The effect of temporal delay and spatial differences on cross-modal object recognition

ANDREW T. WOODS
University of Dublin, Trinity College, Dublin, Ireland

SILE O'MODHRAIN
Media Lab Europe, Dublin, Ireland

and

FIONA N. NEWELL
University of Dublin, Trinity College, Dublin, Ireland

In a series of experiments, we investigated the matching of objects across visual and haptic modalities across different time delays and spatial dimensions. In all of the experiments, we used simple L-shaped figures as stimuli that varied in either the x or the y dimension or in both dimensions. In Experiment 1, we found that cross-modal matching performance decreased as a function of the time delay between the presentation of the objects. We found no difference in performance between the visual–haptic (VH) and haptic–visual (HV) conditions. Cross-modal performance was better when objects differed in both the x and y dimensions rather than in one dimension alone. In Experiment 2, we investigated the relative contribution of each modality to performance across different interstimulus delays. We found no differential effect of delay between the HH and VV conditions, although general performance was better for the VV condition than for the HH condition. Again, responses to xy changes were better than changes in the x or y dimensions alone. Finally, in Experiment 3, we examined performance in a matching task with simultaneous and successive presentation conditions. We failed to find any difference between simultaneous and successive presentation conditions. Our findings suggest that the short-term retention of object representations is similar in both the visual and haptic modalities. Moreover, these results suggest that recognition is best within a temporal window that includes simultaneous or rapidly successive presentation of stimuli across the modalities and is also best when objects are more discriminable from each other.

The fundamental differences between the encoding mechanisms for the visual and haptic senses would suggest that these modalities might also use different underlying mechanisms for the storage of information. For example, in a cross-modal recognition task, encoding time is generally longer for haptics than for vision (Newell, in press). Once an object has been encoded, active touch is both a proficient and efficient method for recognizing common objects (Klatzky, Loomis, Lederman, Wake, & Fujita, 1993) and for exploring and recognizing novel objects (Newell, Ernst, Tjan, & Bülthoff, 2001). However, the fact that vision and touch encode and store object properties in different ways does not necessarily imply the existence of different storage systems.

The present studies concern how objects are stored in visual and haptic memory and whether these memory systems are functionally similar and thus allow for efficient cross-modal recognition. Specifically, we investigate the effect of temporal delay between learning the stimulus in one modality and retrieving its representation in the other. We also sought to determine whether spatial differences along one or two dimensions between objects would affect cross-modal recognition.

A large volume of literature has explored the visual and haptic memory systems, and two broad accounts have emerged. The first of these accounts is that object memory is multisensory, such that object representations are easily shared across modalities, provided the same properties are encoded. Here we would expect no cost for cross-modal performance as opposed to within-modal performance. Furthermore, any observable effects on recognition performance with manipulations such as time delays should be common to all modalities. The second account would be that objects are stored as modality-specific representations that require a recoding from one memory system to the other. Such recoding would result in a cost in terms of time and error rates for cross-modal performance relative to within-modal performance. Moreover, there would be no reason to expect similar effects on recog-

nition performance across modalities with manipulations such as time delay. In other words, the memory systems may be functionally independent in such a way that performance in one might not be related to performance in the other.

Many recent studies support the idea that visual and haptic memory for objects is shared. For example, neuro-imaging evidence has emerged to show that the cortical area involved in visual object recognition is also involved in haptic object recognition. Previous studies have reported that the lateral occipital complex (area LOC) is active during visual object recognition tasks (see Grill-Spector, Kourtzi, & Kanwisher, 2001, for a review). More pertinently however, Amedi, Malach, Hendler, Peled, and Zohary (2001) reported that activation in the LOC to objects is not specific to the visual modality since the LOC is also active during haptic object recognition. Similarly, James et al. (2002) reported activation in area LOC to the visual presentation of novel objects that were previously learned through touch. Furthermore, this area was not active to auditory information about objects (Amedi, Jacobson, Hendler, Malach, & Zohary, 2002), suggesting that area LOC (or LOtv, as termed by Amedi et al., 2002) is involved in recovering the geometrical shape of objects.

Some recent behavioral studies have also suggested that object representations are shared across modalities. For example, Easton, Greene, and Srinivas (1997) reported cross-modal priming effects for 2-D patterns and 3-D objects, suggesting that vision and haptics share representations of the objects based on their shapes and structures. Reales and Ballesteros (1999) also found that when an object was presented in one modality (vision or haptics), the recognition of the same shape in the other modality was facilitated. Cross-modal object representations can also complement each other in order to achieve object constancy with changes in viewpoint (Newell et al., 2001). Moreover, bimodal encoding of objects can result in recognition performance that is superior to unimodal encoding of objects (Newell, Bülthoff, & Ernst, 2004), suggesting that a rich representation of objects is created in memory across the modalities. Taken together, these studies support the idea that object representations might reside in a common storage location easily accessible to both vision and touch.

However, other studies suggest that memory for touch is not equivalent to visual memory. Research on tactile memory provides tentative support for a tactile memory independent of other sensory memory systems and provides a framework for the characterization of haptic memory. For example, Gilson and Baddeley (1969) found evidence for a tactile memory that comprises two elements, one affected by articulatory suppression, and the other unaffected. They applied a single tactile stimulus to a specific location on the underside of a participant's forearm, and the participant's task was to point to the location of that stimulus. The authors introduced a delay between stimulus and response ranging from 0 to 60 sec. In one condition, participants conducted an articulatory suppression task during the delay, and in the other condition, they were silent.

