2012

Spatial working memory for locations specified by vision and audition: Testing the amodality hypothesis

Jack M. Loomis

Roberta L. Klatzky

Brendan McHugh

Nicholas A. Giudice
University of Maine - Main, nicholas.giudice@maine.edu

Atten Percept Psychophys (2012) 74:1260–1267
DOI 10.3758/s13414-012-0311-2

Spatial working memory for locations specified by vision and audition: Testing the amodality hypothesis

Jack M. Loomis · Roberta L. Klatzky · Brendan McHugh · Nicholas A. Giudice

Published online: 3 May 2012

Abstract Spatial working memory can maintain representations from vision, hearing, and touch, representations referred to here as *spatial images*. The present experiment addressed whether spatial images from vision and hearing that are simultaneously present within working memory retain modality-specific tags or are amodal. Observers were presented with short sequences of targets varying in angular direction, with the targets in a given sequence being all auditory, all visual, or a sequential mixture of the two. On two thirds of the trials, one of the locations was repeated, and observers had to respond as quickly as possible when detecting this repetition. Ancillary detection and localization tasks confirmed that the visual and auditory targets were perceptually comparable. Response latencies in the working memory task showed small but reliable costs in performance on trials involving a sequential mixture of auditory and visual targets, as compared with trials of pure vision or pure audition. These deficits were statistically reliable only for trials on which the modalities of the matching location switched from the penultimate to the final target in the sequence, indicating a switching cost. The switching cost for the pair in immediate succession means that the spatial images representing the target locations retain features of the visual or auditory representations from which they were derived. However, there was no reliable evidence of a performance cost for mixed modalities in the matching pair when the second of the two did not immediately follow the first, suggesting that more enduring spatial images in working memory may be amodal.

Keywords Working memory · Visual perception · Audition

J. M. Loomis (✉) · B. McHugh
Department of Psychological and Brain Sciences,
University of California, Santa Barbara,
Santa Barbara, CA 93106, USA
e-mail: jmloomis99@gmail.com

R. L. Klatzky
Department of Psychology, Carnegie Mellon University,
Pittsburgh, PA 15213, USA

N. A. Giudice
Spatial Informatics Program, School of Computing and
Information Science, University of Maine,
Orono, ME 04469, USA

Over the last 3 decades, much research has been devoted to spatial working memory (e.g., Awh & Jonides, 2001; Baddeley & Lieberman, 1980; Jonides et al., 1993; Logie, 1995; McCarthy et al., 1994; Shah & Miyake, 1996). The vast majority of behavioral and brain-imaging studies have focused on vision, and of these studies, most have made use of 2-D computer displays for the research. Because spatial working memory plays a role in the control of action, it would be expected to involve 3-D space and to retain information about auditory and haptic targets as well. Indeed, an abundance of studies using visual targets has demonstrated the ability to spatially update the 3-D locations of perceived target locations stored in working memory during blind locomotion (e.g., Loomis, Da Silva, Fujita, & Fukusima, 1992; Ooi, Wu, & He, 2001; Rieser, 1989). As well, several studies have shown that haptic and auditory locations can be updated in working memory during observer locomotion (e.g., audition: Ashmead, DeFord, & Northington, 1995; Klatzky, Lippa, Loomis, & Golledge, 2003; Loomis, Klatzky, Philbeck, & Golledge, 1998; Loomis, Lippa, Klatzky, & Golledge, 2002; touch: Giudice, Betty, & Loomis, 2011; Hollins & Kelley, 1988; Pasqualotto, Finucane, & Newell, 2005). Our group has coined the term *spatial image* to refer to the contents of spatial working memory, whether deriving

from visual, auditory, or haptic input, within the context of 3-D space. The spatial image is a representation of the location and other spatial properties (e.g., orientation) of one or more targets. Furthermore, besides arising from visual, auditory, and haptic input, the spatial image can be instantiated in working memory from long-term memory (Easton & Sholl, 1995; Giudice, Klatzky, Bennett, & Loomis, in press; Rieser, Garing, & Young, 1994; Wang, 2004) and from linguistic input (Avraamides, Loomis, Klatzky, & Golledge, 2004; Klatzky et al., 2003; Loomis et al., 2002). For a general overview of our theoretical framework and empirical research, see Loomis, Klatzky, and Giudice (in press).

There is general interest in commonalities and interactions across the sensory modalities that goes beyond our direct concern here with multisensory inputs to spatial working memory. This interest is exemplified by the extensive literature on multisensory integration, intersensory conflict, and cross-modal plasticity (e.g., Bowen, Ramachandran, Muday, & Schirillo, 2011; Calvert, Spence, & Stein, 2004; Ernst & Banks, 2002; Sadato et al., 1996; Sathian & Lacey, 2007; Spence & Driver, 2004; Stein & Stanford, 2008). Although research on these topics is not of direct relevance to the present study, we note that there are important ties between research on multisensory perception and research on working memory in tasks where sensory information is acquired over successive eye fixations, over successive haptic samples during hand exploration, and over auditory samples obtained during head rotation.

