Spatial Cognition & Computation, 9:287–304, 2009

ISSN: 1387-5868 print/1542-7633 online
DOI: 10.1080/13875860903305664

Taylor & Francis
Taylor & Francis Group

Evidence for Amodal Representations after Bimodal Learning: Integration of Haptic-Visual Layouts into a Common Spatial Image

Nicholas A. Giudice,[1] Roberta L. Klatzky,[2] and Jack M. Loomis[3]

[1]Department of Spatial Information Science and Engineering, University of Maine, Orono, Maine, USA
[2]Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[3]Department of Psychology, University of California, Santa Barbara, California, USA

Abstract: Participants learned circular layouts of six objects presented haptically or visually, then indicated the direction from a start target to an end target of the same or different modality (intramodal versus intermodal). When objects from the two modalities were learned separately, superior performance for intramodal trials indicated a cost of switching between modalities. When a bimodal layout intermixing modalities was learned, intra- and intermodal trials did not differ reliably. These findings indicate that a spatial image, independent of input modality, can be formed when inputs are spatially and temporally congruent, but not when modalities are temporally segregated in learning.

Keywords: haptics, visual perception, memory, orientation

1. INTRODUCTION

Although the spatial cognition literature has tended to focus on visual apprehension of the environment, there is growing interest in how mental representations built up from vision compare to those from nonvisual inputs. Since the primary spatial senses of hearing, touch, and vision encode 3-D information and support spatial behavior in a common physical world, the representations built up from each input must interact so as to seamlessly support action. As to the form of this interaction, it has been proposed that at some

Correspondence should be sent to Nicholas Giudice, Department of Spatial Information Science and Engineering, University of Maine, Orono, Orono, ME 04469. E-mail: Giudice@spatial.maine.edu

level, inputs from different channels lead to a common spatial representation (Bryant, 1997; Jackendoff, 1987; Miller & Johnson-Laird, 1976). In addition to support from several lines of behavioral data, discussed next, this proposal is consistent with a growing body of electrophysiological and neuroimaging literature demonstrating convergence of visual, auditory, and tactile inputs in common brain regions subserved by multisensory information processing (see Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005; Driver & Noesselt, 2008; Macaluso & Driver, 2005 for reviews).

Critical evidence for a common spatial representation is provided when performance in a spatial task is independent of the modality through which the information was initially encoded, an outcome referred to as *functional equivalence* (see Loomis, Klatzky, Avraamides, Lippa, & Golledge, 2007 for discussion). In particular, equivalent performance has been demonstrated when people update their self-position relative to target arrays learned from vision, spatialized sound, touch, and even spatial language (Avraamides, Loomis, Klatzky, & Golledge, 2004; Giudice, Betty, & Loomis, under revision; Klatzky, Lippa, Loomis, & Golledge, 2003; Loomis, Lippa, Golledge, & Klatzky, 2002).

In accounting for equivalent performance across modalities, Loomis and associates introduced the concept of a spatial image, a three-dimensional representation of external space which resides in working memory (Loomis et al., 2002). The spatial image can be formed in multiple ways: from spatial perception within sensory modalities, read-in from long-term memory, or from constructive spatial processes arising through imagination and language (Loomis & Klatzky, 2007).

The spatial image is different from a 2D "depictive" image (e.g., Kosslyn, 1980, 1994), in which the mental representation of a source constitutes a direct mapping of its 2D projection. These latter images are intrinsically visual and move with the person who holds them. Spatial images, in contrast, are externalized in the world. They remain stationary when people move, allowing updating of self-position. They need not be visual; they could be amodal, as discussed next.

A spatial image is proposed to be the result of perceptual and/or cognitive encoding processes, and as such, there is no requirement that it be veridical. For example, distance tends to be compressed in both auditory spatial perception (Ashmead, Davis, & Northington, 1995; Loomis, Klatzky, Philbeck, & Golledge, 1998) and haptic perception (Abravanel, 1971). Corresponding biases would also be expected in the spatial images formed from these inputs, as has been found with auditory encoding (Klatzky, Lippa, Loomis, & Golledge, 2003; Loomis, Lippa, Golledge, & Klatzky, 2002) and from alignment biases of representations built up from touch (Giudice et al., under revision) or verbal descriptions (Avraamides & Pantelidou, 2008; Shelton & McNamara, 2004; Wilson, Tlauka, & Wildbur, 1999). As discussed further later, for tests of functional equivalence it is important to adjust the objective input so as to compensate for modality-specific biases in encoding that might

otherwise produce differences in performance. It is also important to ensure that spatial layout is learned to the same criterion across modalities, so that spatial images are equally accessible and precise.

If a spatial image encompasses multiple objects, it would be expected to exhibit grouping effects imposed by the input modality. One such effect is the tendency for grouping to follow temporal proximity (e.g., in audition, Warren & Verbrugge, 1984; in vision, Lee & Blake, 1999; Palmer, 2002). Grouping by time is exploited here, albeit at a scale much coarser than typical effects for spatial perception. Specifically, we ask whether separate spatial images are formed when inputs are presented during discrete time-periods.