Gilson and Baddeley found that accuracy levels were maintained for up to 15 sec in the silent condition. After this 15-sec delay, accuracy began to diminish. Articulatory suppression resulted in worse performance than in the silent condition, but these differences emerge only after a 15-sec delay. Using the same paradigm, however, Sullivan and Turvey (1972) found that accuracy was reduced after a delay of only 5 sec. There were two task conditions, one with an unfilled delay between stimulation and recall, and another in which the delay was filled with a subsidiary arithmetic task. The authors found that task accuracy was lower for the condition with the arithmetic task. Mahrer and Miles (2002) found that articulatory suppression impairs accuracy in a task that compared two temporally separated sequences of tactile stimuli applied to different fingers. This result suggests, as in the previous studies, a verbal component to the memory of tactile sequences. Although together these studies involve an investigation of tactile memory, it might be that tactile stimulation of body parts involves processes that are independent of those involved in haptic object recognition and may not reveal the precise nature of object representations per se more than the nature of body representations. Furthermore, studies on passive touch may reveal very little about the processes involved in active touch, such as those involved in object recognition tasks (Gibson, 1962; Heller, 1984).

Some further studies have investigated active haptic memory for objects. For example, Kiphart, Hughes, Simmons, and Cross (1992) found that the haptic memory of complex objects decayed only after a delay of 15 sec (i.e., a decay function similar to that described by Gilson and Baddeley, 1969). In a study on 3-D object recognition in children, Millar (1974, Experiment 2) found that haptic matching performance for nonsense objects was better with interstimulus delays of 1 sec than with delays of 5 and 30 sec, suggesting that haptic memory starts to decay immediately after the exploration of an object. On the other hand, Garvill and Molander (1973) reported no effects of interstimulus delays from 1 to 30 sec in haptic, visual, or cross-modal matching of 3-D nonsense objects. Differences between these studies may be accounted for by differences in task demands or stimulus complexity which may influence the encoding, storage, and retrieval of object information. For example, Kiphart et al. reported using complex objects made from 15 Lego bricks, and differences between objects were characterized by changes in the position of 3 of these bricks. In comparison, the shapes used in the Garvill and Molander study were simple. Thus, the differences between these studies may reflect differences in the stimulus complexity, although it is difficult to quantify the differences in stimuli across these studies. Additionally, the degree of familiarity with the objects being explored and whether or not they can be verbalized may affect performance (Millar, 1981).

In our experiments, we investigated the effect of temporal delay on within- and cross-modal matching performance in an object recognition task. Unlike most previous studies on haptic and visual memory performance, which

used object stimuli sets whose objects were categorically different from each other (e.g., a square vs. a circle), in our experiments we used a set of L-shaped object stimuli that systematically varied along one or two stimulus dimensions. Such objects have been previously used by Millar and Al-Attar (2000) to investigate vertical and bisection illusions in vision and touch. They found that these L-shapes elicit the same illusion in touch as in vision, suggesting that differences between the shapes provide the same spatial effects. Furthermore, verbal recoding has been found to improve tactile memory, especially for similar tactile items (Millar, 1975); therefore, we wanted to ensure that the differences between our stimuli were difficult to verbalize. Since our stimuli constituted metric rather than categorical differences, such differences are difficult to verbalize. We were interested in investigating whether objects that varied from each other in more than one dimension would be more easily discriminated across modalities and whether objects that are more easily discriminable are also remembered more accurately over time.

EXPERIMENT 1

In this experiment, we used a cross-modal matching task to measure object recognition performance. We were particularly interested in investigating memory performance in object recognition; therefore, traditional tests of short-term memory (e.g., memory span tests or Brown–Peterson designs) were not appropriate here. We were interested in whether a time delay between matching objects had the same effect on objects that were encoded visually and later presented haptically versus objects that were encoded haptically and then presented visually. If visual memory is independent from haptic memory, then we would expect different effects of delay on cross-modal performance. If, on the other hand, object memory is multisensory, then we would expect similar effects of delay, irrespective of the encoding modality. Furthermore, we were interested in whether effects of delay are influenced by stimulus complexity. We might assume, as suggested by previous literature, that the representations of more discriminable objects would be retained across large delays between presentation of the stimulus to be encoded and presentation of the comparison stimulus (e.g., Garvill and Molander, 1973) whereas the representation of less discriminable objects might be more susceptible to decay.

Method

Participants. Sixteen undergraduate students (4 male and 12 female) from Trinity College Dublin participated in the experiment for course credit. Their ages ranged from 17 to 26 years, with a mean age of 22.2 years. All of the participants had normal or corrected-to-normal vision, and none reported any somatosensory impairment.

Stimuli. Twenty-five L-shaped objects were created as stimuli. The objects differed from each other according to the lengths of their x- and y-axes. The smallest object in the set measured 2 cm in the y-axis and 2 cm in the x-axis. The largest object measured 6 × 6 cm. The length along each axis increased in increments of 1 cm. The x and y lengths paired for each object ranged from 2 to 6 cm, thus creating a complete set of 25 objects. Figure 1 illustrates the set of stimuli used in this experiment.
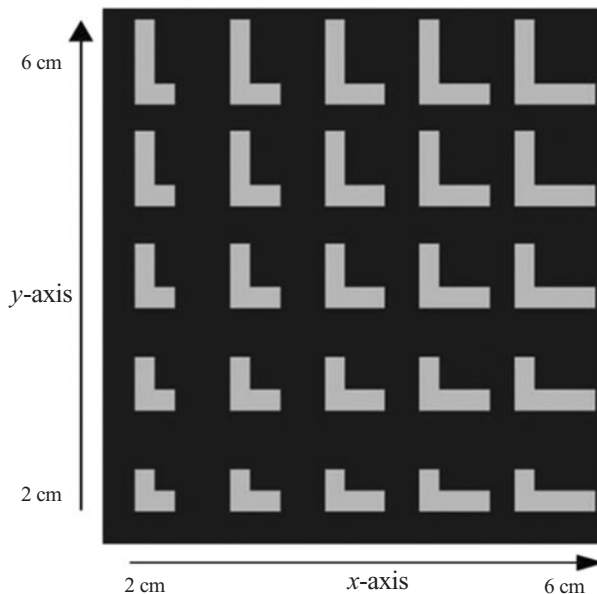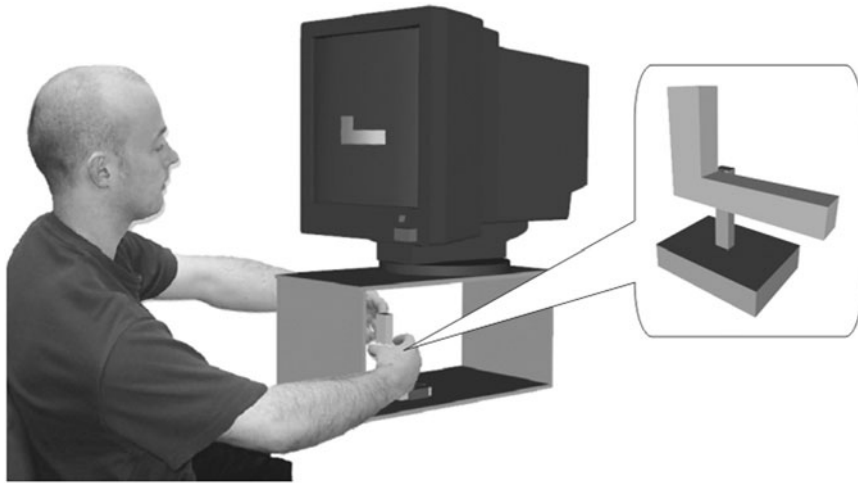

