective in visually impaired people. Perception 31:747–769

Heller MA, Kennedy JM, Clark A, McCarthy M, Borgert A, Wemple L, Fulkerson E, Kaffel N, Duncan A, Riddle T (2006) Viewpoint and orientation influence picture recognition in the blind. Perception 35:1397–1420

James TW, Humphrey GK, Gati JS, Menon RS, Goodale MA (2002) Differential effects of view on object-driven activation in dorsal and ventral streams. Neuron 35:793–801

Jolicoeur P (1985) The time to name disoriented natural objects. Mem Cognit 13:289–303

Kaas AL, van Mier HI (2006) Haptic spatial matching in near peripersonal space. Exp Brain Res 170:403–413

Kraebel KS, Gerhardstein PC (2006) Three-month-old infants' object recognition across changes in viewpoint using an operant learning procedure. Infant Behav Dev 29:11–23

Lacey S, Campbell C (2006) Mental representation in visual/haptic crossmodal memory: Evidence from interference effects. Q J Exp Psychol 59:361–376

Lacey S, Peters A, Sathian K (2007) Cross-modal object recognition is view-independent. PLoS ONE 2(9):e890. doi:10.1371/journal.pone.0000890

Lawson R (2009) A comparison of the effects of depth rotation on visual and haptic 3D object recognition. J Exp Psychol Human (in press)

Lederman SJ, Klatzky RL (1987) Hand movements: a window into haptic object recognition. Cogn Psychol 19:342–368

Liu T (2007) Learning sequence of views of three-dimensional objects: the effect of temporal coherence on object memory. Perception 36:1320–1333

Newell FN, Ernst MO, Tjan BS, Bülthoff HH (2001) View dependence in visual and haptic object recognition. Psychol Sci 12:37–42

Newell FN, Woods AT, Mernagh M, Bülthoff HH (2005) Visual, haptic and crossmodal recognition of scenes. Exp Brain Res 161:233–242

Norman JF, Clayton AM, Norman HF, Crabtree CE (2008) Learning to perceive differences in solid shape through vision and touch. Perception 37:185–196

Peissig JJ, Tarr MJ (2007) Visual object recognition: do we know more now than we did 20 years ago? Annu Rev Psychol 58:75–96

Perry G, Rolls ET, Stringer SM (2006) Spatial vs temporal continuity in view invariant visual object recognition learning. Vis Res 46:3994–4006

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025

Robert M, Pelletier J, St-Onge R, Berthiaume F (1994) Women's deficiency in water-level representation: present in visual conditions yet absent in haptic contexts. Acta Psychol 87:19–32

Sathian K (2005) Visual cortical activity during tactile perception in the sighted and the visually deprived. Dev Psychobiol 46:279–286

Stilla R, Sathian K (2008) Selective visuo-haptic processing of shape and texture. Hum Brain Mapp 29:1123–1138

Tarr MJ, Bülthoff HH (1995) Is human object recognition better described by geon structural descriptions or by multiple views: comment on Biederman and Gerhardstein (1993). J Exp Psychol Human 21:1494–1505

Ueda Y, Saiki J (2007) View independence in visual and haptic object recognition. Jpn J Psychon Sci 26:11–19

Wallis G, Bülthoff HH (1999) Learning to recognize objects. Trends Cogn Sci 3:22–31

Woods AT, Moore A, Newell FN (2008) Canonical views in haptic object representation. Perception 37:1867–1878

Zhang M, Weisser VD, Stilla R, Prather SC, Sathian K (2004) Multi-sensory cortical processing of shape and its relation to mental imagery. Cogn Affect Behav Neurosci 4:251–259

Zuidhoek S, Kappers AML, Postma A (2007) Haptic orientation perception: Sex differences and lateralization of functions. Neuropsychologia 45:332–341