Given the conception of a spatial image as functioning equivalently across modalities, three competing explanations could account for this phenomenon. The Separate-But-Equal hypothesis attributes equivalent spatial behavior across different inputs to the formation of sensory-specific spatial images that afford common behaviors. The hypothesis suffers, however, from lack of explanatory or predictive power, as it offers no fundamental principle by which sensory-specific images would result in equivalent performance across modalities.

Loomis, Klatzky, and associates (Avraamides, Loomis, Klatzky, & Golledge, 2004; Klatzky et al., 2003; Loomis et al., 2002) proposed that a sufficient condition for functional equivalence is that different input channels converge on a common representation. This leads to two other explanations of the phenomenon, as follows. The Common-Recoding hypothesis explains equivalence by assuming that all inputs are recoded into spatial images within the same modality, most probably visual. Evidence against this hypothesis comes from experiments where participants were asked to walk to or orient toward known spatial locations, either directly or after a rotation or translation. Comparable performance with indirect and direct responses indicates spatial updating during the movement. Contrary to the idea of Common Recoding, similar spatial updating has been demonstrated between sighted and blind participants, the latter presumably not relying on a visually recoded image (Giudice et al., under revision; Loomis et al., 2002). The Amodal Hypothesis proposes that if inputs are matched so as to compensate for modality-specific biases and learning rates, they will be encoded into a spatial image that is invariant across the input channel and that is not tied to any sensory or cognitive input source. Our own approach favors this hypothesis, and the results of the current experiments provide further empirical support of its efficacy.

The present experiments are derived from the account of functional equivalence in terms of amodal spatial images. They test conditions that promote or inhibit the encoding of multiple objects into a single, integrated spatial image of their layout. In particular, the idea of amodality suggests that integration of objects into a spatial image of their layout should not depend on the objects' input modality. This has not been tested in previous research on spatial updating. Rather, the body of research showing equivalent

performance in spatial updating adopted a design whereby participants first learned the location of one or more targets from a single input channel, and then were required to localize the objects from a new position in space. Functional equivalence was assessed by comparing the results of updating in layouts learned entirely by vision, audition, touch, or language. If functional equivalence results from use of amodal spatial images, however, it should not require isolating sensory modalities at the time the images are encoded. To the contrary, we commonly access different points in the world by different senses. We may hear a sound emanating from one location in space while we see an object at a second location and touch another object at a third. Accordingly, the concept of functional equivalence suggests that inputs from multiple channels could be combined into an integrated representation of space that contains amodal spatial images.

Although the spatial image is proposed to be independent of whether inputs are segregated or mixed by modality, there are likely constraints on this independence. One such constraint, addressed in these experiments, is the temporal contiguity of encoding experience. That is, if objects from different modalities are encoded from distinct temporal events, their integration into a common spatial image is likely to fail. This prediction derives from the idea that a spatial image groups the layout of objects in the world by time as well as space, consistently with temporal organization commonly found in perception. The layout may change over time, if the observer or the objects move; however, the spatial images represent some coherent world event. Accordingly, circumstances that support the encoding of multiple objects as a single event should facilitate their integration into a single image, whereas separation of object presentations into temporally distinct events should discourage integration.

In short, our hypotheses are that mixing modalities during encoding should not impair the formation of a spatial image, unless the modalities are also segregated into distinct encoding events. To pursue these predictions, the current studies investigated functional equivalence using a different paradigm from the previous work. Participants learned a layout of six objects by rotating in the center of a circular environment. Half of the objects were apprehended haptically on a table in the physical environment and half were exposed visually on a shelf in a virtual cylindrical environment. After learning all targets to criterion, participants were tested on their ability to make pointing judgments between target pairs. For these test trials, a Start and End target location were named, and the participant was asked to first imagine facing the Start location and then to point to the End location. The task requires that people imaginally update their position relative to multiple objects in the array, and has previously been found to be highly effective in tests of spatial processing within visual virtual reality (Kelly, Avraamides, & Loomis, 2007).

The task allows for the recording of two latency measures and an error measure. *Orientation time* represents the time to imaginally update self-

position relative to the named start target. *Processing time* represents the time to determine one's spatial relation to the end target, when facing the start. The error measure, namely, the absolute difference between the angle of response and correct angle between the start and end targets, represents noise in the spatial judgment. (In principle, absolute error also includes systematic bias, but as we did not observe consistent bias in the signed error data, we interpret absolute error as noise.) The principal analysis is directed toward differences in these measures when the start and end targets are in the same vs. different modalities.

If the two modalities have been integrated into a common amodal spatial image, then the computation of interobject angles should not depend on their modal source. If, however, the modalities are separated into distinct spatial images, intermodal computation should produce longer processing times and higher errors than intramodal, due to the need to register the two images for the computation of the response. We call the difference between intermodal and intramodal response times the *switching cost.*

The use of multiple measures allows more detailed analyses to be performed. It is possible, for example, that orientation time will be lower for start objects in one of the modalities, regardless of the end object, suggesting intermodal differences in the accessibility of objects in the spatial image. There may be differences between the intermodal pairs (e.g., visual-haptic trials produce faster processing times than haptic-visual trials), suggesting that the cost of switching between two modality-discrepant spatial images depends on the direction of the transition.