Figure 1. An illustration of the stimulus set. All the stimuli were L-shaped. The lengths of their y- and x-axis extensions were measured from the inside edge of each arm (the minimum extension was 2 cm, the maximum 6 cm). Stimuli that are horizontally or vertically adjacent on the matrix differ from each other by 1 cm.

We created two sets of stimuli based on these size criteria: a haptic and a visual set. The haptic set was made from four fused sheets of 0.5-cm-thick laminate, measuring a total of 2 cm in depth. The visual set was presented on a computer screen and was made using the CorelDraw graphics package. The visual set was created so that it matched the physical dimensions of the haptic set exactly.

Apparatus. The apparatus consisted of a computer monitor placed on a platform which was used to occlude the haptic stimuli from view. The center of the monitor was then positioned approximately at eye level for each participant, at a distance of 57 cm. Figure 2 illustrates our experimental set-up. Visual stimuli were presented centrally on the 21-inch CRT monitor, and haptic stimuli were presented underneath the monitor behind the front of the platform and thus hidden from view. Each haptic stimulus was affixed onto the end of a supporting structure so that each object was suspended 10 cm above the table to allow the participants to freely explore each object. The stimuli were held in place by two prongs, which projected from the supporting structure and fitted into two holes in each of the stimuli. These holes were located at the junction of each stimulus's arms, allowing each length of the L-shaped stimulus to be perceived. The supporting structure's base fitted into a specially made socket on the table to keep the stimuli in a stable vertical orientation. Once positioned, the haptic stimulus was directly aligned with the front of the computer monitor.

The experiment was run using DMDX software for the PC (Forster & Forster, 2003). A head-mounted microphone was used to record the participants' responses during the experiment.

Design. The experiment was based on a mixed-subjects design with two within- and one between-subjects factors. The between-subjects variable was the order of modality between learning and test. There were two levels to this factor: visual learning followed by haptic testing (VH) and haptic learning followed by visual testing (HV). Because the main purpose of our research was to compare multisensory object memory performance across different delays and not to compare memory performance across modalities per se, it was not necessary to have VV and HH controls. The within-subjects factors were the interstimulus interval (ISI) and spatial differences (spa-

Figure 2. An illustration of our experimental set-up, including the monitor and the supporting structure. An illustration of the stimulus is shown in the inset.

tial). For ISI, the levels were 0, 15, and 30 sec and for *spatial*, they were differences in the *x*-axis, *y*-axis or *xy*-axes between the stimuli. On *different* trials, there were equal numbers of 1-cm, 2-cm, and 3-cm size differences along the *x*- or *y*-axes. For differences along both *x*- and *y*-axes, the size difference was an average of 4 cm, but differences in scale between two objects were avoided (e.g., if one object was 2 × 2 cm, then the other could not be 3 × 3 cm). The sum of *x*-axis and *y*-axis changes equaled the sum of *xy*-axis changes. This allowed us to make a direct comparison between recognition performance to differences in the *x*- or *y*-axis to differences in changes of the *xy*-axes.

The dependent variables were participant error rate (*error*) and voice-activated response time (RT). All trials were randomly presented across the participants.

Procedure

Before the experiment, the participants were randomly assigned to the two experimental groups of the modality order condition—namely, VH and HV learning and testing conditions.

The experiment was based on a same/different matching protocol, and the participants were instructed to decide whether two successively presented stimuli were the same or different as fast and as accurately as possible.

Prior to the experiment, the participants in the HV condition were instructed to match objects by first feeling and then looking at the objects in each trial. The participants in the VH condition were instructed to first look at and then feel the objects. Prior to haptic presentation, all participants received a 500-msec warning tone (a series of beeps), which alerted them to prepare to feel the stimulus. They then heard a long continuous tone during which they were instructed to feel the haptic stimulus. The tone lasted for 3 sec and on offset, the participants had to remove their hands from the object. Thus, haptic exploration of an object lasted for 3 sec, which was considered adequate time to encode the properties of the object through touch (Kiphart et al., 1992). A visual presentation of a stimulus was preceded by a fixation cross for 500 msec, which alerted the participant and prepared him or her for the subsequent presentation of a visual stimulus. A visual object remained on the screen for 250 msec.[1] The delay between the presentation of the first and second stimuli in a trial was 0, 15, or 30 sec. There was an intertrial interval (ITI) of 15 sec, during which the experimenter changed the haptic stimulus for the next trial.

For both the VH and the HV conditions, the second stimulus had 50% likelihood of being identical to the first. On the trials in which the two stimuli were different, the second stimulus could differ along the *x*-axis, the *y*-axis, or both (see Design section). There was a total of 108 trials in the experiment (3 spatial conditions, 3 ISIs, repeated six times, totaling 54 *different* trials and a further 54 *same* trials). To familiarize the participant with the procedure, 10 practice trials were given at the start of the experiment. The voice-activated response recorder was calibrated for each participant. The experiment lasted for approximately 80 min.

Results

Our main measure of performance was the mean hit and correct rejection scores for each participant across the modality order and ISI conditions (see Table 1). A mixed-factorial analysis of variance (ANOVA) was conducted, using modality order as a between-subjects factor and ISI as a within-subjects factor.[2] We found no effect of modality order [$F(1,14) = 1.665$, n.s.]. A main effect of ISI was found [$F(2,28) = 7.183, p < .005$]. A post hoc Newman–Keuls analysis revealed that delays of 0 sec produced significantly better matching performance than did delays of

Table 1
Mean Percentage Correct Responses (Hit and Correct Rejection Scores) and Mean Response Times (RTs, in Milliseconds) Across the Different ISIs In Each Order of Modality in Experiment 1

ISI	% Correct	*SE*	RT	*SE*
		Vision–Haptics		
0 msec	86.89	1.40	1,929.7	272.9
15 msec	81.94	2.16	2,044.8	245.6
30 msec	76.82	1.84	2,226.6	282.8
		Haptics–Vision		
0 msec	82.70	1.71	770.3	40.1
15 msec	79.51	4.13	878.1	80.6
30 msec	75.22	2.23	869.9	79.8

Note—ISI, interstimulus interval.

30 sec ($p < .005$) but not delays of 15 sec. There was a nonsignificant trend toward a difference in performance between delays of 15 and 30 sec ($p = .052$). We found no interaction between the factors [$F(2,28) < 1$].

The mean response times (across hits and correct rejections) for each modality order across the different ISIs are shown in Table 1. We conducted a two-way, mixed-subjects ANOVA with modality order as the between-subjects factor (VH and HV) and ISI (0, 15, and 30 sec) as the within-subjects factor. We found a main effect of order [$F(1,14) = 20.329, p < .01$]. Response times to the HV condition were significantly faster than response times to the VH condition. We found a main effect of ISI [$F(2,28) = 8.304, p < .01$]. There was no significant interaction between the factors [$F(2,28) = 1.805$, n.s.].

A three-way, mixed-subjects ANOVA was conducted on the correct rejection responses with modality order, ISI, and spatial differences as factors. We found no effect of modality order [$F(1,14) < 1$] or of ISI [$F(2,28) = 1.658$, n.s.]. The effect of spatial differences was significant [$F(2,28) = 43.200, p < .001$]. The mean percentage performance for each of the spatial differences was as follows: 69.9% (3.1% SE) accuracy for x differences; 66.4% (2.9% SE) accuracy for y differences, and 90.6% (2.0% SE) accuracy for xy differences. A Newman–Keuls test revealed that performance was significantly better to changes in the xy dimensions than to changes in either the x ($p < .001$) or y ($p < .001$) dimensions and no difference between the x and y dimensions. There were no significant interactions between the factors [$F(2,28) < 1$].

Discussion

We found that the effect of delay between cross-modal presentations of object stimuli was the same whether the initial object was encoded visually or haptically. This finding suggests that visual and haptic memory for objects is equivalent and that representations decay at the same rate. The fact that response times were slower in the VH condition than in the HV condition is simply a reflection of haptics being relatively slow in encoding shape properties when compared with vision. A more interesting finding is that there was no difference in discrimination performance between the cross-modal conditions.

We also found that changes in the xy dimensions resulted in better performance than changes in either the x or y dimension alone, with no difference between the unidimensional conditions. This shows, perhaps not surprisingly, that the more dimensions the two objects differ along, the easier it is to discriminate between the objects. What is interesting, however, is that there was no interaction between ISI and spatial differences. We expected that the effects of ISI would be attenuated when the objects were more discriminable, but there was no main effect of ISI on the correct rejections scores, which may have obscured any potential effects of an interaction between the factors.

Our findings suggest that the visual and haptic modalities share the same memory characteristics. The results

might also indicate that objects are encoded into a common representation shared by both modalities and that performance reflects the characteristics of this multisensory memory. We cannot, however, conclude that memory is multisensory without first comparing recognition performance between the modalities. We therefore conducted a further experiment to investigate differences between the modalities.

EXPERIMENT 2

In this experiment, we directly compared matching performance between the visual and haptic modalities. In the previous experiment, we found no effect of order of the encoding and comparison modalities in the matching task, suggesting that memory is the same across modalities. However, performance may also have been related to one memory system if, for example, sensory capture had occurred. Thus, in all cases, the representation of the object may have been recoded into visual memory in order for cross-modal recognition to have occurred. In this case, we would expect to find differences between visual and haptic matching tasks in terms of performance across time delays. On the other hand, visual and haptic memory may indeed be similar; therefore, we would expect the same effect of time delay on performance, irrespective of modality. Finally, as in Experiment 1, we tested the effect of time delay across modalities with objects of varying discriminability to see whether the representation of objects that are more easily discriminated would be more robust over longer time delays.

Method

Participants. Twenty-four undergraduate students (14 male and 10 female) from Trinity College Dublin participated in the experiment for course credit. Their ages ranged from 19 to 47 years, with a mean age of 24 years. All of the participants had normal or corrected-to-normal vision and none reported any somatosensory impairment.

Stimuli and Apparatus. We created an exact copy of our haptic set of stimuli in order to have identical, duplicate haptic objects for use in the *same* trials. In all other aspects, the stimuli were the same as those used in Experiment 1.

We adapted our apparatus slightly in order to present two haptic stimuli at the same time. We created two supporting structures and set one behind the other, onto which haptic objects could be placed. The distance between these supports was 10 cm, which allowed for palpation of each object without interference from the other. The center of the computer monitor screen was aligned with the midpoint between the two haptic stimuli.