In Experiment 1, we asked whether temporally segregating the haptic and visual objects into two distinct learning blocks would impede the formation of a common spatial image. If so, performance requiring judgments of relative direction between two objects in the same modality (visual-visual or haptic-haptic) should be faster and/or more accurate than that for updating between modalities (haptic-visual or visual-haptic), which requires that two spatial images be brought into registration. That is, there should be switching cost. In Experiment 2, we investigated whether mixing haptic and visual objects within a single learning event would lead to their integration within a single spatial image. If so, performance requiring judgments across two objects in the same modality (e.g., visual-visual or haptic-haptic) should be equivalent to that for updating between modalities (e.g., haptic-visual or visual-haptic); that is, there should be no switching cost.

In addition to testing the foregoing hypotheses, these studies were intended to extend tests of functional equivalence to comparisons of the haptic and visual modalities. While haptic spatial updating has been demonstrated for object arrays and scenes learned by touch (Barber & Lederman, 1988; Hollins & Kelley, 1988; Newell, Woods, Mernagh, & Bulthoff, 2005; Pasqualotto, Finucane, & Newell, 2005), there has been relatively little effort to demonstrate equivalent performance in spatial tasks after haptic and visual learning (but see Giudice et al., in revision).

2. EXPERIMENT 1

The method of Experiment 1 was intended to produce distinct spatial images for vision and haptics and thus to demonstrate a strong switching cost. Participants learned the haptic and visual objects as two layouts within the same region of space, but separated by a 30-second interval. They were then tested on relative orientations of object pairs drawn either within the same modality or across the two modalities.

2.1. Method

2.1.1. Participants. Eighteen participants, ages 18–26 (mean = 19.2), balanced equally by gender, took part in the study. All gave informed consent and received course credit for their time.

2.1.2. Apparatus. Participants stood in the center of a circular table (0.91 m inner diameter and 0.91 m high), which was also depicted in a virtual cylindrical environment. Six objects (ball, cup, spoon, wrench, stapler, and glasses) were placed around the physical/virtual layout at the 1, 3, 5, 7, 9, and 11 o'clock positions. To ensure they could be identified, the objects were shown to participants, who then were required to name each one by touch and virtual vision before starting the experiment. Object layouts were learned separately, with three of the objects apprehended haptically on the physical table and three presented visually on a virtual shelf slightly above the table. The targets were spatially intermixed, but were only felt or seen one at a time, as the participant rotated through 360° (see Figure 1 for an illustration).

During learning, participants wore a Virtual Research V8 head-mounted display (HMD) to see the virtual environment and visual targets. The HMD's earphones were also used to deliver audio messages during the testing period. An Intersense IS-300 inertial tracker was mounted on the HMD to update the virtual environment during rotation. The tracker had an accuracy of 3° RMS and a resolution of 0.02°. During testing, pointing responses were made using a joystick affixed with a 1 m extension designed to improve haptic cues about the stick's position. The stick was held in the dominant hand; pushing a button mounted on the top of the stick logged the change in time and facing direction from stimulus onset. Version 3.0 of the Vizard software package (WorldViz, Santa Barbara, CA; www.worldviz.com) was used to coordinate the sequence of experimental trials and record latency and pointing data.

2.1.3. Design and Procedure. The experiment comprised three phases: practice, learning, and testing. A within-participants design was used, with all participants learning one layout of three haptic objects and one layout of three visual objects. Testing comprised 24 pointing trials, where the start and end

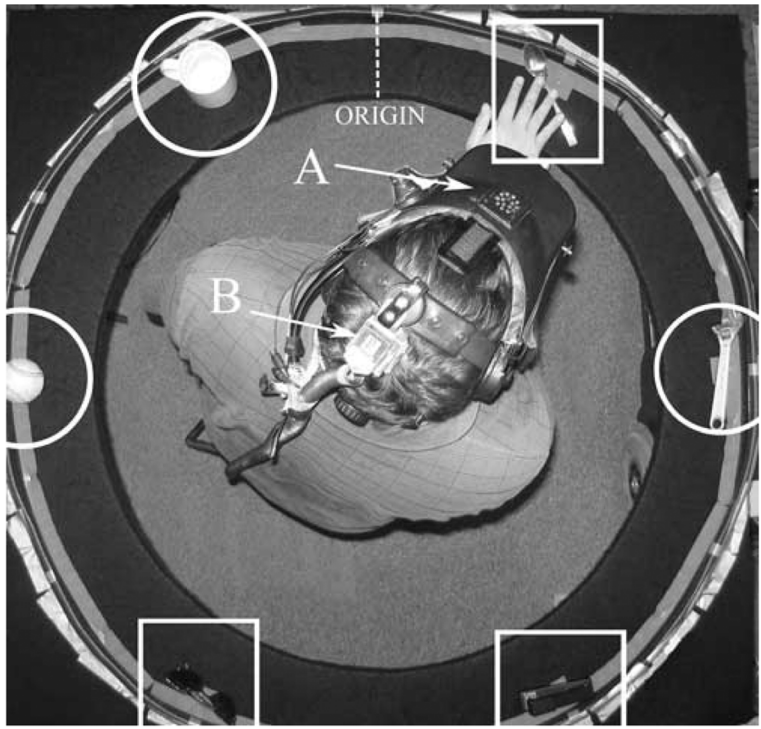

Figure 1. An overhead view of a participant in the bimodal training environment. Apparatus: (A) head-mounted display for viewing the virtual objects and (B) inertial tracker for updating the display during user rotation. Six interspersed visual targets (circled) and haptic targets (squared) are placed at 60° angular separation around the circular environment. In the actual experiment, objects are perceived from a first person perspective and exposed individually during rotation, with haptic objects felt on the physical table and visual objects seen through the HMD on a shelf in the virtual layout.

targets constituted four modality combinations: haptic-haptic, visual-visual, haptic-visual, and visual-haptic. Each test combination occurred equally often.