Design. The experiment was based on a two-factor, within-subjects design with modality and ISI as factors. The spatial factor was nested under these factors. For the modality factor, the levels were intramodal visual (VV) and intramodal haptic (HH). The levels for the ISI factor were a delay of either 0, 15, or 30 sec. Again we used differences along the x-axis, y-axis, or xy-axes together as levels in the spatial factor.

Procedure. With the following exceptions, the procedure was the same as in Experiment 1: All trials were within modalities. Also, there was an ITI of 15-sec for the HH trials and 1.5 sec for the VV trials. During the haptic presentation of the stimuli, the participants were instructed to first feel the object closest to them. The second

haptic stimulus was presented 10 cm after the first, and participants were instructed to reach behind the first object to feel the second. Haptic presentation was always 3 sec, and visual presentation was 0.25 sec (as in Experiment 1). The second visual stimulus was presented in a location on the monitor that was offset from the location of the first stimulus by 1 cm in a random direction. We did this to ensure that participants were not using apparent motion to match between the visual stimuli.

We reduced the number of trials overall from Experiment 1 in order not to overburden our participants. Thus there were slightly different numbers of trials, depending on the differing number of presentations of each of the *x*, *y*, and *xy* dimensions across participants, but the same number of trials was retained per level across the experiment. Trials were blocked per modality, and the order of these blocks was counterbalanced across participants. There were 18 *different* trials and 18 *same* trials per modality, yielding a total of 72 trials in the experiment. The experiment took approximately 45 min to complete.

Results

The mean of each participant's hit and correct rejection scores was calculated, and we conducted a two-way, within-subjects ANOVA with modality and ISI as factors. See Table 2 for the mean scores across the modality and ISI factors. We found a main effect of modality [$F(1,23) = 9.959, p < .01$]. Performance to the VV modality was better than performance to the HH modality. A main effect of ISI was also found [$F(2,46) = 4.057, p < .05$]. Post hoc Newman–Keuls analyses revealed that performance to an ISI of 0 sec was better than performance to ISIs of 15 sec ($p < .05$) and 30 sec ($p < .05$). Performance did not differ between ISIs of 15 and 30 sec. There was no interaction between modality and ISI [$F(2,46) < 1$].

The mean response times (RTs, for hits and correct rejections) for each of the modalities across the different ISIs are shown in Table 2. We conducted a two-way within-subjects ANOVA on the RTs across the modality and ISI factors. We found a main effect of modality [$F(1,23) = 100.87, p < .001$] with faster RTs to the VV condition than to the HH condition. We also found a main effect of ISI [$F(2,46) = 5.72, p < .01$]. RTs increased with longer interstimulus delays such that RTs to the 0-sec ISI were faster than to the 30-sec ISI ($p < .05$). There was no difference in responses between either the 0- and 15-sec ISIs or the 15- and 30-sec ISIs. There was no interaction between the factors, although the interaction approached significance [$F(2,46) = 2.778, p = .073$].

Table 2
Mean Percentage Correct Responses (Hit and Correct Rejection Scores) and Mean Response Times (RTs, in Milliseconds) Across the Different ISIs in Each Modality in Experiment 2

ISI	% Correct	SE	RT	SE
Vision–Vision				
0 msec	93.19	1.76	951.5	28.8
15 msec	87.50	2.01	959.7	40.4
30 msec	87.36	2.09	1,038.0	68.7
Haptics–Haptics				
0 msec	83.47	2.26	2,095.2	151.6
15 msec	80.66	2.33	2,344.8	129.0
30 msec	79.27	3.25	2,481.8	174.1

Note—ISI, interstimulus interval.

We conducted a separate three-way ANOVA on the correct rejection responses with modality, ISI, and spatial differences as factors. There was an effect of modality [$F(1,22) = 7.306, p < .05$] with better performance to the VV condition than to the HH condition. We also found an effect of ISI [$F(2,44) = 3.777, p < .05$]. A post hoc Newman–Keuls analysis revealed that performance to the 0-sec ISI was significantly better than performance to an ISI of 30 sec ($p < .05$). There was no difference between the 0- and 15-sec ISIs and between the 15- and 30-sec ISIs. We found a main effect of spatial differences [$F(2,44) = 9.214, p < .0005$]. A post hoc Newman–Keuls analysis on the spatial effect revealed that performance was better to differences between the *xy*-axes than to differences between either the *x* ($p < .01$) or *y*-axes ($p < .01$) axes, and there was no difference between responses to the *x*- and *y*-axes.

Interestingly, we found an interaction between spatial differences and ISI [$F(4,88) = 12.200, p < .001$]. This interaction is depicted in Figure 3B. Post hoc tests revealed no effect of ISI with changes in the *xy* dimensions. For changes in the *x* dimension, performance to the ISI of 30 sec was significantly worse than performance to both 0-sec ($p < .001$) and 15-sec ISIs ($p < .001$) and no difference between 0- and 15-sec ISIs. Finally, for changes in the *y* dimension only, performance was worse to an ISI of 15 sec relative to ISIs of 0 sec ($p < .05$) and 30 sec ($p < .05$), with no difference between 0- and 30-sec ISIs.

We found a significant interaction between modality and ISI [$F(2,44) = 3.856, p < .05$]. This interaction is depicted in Figure 3B. Post hoc tests revealed no difference between the ISI conditions in the visual modality, but in the haptic modality, performance was significantly worse for ISIs of 15 sec and 30 sec relative to 0 sec ($p < .05$). The modality and spatial interaction was not significant [$F(2,44) < 1$], but the three-way interaction was significant [$F(4,88) = 3.412, p < .05$].

Discussion

In this experiment, we found that the effect of delay is the same for visual matching tasks as for haptic matching tasks. Overall performance was better when there was a short delay between the stimuli than when there was a longer delay, and this effect was not different across the modalities. This finding suggests that the effect of delay found in the previous experiment across modalities was not due to any differences between the modalities, since both show the same effects of ISI.

We also observed an interesting effect with respect to the correct rejection responses across the different spatial dimensions. Our results show that when objects are more discriminable, they are also less affected by delay. Only differences in the *x* and *y* dimensions were affected by ISI, whereas differences in the *xy* dimensions together were not affected by ISI, even as long as 30 sec. We suggest that these differences in the effect of time delay across object discriminability might account for the inconsistencies in the literature on the effects of delay for object matching tasks. Our finding also makes intuitive sense; if the task is to dis-

A

B
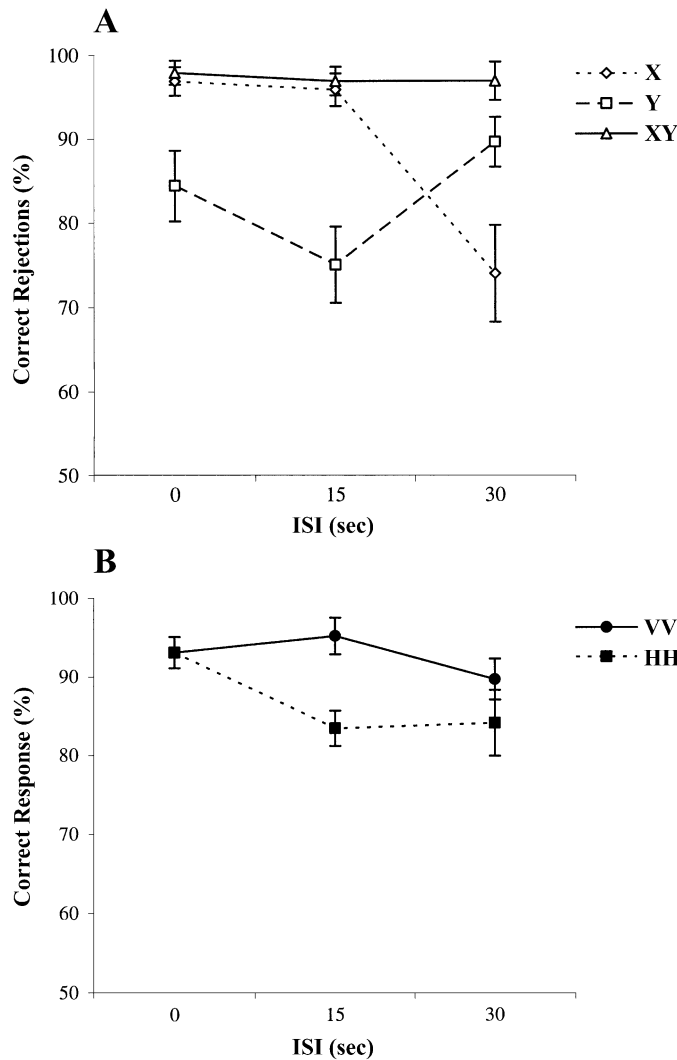
Figure 3. Plot showing percentage correct rejection scores for each spatial change (A) and modality (B) across the different ISIs in Experiment 2.

criminate between two objects from different categories (e.g., a pen and a coffee mug), then we might expect this task to be less sensitive to delay simply because gross feature differences may be more robust to time delays. On the other hand, a task involving objects from within the same category (e.g., between two different coffee mugs) may need to rely on small differences between the objects and may therefore be more sensitive to time delay. Previous studies have found that the recognition of highly similar objects is not only more difficult but more sensitive to small image changes such as changes in viewpoint (Edelman, 1995; Newell, 1998).

EXPERIMENT 3

In the third experiment, we tested simultaneous presentation against successive presentation of objects across

modalities. In the real world, an object of interest is encoded at the same time across the visual and haptic modalities. Thus, temporal synchrony might be key to integrating information across the modalities into a single multisensory representation. If objects are represented into the same multisensory memory system as a result of simultaneous presentation, then we would expect performance in our matching task to be worse for simultaneous presentation than for successive presentation. Our reasoning is as follows: If simultaneous presentation results in mandatory fusion across the modalities, then the consequence of this is that two different objects might be represented as one. Therefore, it would be more difficult to discriminate different objects. On the other hand, separate memory systems may still exist independently of presentation timing. In this case, objects should be equally discriminable, irrespective of presentation order. As in the

previous experiments, we also tested for effects of spatial differences across the presentation conditions.

Method

Participants. Twenty-two undergraduate students (7 male and 15 female) from Trinity College Dublin participated in the experiment for course credit. Their ages ranged from 18 to 28 years, with a mean age of 20 years. All of the participants had normal or corrected-to-normal vision, and none reported any somatosensory impairments.

Stimuli and Apparatus. The stimuli and apparatus were the same as in Experiment 1.

Design. The experiment was based on a two-factor, within-subjects design with presentation order of the modalities and spatial differences as factors. Unlike in Experiment 1, there were three levels to the presentation order factor: VH, HV, and simultaneous presentation. The spatial levels were the same as those described in Experiment 1 (i.e., differences in the x, y, and xy dimensions). Trials were blocked according to the presentation order levels, and the sequence of the blocks was counterbalanced across participants.

Procedure. With the following exception, the procedure was the same as that outlined in Experiment 1: Both the visual and haptic stimuli were presented for 3 sec during the experiment. Thus, in the simultaneous condition, the stimuli were presented for 3 sec in total. We used an ISI of 0 sec for the VH and HV trials and compared performance in these conditions with performance on simultaneous presentation of the stimuli across modalities. The participants' eye movements were informally monitored by the experimenter during the simultaneous presentation condition to ensure that the participants were viewing while touching the objects.

There were 108 trials in the experiment: 36 VH, 36 HV, and 36 simultaneous trials (there were 18 *same* trials and 18 *different* trials for each of these blocks). To familiarize the participant with the procedure, 3 practice trials were given at the start of each block of the presentation order condition. In all other ways, the procedure was the same as described in Experiment 1. The experiment took approximately 80 min to complete.