During the *practice phase*, participants stood in the center of the circular table, represented as a virtual cylinder in the VE, and were required to point correctly to the six clock-face locations used in the experiment. They were instructed to point the joystick as if it was fixed in the center of the circle (in actuality, it was slightly offset as the participant stood in the center and the stick was placed on the floor in front of them). Meeting the pointing criterion ensured that participants had correctly calibrated their pointing judgments for this offset and could accurately indicate all target directions with the same deflection speed and throw distance of the stick for each. They were then

given a sample bimodal environment and run through the complete learning and testing sequence with corrective feedback provided.

Participants began the *learning phase* from a 0° origin in either the haptic or visual layout, designated by a tactile or visual line indicator respectively. The joystick was removed. To be exposed to targets, they rotated in place by following an arrow displayed through the HMD until they reached a randomly chosen target location. The actual target objects were not revealed during this outbound guidance. Upon reaching the target location, the arrow disappeared, and the target was made visible on the virtual shelf (visual condition) or felt by reaching forward to touch it on the physical table (haptic condition). Targets could be seen or felt, but they could not be manipulated.

For both haptic and visual layouts, participants rotated through 360° in order to be exposed to each of the target locations (left-right rotations were counterbalanced between participants). During rotation, participants either kept their hands on the table to find the 3 haptic targets, or looked at the virtual shelf, slightly above the physical table, to find the 3 visual targets. To ensure that only one target could be apprehended at a time, participants were instructed to keep their hands and head fixed in the forward position as they rotated. In the haptic condition, use of both hands was allowed as long as they were held side by side on the table.

In the visual condition, the virtual shelf was seen through an aperture providing approximately the same field of view as was available from touch. Backtracking was not allowed. After being exposed to either the haptic or visual three-object layout (layout order was counterbalanced between participants), participants were guided back to the 0° start position for a 30-second wait interval. They were then guided to the first target position in the second layout and performed the same task (direction of outbound guidance alternated between layouts).

Following the two exposures, participants were guided back to the 0° origin position and asked to point, via the joystick, to each of the three target positions for each layout (target order was randomized). If they passed the learning criterion (mean absolute pointing error less than 15° for both haptic and visual layouts), they moved on to the test phase. If they failed on one or both modalities, they relearned the failed layout(s) until they either met the learning criterion or completed six new exposure trials. (When both layouts failed, the re-learning order alternated between modalities.) If one modality was mastered before the other, requiring further exposures of the failed modality, the completed condition was re-exposed before moving on to the test.

During the *testing phase*, the HMD was used as a blindfold and the joystick and all targets were removed from the environment. Participants were disoriented by having them rotate in several directions in place, ending in an arbitrary direction (facing direction at test was never at a target location or at the 0° origin). The joystick was then placed in front of the participant and the test trials proceeded with judgments of relative direction (JRDs). For each JRD, the participant was first given an auditory instruction, delivered

through the HMD's headphones, "You are facing target x." The participant was instructed to imagine facing the designated start position, and when ready, to press the button mounted on the joystick. The next auditory instruction was then delivered, in the form, "point to target y." The participant again pressed the button after pointing.

The procedure allowed us to record two latency measures and an error measure. *Orientation time* is defined as the interval between the onset of the start target and the button press indicating the participant had imagined facing this target. *Processing time* is the interval between naming of the end target, which follows the first button press, and the second button press, made after completion of the pointing judgment. The latter measure includes not only the time needed to determine the response direction, but also the latency of the actual pointing response. The pointing latency is expected to play a minimal role, however, as participants were instructed to execute their response only after they clearly knew the intended direction, and variations in joystick control across angle were minimal. The *absolute error* measure is defined as the absolute difference between the response angle and the objective angle between the start and end targets.

Six replications of the four start-end modality combinations (visual-visual, visual-haptic, haptic-visual, haptic-haptic) were given at test, divided into two blocks of 12 trials each. Trials were randomized within block and balanced equally across 60°, 120°, and 180° angular separation between start and end target pairs for each condition.

2.2. Results and Discussion

Participants required an average of 2.8 haptic trials and 3.1 visual trials to reach the learning criterion, a nonsignificant difference $t(17) = 2.11$, $p = 0.64$. Although combined error was less than 15° for all participants after six learning trials, three participants did not meet criterion within the six-trial limit for visual layouts and two did not meet criterion for the haptic layouts (all were within 7° of the 15° passing threshold). Since the individual test performance of these participants did not reliably differ from the overall sample, their data were included in all subsequent analyses. The last trial of several participants' test data was lost due to a corrupted logging function, and outliers outside the range of ± 3 standard deviations of the mean for the latency data were removed from the analyses, representing less than 2% of the complete dataset.