Results

As in the previous experiments, the mean of each participant's hit and correct rejection scores was calculated. The mean percentage correct (i.e., hits plus correct rejections) scores for the different presentation order conditions are given in Table 3. We conducted a one-way repeated measures ANOVA with presentation as the factor (VH, HV, and simultaneous). No main effect of presentation was found [$F(2,42) = 1.181$, n.s.].

Table 3 also shows the mean RT performance across the presentation conditions. We conducted a one-way ANOVA on the RT data and found a main effect of presentation [$F(2,42) = 79.335$, $p < .001$]. A post hoc Newman–Keuls analysis revealed that performance to the HV condition was significantly faster than to the VH ($p < .01$) or the simultaneous condition ($p < .001$). There was no difference between the HV and simultaneous conditions. This relatively fast performance in the HV condition is not surprising, because participants were faster at encoding and responding to the test stimulus when it was presented as a visual rather than a haptic stimulus.

We then conducted a separate two-way ANOVA on the correct rejection responses with spatial differences and presentation order conditions as factors. We found a main effect of spatial differences [$F(2,42) = 49.827$, $p < .001$].

Table 3
Mean Percentage Correct Responses (Hit and Correct Rejection Scores) and Mean Response Times (RTs, in Milliseconds) Across the Different Presentation Order Conditions in Experiment 3

Presentation Order	% Correct	SE	RT	SE
Simultaneous	84.96	1.56	2,275.4	130.0
Visual–haptic	86.11	1.39	2,151.0	151.8
Haptic–visual	83.81	1.62	951.1	35.2

The mean percentage performance for each of the spatial differences was as follows: 72.7% (1.9% SE) accuracy for x differences, 70.9% (2.3% SE) accuracy for y differences and 93.4% (1.5% SE) accuracy for xy differences. Performance to the changes in the xy dimension was significantly better than performance to changes in either the x or the y dimension (Newman–Keuls, $p < .05$). There was no effect of presentation order [$F(2,42) = 2.272$, n.s.] and no interaction between the factors [$F(4,84) = 0.829$, n.s.].

Discussion

Our finding that there was no difference in performance between simultaneous and successive presentation of objects in our cross-modal matching task suggests that the modalities may work independently and allow for efficient comparison across objects. If objects were stored as multisensory representations, then we would have expected performance to be worse for simultaneous presentation than for successive presentation. But we did not find any differences between the presentation conditions. This finding suggests that the representations of the objects within each modality are sufficiently discrete to allow for cross-modal comparison, even when the objects are presented at the same time to the different modalities.

We also found no difference between the spatial conditions across each of the presentation orders. That is to say, differences in the x, y, or xy dimensions were not any more or less difficult when presented simultaneously than when presented in succession. Overall, however, as in the previous experiments, the results show that performance to differences in xy was better than to differences in the x and y dimensions separately.

GENERAL DISCUSSION

We investigated differences in matching performance across interstimulus time delays between vision and haptics. In Experiment 1, we found that cross-modal matching performance decreased with an increase in time delay between one object stimulus and the next. Furthermore, the rate of performance decrease was the same, irrespective of the modality to which the first stimulus was presented. This result suggested that both the visual and haptic modalities share the same memory characteristics in terms of the effects of delay between stimulus presentations. We tested this suggestion directly in Experiment 2 by comparing performance with increasing ISI across each of the modalities. Again we found no difference in the

effect of ISI across the modalities, although visual performance was better overall than haptic performance. Finally, in Experiment 3, we investigated whether our findings in the previous experiments were due to a single multisensory representation of objects or whether separate, modality-specific representations are maintained. If objects were represented in a single multisensory representation with temporal synchrony, we predicted that simultaneous presentation of object stimuli across the modalities would result in a cost in matching performance in comparison with successive presentation of the objects. We found no difference between simultaneous and successive matching across modalities, suggesting that representations are sufficiently modality specific. We conclude that objects are mainly stored in separate modalities that share similar functional properties of their memory systems. This similarity in properties may allow for efficient access between object representations across the modalities.

In all our experiments, we found that objects which differed by more than one dimension were also more discriminable. This is perhaps not a surprising finding. More interesting, we found that in each modality, there was no effect of interstimulus delay on objects that were more discriminable than on objects that changed on either the *x*- or *y*-axis only (Experiment 2). We suggest that interstimulus similarity may affect performance over different delays and that this may explain the inconsistencies in the literature between effects of delay on cross-modal recognition tasks (Garvill & Molander, 1973; Kiphart et al., 1992). Spatial differences, however, interacted with interstimulus intervals in our within-modal experiment only (Experiment 2) and not in our cross-modal experiment (Experiment 1), although changes in *xy* resulted in better performance in both experiments. In Experiment 1, we failed to find a main effect of ISI on the correct rejections, and this result may have obscured any interactions with the spatial factor. Given that we found an effect of ISI with both hits and correct rejection responses together, it certainly seems that for analysis using only correct rejection responses, our experiment lacked sufficient power for these effects to emerge.

In Experiment 3, we expected to find an interference effect with simultaneous presentation if the objects were stored as a single multisensory representation due to temporal synchrony. Instead, we found no difference between presenting stimuli to the modalities simultaneously or successively. We argue that this finding suggests that modality-specific representations might be maintained such that comparisons between the stimuli can occur. However, it might still be the case that temporal synchrony across modalities results in multisensory representation. In Experiment 3, interference effects may have been avoided because of the difference in the spatial location between the visual and haptic stimuli or a difference between the stimuli themselves, irrespective of the timing of the presentation of the stimuli. Generally, a single object of interest not only will be simultaneously perceived by the modalities, but also will occupy the same spatial location and share the same spatial characteristics. Thus, in our ex-

periment, the objects may not have been integrated because of spatial disparity. The role of spatial disparity in object integration across modalities is currently being investigated.

Our finding that visual and haptic memory for objects share similar functional properties (see also Woods & Newell, in press, for a similar argument) fits with current neuroimaging literature demonstrating similar cortical mechanisms for visual and haptic object recognition (Amedi et al., 2001; James et al., 2002). One interesting observation about these imaging studies is that activation to haptic object recognition resides in a cortical visual area (LOC) implicated in visual object recognition and not in a somatosensory area. This begs the question of whether haptic objects are "recoded" into visual coordinates and stored as visual representations. If so, then our findings may mean that haptic memory for objects is the same as visual memory because objects are recoded into visual memory. Indeed, in Experiment 2, we found that within-modal haptic memory was less efficient than within modal visual memory, suggesting that a recoding was occurring in the previous cross-modal experiments (since there was no difference between the modality orders). Furthermore, our findings from Experiment 3 suggest that visual recoding of haptic stimuli is not mandatory, since we found no interference with simultaneous, cross-modal presentation. Neuropsychological literature also suggests that haptic representations are not necessarily recoded into visual coordinates since impaired memory for recognition in one modality is generally dissociated from performance in the other modality (Farah, 1990; Reed, Caselli, & Farah, 1996). Nevertheless, visual recoding seems to be a good candidate model for cross-modal efficiency, although this may depend on the task. Certainly, in tactile tasks where participants are encouraged to use noninformative vision or a visually determined reference frame, then judgments of spatial differences improve in touch (Millar & Al-Attar, 2004; Newport, Rabb, & Jackson, 2002). One way to test the contribution of visual memory in cross-modal object recognition tasks is to compare matching performance across ISIs in congenitally blind participants and sighted participants. If haptic memory performance is not affected by visual experience, then we would expect no difference between the participant groups. Such studies can further our understanding of the role of visual processing in haptic and cross-modal object recognition.

REFERENCES

AMEDI, A., JACOBSON, G., HENDLER, T., MALACH, R., & ZOHARY, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex*, **12**, 1202-1212.

AMEDI, A., MALACH, R., HENDLER, T., PELED, S., & ZOHARY, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, **4**, 324-330.

EASTON, R. D., GREENE, A. J., & SRINIVAS, K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects. *Psychonomic Bulletin & Review*, **4**, 403-410.

EDELMAN, S. (1995). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics*, **72**, 207-220.

FARAH, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computer*, **35**, 116-124.

Garvill, J., & Molander, B. (1973). Effects of standard modality, comparison modality and retention interval on matching of form. *Scandinavian Journal of Psychology*, **14**, 203-206.

Gibson J. J. (1962). Observations on active touch. *Psychological Review*, **69**, 477-491.

Gilson, E. Q., & Baddeley, A. D. (1969). Tactile short-term memory. *Quarterly Journal of Experimental Psychology*, **21**, 180-184.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, **41**, 1409-1422.

Heller, M. A. (1984). Active and passive touch: The influence of exploration time on form recognition. *Journal of General Psychology*, **110**, 243-249.

James, T. W., Humphrey G. K., Gati, J. S., Servos, P., Menon, R. S., & Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, **40**, 1706-1714.

Kiphart, M. J., Hughes, J. L., Simmons, J. P., & Cross, H. A. (1992). Short-term haptic memory for complex objects. *Bulletin of the Psychonomic Society*, **30**, 212-214.

Klatzky, R. L., Loomis, J. M., Lederman S. J., Wake, H., & Fujita, N. (1993). Haptic identification of objects and their depictions. *Perception & Psychophysics*, **52**, 170-178.

Mahrer, P., & Miles, C. (2002). Recognition for tactile sequences. *Memory*, **10**, 7-20.

Millar, S. (1974). Tactile short-term memory by blind and sighted children. *British Journal of Psychology*, **65**, 253-263.

Millar, S. (1975). Effects of tactual and phonological similarity on the recall of Braille letters by blind children. *British Journal of Psychology*, **66**, 193-201.

Millar, S. (1981). Crossmodal and intersensory perception and the blind. In R. D. Walk & H. L. Pick (Eds.), *Intersensory perception and sensory integration* (pp. 281-314). New York: Plenum.

Millar S., & Al-Attar, Z. (2000). Vertical and bisection bias in active touch. *Perception*, **29**, 481-500.

Millar S., & Al-Attar, Z. (2004). External and body-centered frames of reference in spatial memory: Evidence from touch. *Perception & Psychophysics*, **66**, 51-59.

Newell, F. N. (1998). Stimulus context and view dependence in object recognition. *Perception*, **27**, 47-68.

Newell, F. N. (in press). Cross-modal object recognition. In C. Spence, G. Calvert, & B. Stein (Eds.), *Handbook of multisensory integration*. Cambridge, MA: MIT Press.

Newell, F. N., Bülthoff, H. H., & Ernst, M. O. (2004). *Multisensory enhancement in the recognition of actively explored objects*. Manuscript in preparation.

Newell, F. N., Ernst, M. O., Tjan, B. S. & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, **12**, 37-42.

Newport, R., Rabb, B., & Jackson, S. R. (2002). Noninformative vision improves haptic spatial perception. *Current Biology*, **12**, 1661-1664.

Reales, J. M., & Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: Cross-modal priming depends on structural descriptions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 644-663.

Reed, C. L., Caselli, R. J., & Farah, M. J. (1996). Tactile agnosia: Underlying impairment and implications for normal tactile object recognition. *Brain*, **119**, 875-888.

Sullivan, E. V., & Turvey, M. T. (1972). Short-term retention of tactile stimulation. *Quarterly Journal of Experimental Psychology*, **24**, 253-261.

Woods, A. T., & Newell, F. N. (in press). Visual, haptic and cross-modal recognition of objects and scenes. *Journal of Physiology*.

NOTES

1. The presentation timings of both the visual and haptic stimuli during learning were carefully determined in a pilot study to ensure that there were no encoding differences across the modalities. Such differences might have caused differential rates of memory decay over time because the representation in one modality may be weak compared to the representation in the other. Thus, pilot studies allowed us to determine the presentation timings across modalities that resulted in equal cross-modal performance for the 0-sec ISI.

2. The spatial factor was analyzed separately because the *same* trials were identical for all levels; therefore, hit data could not be separated across the levels.