No effect was found for block order during test on the variables of interest; thus, the data were collapsed across block for all of the analyses. Orientation times did not reliably differ between modalities: The average latency for imagining start targets was 1.89 sec for haptic targets vs. 2.07 for visual targets, $t(35) = 1.66$, $p = 0.11$. These findings suggest that retrieval and instantiation of a named perspective did not depend on the modality being queried.

Absolute error and processing time directly measure participants' ability to access the spatial relations between the start and end target. These data are shown in Figure 2 as a function of the angular difference between the targets, for each combination of the two start and end target modalities (haptic vs. visual). Of greatest interest was the predicted interaction between start and end modality, such that performance should be facilitated when the two targets were of the same modality (indicating a switching cost for inter-modal trials). This trend can be seen in the figure. ANOVAs were performed on each

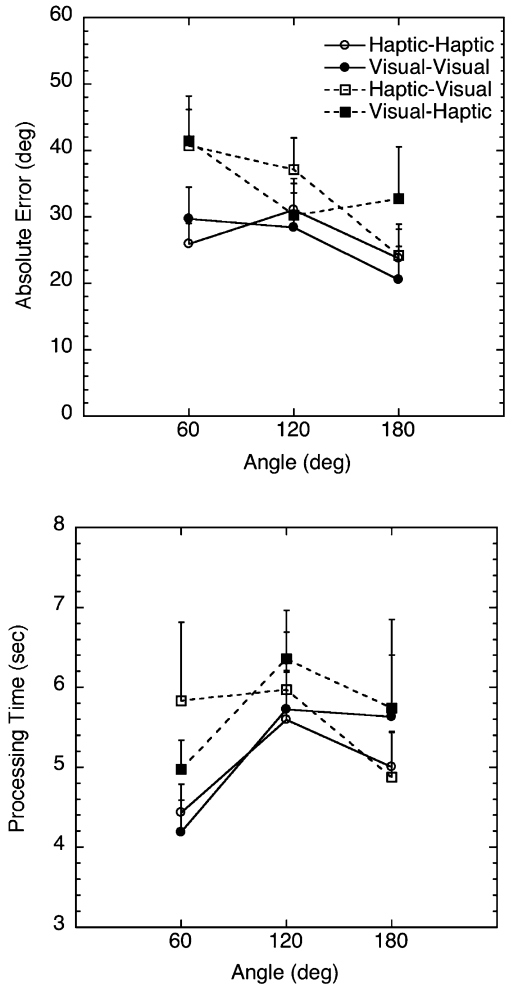

Figure 2. Absolute error (top) and processing time (bottom) in Experiment 1 by angle and by combination of start and end modalities. Error bars are 1 s.e.m.

of the measures with factors angle (60°, 120°, 180°), start modality (haptic, vision) and end modality (haptic, vision).

For the error measure, the predicted interaction was the only significant effect, $F(1, 17) = 7.00$, $p = .017$, $\eta^2 = .29$. For the processing time measure, there was a significant effect of angle, $F(2, 34) = 3.70$, $p = .035$, $\eta^2 = .18$. There was a clear trend for within-modality processing to be faster than between-modality at the oblique angles (60° and 120°), which were more difficult than the 180° angle, but with all angles included, the predicted interaction between start and end targets only approached significance, $F(1, 17) = 3.76$, $p = .069$, $\eta^2 = .18$. When only the oblique angles were considered, the predicted interaction was significant for processing time, $F(1, 17) = 5.88$, $p = .027$, $\eta^2 = .26$. The performance differences as a function of angle are not surprising, as canonical angles such as 90° and 180° are generally computed more accurately (Tversky, 1981) and show less variability than obliques on angle estimation tasks after haptic and visual learning (Appelle, 1971; Lakatos & Marks, 1998).

Paired *t*-tests (two-tail) were used to compare the cost of switching from vision to haptics (visual-start-haptic-end minus both visual) versus the reverse (haptic-start-visual-end minus both haptic). The difference was not significant for either error (cost = 8.6° vs. 7.1°, respectively, $t(17) = 0.32$) or processing time (cost = 510 msec vs. 548 msec, respectively, $t(17) = 0.08$), $ps > .75$. Thus switching costs exhibit no dependence on the start and end modalities.

In short, the results indicate that the two modalities produced equally accessible spatial images, as measured by orientation time. Moreover, when the start and end targets were both of the same modality, processing the angle between targets was equivalent in time and error for vision and haptics, indicating that the accessibility of one target from another within a modality did not differ. Importantly, however, the finding that intermodal judgments produced significantly greater error, and for the oblique angles, greater processing time, supports the hypothesis that distinct spatial images were developed for the two modalities. The observed switching costs then presumably reflect the extra processing necessary to retrieve images from different layouts and register them relative to one another, as will be explained in more detail later. We next ask whether it is possible to form a single, coherent spatial image of objects encoded from different modalities, when the modalities are not segregated in time.

3. EXPERIMENT 2

Having demonstrated evidence against integration of modalities into a single image when they were segregated as encoding events, we now turn to an experiment that promotes their integration and address whether the previous switching costs are mitigated. In Experiment 2, modalities were intermixed during encoding, and judgments of relative angle were again used to assess

whether a common spatial image was formed. If so, contrasting with Experiment 1, judgments of angle that cross-modalities should lead to the same time and error as intramodal judgments.

3.1. Method

Eighteen young adults (ages 18–22, mean = 18.9) participated, 10 male and 8 female. All gave informed consent and received course credit for their time. The apparatus and procedure were identical to Experiment 1, except that learning of visual and haptic targets was interleaved in a single bimodal layout. The virtual shelf with the visual objects occluded the locations of the haptic objects on the table and the participant's hands (see Figure 1 for an illustration). This arrangement facilitated the experience of visual and haptic objects as being in a common spatial layout. During learning, participants rotated through 360° in order to be exposed to each of the six target locations. While rotating, they kept their hands on the table to find the three haptic targets, and looked at the virtual shelf, slightly above the physical table, to find the three interspersed visual targets.

If the learning criterion (point to each target location with an average absolute error of 15° or less) was not met after the first exposure of the six target locations, participants were guided to the start position and made another 360° rotation to re-learn the locations (direction of turning alternated between exposures). The learn/test sequence continued until the learning criterion was met or six exposure periods were completed. Judgments of relative direction then proceeded as in Experiment 1.

3.2. Results and Discussion

During the learning phase, participants required an average of 3.9 trials to reach the criterion. Three participants did not pass within the six-trial limit (but all were within 6° of the 15° passing threshold). Since the individual test performance of these participants did not reliably differ from the overall sample, their data were included in all subsequent analyses. The last trial of several participants' data was lost due to a corrupted logging function, and outliers outside the range of ±3 standard deviations of the mean for the latency data were removed from the analyses, representing less than 2% of the complete data set. No effect was found for block order during test on the variables of interest; thus, the data were collapsed across block for all of the analyses.

Orientation times did not reliably differ between modalities: Virtually identical latencies were observed for imagining haptic start targets (m = 1.49 sec) and visual start targets (m = 1.53 sec), $t(35) = 0.43$, $p = 0.73$. Note that the orientation time was significantly less than for Experiment 1, $t(17) = 2.80$, $p = .0082$, indicating that orienting to the first stimulus took less time when the objects had been integrated into a single array.

The magnitude of this interexperiment difference in orientation time was 470 msec; this effect will be discussed further later.

The absolute error and processing time measures, shown in Figure 3, were again examined with repeated-measures ANOVAs having factors of angle, start modality, and end modality. In the analysis of error, the only effect was a marginal effect of angle, $F(2, 34) = 3.12$, $p = .057$, $\eta^2 = .16$. The interaction of interest, between start modality and end modality, was not statistically reliable, $F(1, 17) = 2.40$, $p = .14$, $\eta^2 = .12$. In the analysis of processing time, only the angle effect was significant, $F(2, 34) = 6.66$,

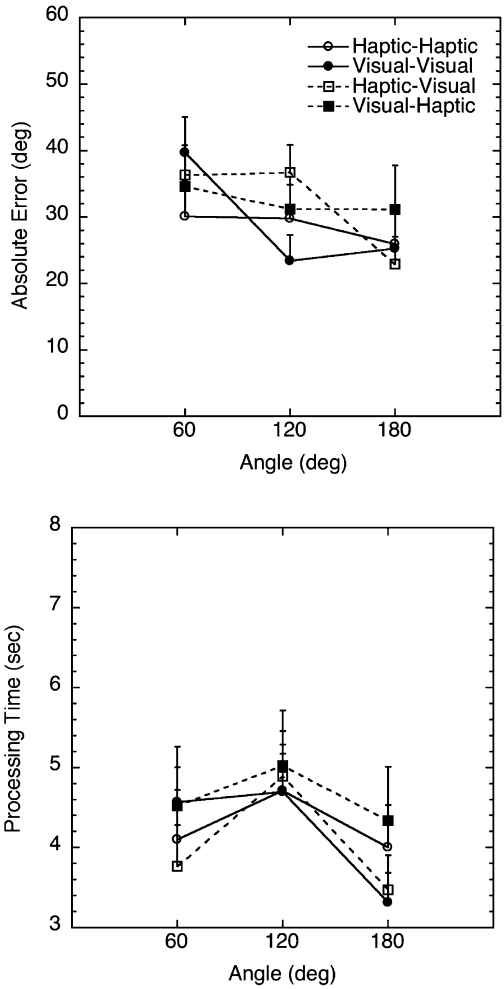

Figure 3. Absolute error (top) and processing time (bottom) in Experiment 2 by angle and by combination of start and end modalities. Error bars are 1 s.e.m.

$p = .004$, $\eta^2 = .28$. The interaction between start and end modality did not approach significance, $F(1, 17) = .38$, $p = .547$, $\eta^2 = .02$.

As can be seen from Figure 3, similar performance was observed for both measures across the four start-end modality pairings. Because the conclusion of functional equivalence requires accepting the null hypothesis, and given that the difference between intermodal and intramodal judgments was in the predicted direction for error and processing time, we pursued the issue with further tests. Paired *t*-tests comparing the average intermodal and intramodal absolute error and processing time proved to be marginal for error, $t(17) = 1.55$, one-tailed $p = .07$; for processing time the effect did not approach significance, $t(17) = .62$, one-tailed $p = .27$. With respect to the absolute error measure, statistical power with alpha set to .05 was 18.1%. A sample size more than an order of magnitude greater than the current one would be necessary to definitively exclude the null hypothesis, while maintaining probability of Type 2 error at .05.

Although acceptance of the null hypothesis may not be fully merited by these data, it is clear that relative to the pronounced switching costs of Experiment 1, which can be taken to be the benchmark for distinct spatial images, the effects in Experiment 2 are small. The mean modality effect (same vs. different modality) was 7.9° for error and 529 msec for processing time in Experiment 1; in Experiment 2 these values were reduced to 3.1° and 104 msec. Thus segregating the modalities within a single event impaired angular judgments substantially, more than doubling error and moving processing time into a range suggestive of cognitively demanding processes such as mental imagery (Farah & Kosslyn, 1991).

4. GENERAL DISCUSSION

The present results show a clear contrast between the cross-modal availability of spatial location information, depending on whether spatial arrays are learned in an intermixed vs. temporally segregated fashion. The data suggest that when two layouts are learned in sequence, judgments of relative direction between the two arrays imposes extra processing load. We interpret these results in the context of a task model in which the directional judgment has the following basic components: forming a spatial image containing the start target, orienting to the start target, accessing the end target, and determining its direction relative to the start target. If the start and end target are in the same array, only one spatial image need be formed, and it can be formed in anticipation of the start target's being named. When there are two separate spatial images, formed from segregated object arrays, there are two consequences: First, the array in which the start target resides is ambiguous until that target is named; hence, there is an additional time needed to retrieve the spatial image containing the start target and orient to it. Second, if the end target is in a different array from the start target,

additional time is needed to retrieve the spatial image containing that target and bring it into spatial correspondence with the spatial image containing the start target.

This model leads to two empirical comparisons to test the idea that learning temporally segregated modalities leads to two different representations, whereas temporal integration creates a unitary array. The first, which was the principal focus of the present experiments, is the switching cost, that is, the processing-time difference between intermodal and intramodal judgments of relative direction. When modalities were intermixed in Experiment 2, no switching cost was expected, under the hypothesis that regardless of modality, the start and end targets are resident in the same spatial image. In contrast, with the segregated modalities of Experiment 1, a cost was expected. This is the case because with an intramodal test, both the start and end objects reside in the initially instantiated array, whereas with the intermodal test, the second array must be called up when the end object is named, adding to the measure of processing time. Thus the switching cost represents the time to instantiate the array of the second object when its modality has changed. Our data show that this effect is approximately 500 msec.

The second comparison is between the orientation times of the two experiments. Specifically, when there is only one array, as in Experiment 2, the spatial image of the objects can be formed in advance, whereas when two arrays have been learned, forming the spatial image of the start object must wait until the object is named. By this reasoning, the time to form the spatial image of an array can be measured by the difference in the initiation time between Experiment 1 and 2, which is again approximately 500 ms.

These results support the idea that modality differences present no barrier to the integration of object locations into a common spatial image. They also indicate that temporal segregation, to the contrary, is an impediment to the formation of a common spatial image. This leads to the question of what other manipulations might lead to separate spatial images, for example, separating different groups of objects into distinct regions of space.

Another question raised by these results is whether spatial images formed at separate times might ever be capable of being integrated. With sufficient testing after learning, for example, would participants in the present Experiment 1 come to bring the spatial images of seen and touched objects together? Relevant data come from Avraamides and associates (Avraamides, Loomis, Klatzky, & Golledge, 2004), who compared judgments of relative direction, made without vision, among sets of objects that had been learned in two conditions. In one case, several objects were visually exposed in a room one at a time, and in the other, all objects were exposed simultaneously.

The judgments of relative direction were slower after sequential learning than after simultaneous exposure; however, this difference vanished when the participants moved backward, then pointed to each object from the new perspective, before making the directional judgments. These results suggest that spatial updating relative to the set of individually learned objects was neces-

sary to bring them into a common spatial image and indicates a manipulation that might be tried with temporally segregated arrays as used in Experiment 1.

Although we conclude that integration of objects across modalities into a common spatial image is supported by these data, there may be a small cost to switching modalities within a single spatial image that cannot be reliably detected with the current paradigm. Such a cost could arise from nonspatial processes that retain memory tags for the modal origins of the objects, which are not precluded by the existence of an amodal image.

As a final point, one could question whether the present data constitute evidence for "amodal," as opposed to "multimodal" representations. By amodal, we mean a single representation that is accessible by multiple modalities but is associated with none of them in particular. The term "multimodal" is more ambiguous. If it is taken to mean a collection of modalities that retain functional autonomy, this would be equivalent to the Separate-but-Equal hypothesis considered and rejected in the introduction. Another sense of multimodal is that sensory origins are retained, but the representations function as a whole at a population level (cf. Stilla & Sathian, 2008). In the present task, for example, haptic and visual objects might be encoded as such, but entered jointly into a common frame of reference, which would itself be amodal. This formulation cannot be ruled out by the present data, but it seems to defer the notion of amodality to a second-order level of processing, not to dismiss it entirely. Moreover, it is not clear how population-based computations encompassing multiple, distinct modalities would lead to the modality-invariant performance found in tasks like spatial updating.

ACKNOWLEDGMENT

The authors thank Brendan McHugh and Kevin Verlatti for their invaluable assistance in running participants. This work was supported by NSF grant BCS-0745328.

REFERENCES

Abravanel, E. (1971). Intersensory integration of selected spatial dimensions: Extension to an adult sample. *Perceptual and Motor Skills, 32*(2), 479–484.

Amedi, A., Von Kriegstein, K., Van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research, 166*(3–4), 559–571.

Appelle, S. (1971). Visual and haptic angle perception in the matching task. *American Journal of Psychology, 84*(4), 487–499.

Ashmead, D. H., Davis, D. L., & Northington, A. (1995). Contribution of listeners' approaching motion to auditory distance perception. *Journal of Experimental Psychology: Human Perception & Performance, 21*(2), 239–256.

Avraamides, M. N., Loomis, J. M., Klatzky, R. L., & Golledge, R. G. (2004). Functional equivalence of spatial representations derived from vision and language: Evidence from allocentric judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 804–814.

Avraamides, M. N., & Pantelidou, S. (2008). Does body orientation matter when reasoning about depicted or described scenes? In C. Freksa, S. N. Newcombe, P. Gärdenfors, & S. Wölfl (Eds.), *Spatial cognition VI: Lecture notes in artificial intelligence* (vol. 5248, pp. 8–21). Berlin: Springer.

Barber, P. O., & Lederman, S. J. (1988). Encoding direction in manipulatory space and the role of visual experience. *Journal of Visual Impairment & Blindness, 82*(3), 99–106.

Bryant, K. J. (1997). Representing space in language and perception. *Mind and Language, 12*(3), 239–264.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron, 57*(1), 11–23.

Farah, M. J., & Kosslyn, S. M. (1981). *Structure and strategy in image generation. Cognitive Science, 5*(4), 371–383.

Giudice, N. A., Betty, M. R., & Loomis, J. M. (under revision). Functional equivalence of spatial images from touch and vision: Evidence from spatial updating in blind and sighted individuals.

Hollins, M., & Kelley, E. K. (1988). Spatial updating in blind and sighted people. *Perception & Psychophysics, 43*(4), 380–388.

Jackendoff, R. (1987). *Consciousness and the computational mind.* Cambridge, MA: MIT Press.

Kelly, J. W., Avraamides, M. N., & Loomis, J. M. (2007). Sensorimotor alignment effects in the learning environment and in novel environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 1092–1107.

Klatzky, R. L., Lippa, Y., Loomis, J. M., & Golledge, R. G. (2003). Encoding, learning, and spatial updating of multiple object locations specified by 3-d sound, spatial language, and vision. *Experimental Brain Research, 149*(1), 48–61.

Kosslyn, S. M. (1980). *Image and Mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and Brain*. Boston: MIT Press.

Lakatos, S., & Marks, L. E. (1998). Haptic underestimation of angular extent. *Perception, 27*, 737–754.

Lee, S.-H., & Blake, R. (1999). Visual form created solely from temporal structure. *Science, 284*, 1165–1168.

Loomis, J. M., & Klatzky, R. L. (2007). Functional equivalence of spatial representations from vision, touch, and hearing: Relevance for sensory substitution. In J. J. Rieser, D. A. Ashmead, F. F. Ebner, & A. L. Corn (Eds.), *Blindness and brain plasticity in navigation and object perception* (pp. 155–184). Mahwah, NJ: Lawrence Erlbaum Associates.

Loomis, J. M., Klatzky, R. L., Avraamides, M., Lippa, Y., & Golledge, R. G. (2007). Functional equivalence of spatial images produced by perception and spatial language. In F. Mast & L. Jäncke (Eds.), *Spatial processing in navigation, imagery, and perception* (pp. 29–48). New York: Springer.

Loomis, J. M., Klatzky, R. L., Philbeck, J. W., & Golledge, R. G. (1998). Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics, 60*(6), 966–980.

Loomis, J. M., Lippa, Y., Golledge, R. G., & Klatzky, R. L. (2002). Spatial updating of locations specified by 3-d sound and spatial language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(2), 335–345.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neurosciences, 28*(5), 264–271.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Newell, F. N., Woods, A. T., Mernagh, M., & Bulthoff, H. H. (2005). Visual, haptic and crossmodal recognition of scenes. *Experimental Brain Research, 161*(2), 233–242.

Palmer, S. (2002). Perceptual grouping: It's later than you think. *Current Directions in Psychological Science, 11*, 101–106.

Pasqualotto, A., Finucane, C. M., & Newell, F. N. (2005). Visual and haptic representations of scenes are updated with observer movement. *Experimental Brain Research, 166*(3–4), 481–488.

Shelton, A. L., & McNamara, T. P. (2004). Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 158–170.

Stilla, R., & Sathian, K. (2008). Selective visuo-haptic processing of shape and texture. *Human Brain Mapping, 29*, 1123–1138.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology, 13*, 407–433.

Warren, W. H., Jr., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception & Performance, 10*, 704–712.

Wilson, P. N., Tlauka, M., & Wildbur, D. (1999). Orientation specificity occurs in both small- and large-scale imagined routes presented as verbal descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 664–679.