We view the spatial image as a representation that enables action in the absence of direct perceptual support. In line with our focus on the action-relevant content of spatial working memory are many other studies on human spatial cognition (e.g., Mou, McNamara, Valiquette, & Rump, 2004; Waller & Hodgson, 2006; Wang, 2004). Especially relevant is the work of Byrne, Becker, and Burgess (2007; see also Burgess, 2006), for it is concerned in part with short-lived spatial representations in working memory suitable for guiding action. The authors presented a functional and neural model of spatial cognition in humans and other species based initially on neurophysiological studies in the rat. It details the neural network underlying the constant interplay between egocentric perceptual representations that are concurrent with sensory stimulation and perspective-free (allocentric) representations of surrounding space (e.g., cognitive maps) that remain well after the perceptual representations are gone. Spatial working memory plays a central role in mediating between the perceptual information and allocentric representations stored in long-term memory within medial-temporal areas of the brain.

Our research has led us to hypothesize that the spatial image is amodal in nature; once a spatial image has been formed in spatial memory, its processing by subsequent mental operations, such as those involved in spatial updating, does not depend on any of the modality-specific features of the input modality (see Bryant, 1997, for a very similar idea). Two studies provide some support for the hypothesis of amodal representations in spatial working memory. Giudice et al. (2011) had participants make spatial judgments based on simple layouts learned by touch or vision; they made these judgments immediately after learning. Across a variety of conditions, some of which involved spatial updating, haptically based performance was remarkably similar to visually based performance in terms of errors and response times. By itself, this result is consistent with three hypotheses: (1) Modality-specific spatial images are modality specific but functionally equivalent (separate but equal hypothesis), (2) haptic percepts are recoded into visually based spatial images (visual recoding hypothesis), and (3) spatial images are amodal (amodality hypothesis). An ancillary experiment showed similar performance by blind observers on the haptic version of the task, ruling out the visual recoding idea. Because we find it implausible that modality-specific spatial images from vision and touch would be so nearly equivalent with respect to response latencies (separate but equal hypothesis), we favor the amodality hypothesis, but the argument is not airtight. Another study by Giudice, Klatzky, and Loomis (2009) had participants make judgments of relative direction of simple layouts of objects that were perceived exclusively by vision, exclusively by touch, or by a sequential mixture of the two senses. The response to indicate the direction to the target object was made by moving an extended joystick in the appropriate direction. There was a nonsignificant trend toward greater absolute error with the mixture of modalities, while the response latencies, averaging over 4.0 s, revealed no significant costs. However, because the joystick response included judgment time, response latency, and movement time, the absence of a reliable cost associated with the mixture might have reflected multiple sources of variability for the response and the consequently limited statistical power.

The present experiment was motivated by the desire for a task addressing the issue of amodality in spatial working memory using a response of minimal complexity, with its variation presumably reflecting just judgment time. Here, observers performed a working memory task involving visual and auditory targets at various locations differing in direction but equal in distance. On a given trial, a sequence of up to four targets was presented, with the four targets being all visual targets, all auditory targets, or a sequential mixture of the two. Observers had to press a button as quickly as possible when detecting a repetition of the same location, regardless of the modalities involved. We were interested in whether accuracy and response latencies would show a cost associated with a mixture of the two modalities, for a performance cost would be evidence against amodality.

In this regard, we considered three possible outcomes and their associated hypotheses. First, if all differences in encoding between modalities have been controlled for and spatial images are amodal, the spatial image for a given trial will always be the same regardless of the modality from which it was encoded (amodality hypothesis). Without any modality-specific features, the performance on the repetition detection task cannot depend on composition of the targets in terms of sensory modality, implying equal error rates and response latencies for constant modality and mixed-modality trials. Second, separate spatial stores might exist, one for each modality (separate but equal hypothesis). If so, mixed-modality trials would require switching between stores when the second target at the matching location is presented, which would increase the time taken to detect the repetition by the amount of the switching time. Errors would also likely increase. Third, spatial images might briefly retain some trace of their modal origin but lose this trace as they are maintained in working memory. We refer to this as the transient modality-specific tags hypothesis. It has elements of both the separate-but-equal hypothesis and the amodality hypothesis. Evidence of this would be a large performance cost when the two targets that match in location immediately follow one another but a reduced performance cost when the two targets appear farther apart in the sequence. As will be seen, the results of the study favor the third alternative.

The spatial image is associated with a location in 3-D space (both distance and direction), not with direction alone. Although the targets varied in direction only, our experiment assumed that the visual and auditory targets were matched in distance and direction. Because visual and auditory targets at the same physical distance can be perceived to lie at different distances when presented without explicit information signifying equidistance, we designed our experiment to ensure that the auditory and visual targets were perceived at the same distance, as is explained further in the Method section.

Method

Observers Twenty undergraduate students (10 female) participated in a single 90-min session for payment; all gave informed consent. None reported having any known deficiency involving vision or hearing.

Design considerations We would like to have used more than five locations varying in direction in our experiment, for doing so would have permitted longer sequences of trials in which only one location matched, but the constraints of visual perception and auditory location limited us to five directions. With visual fixation straight ahead, we needed to confine the visual targets to retinal eccentricities that were not too extreme (greater than 60º). With directional localization of audition being much less precise than that of vision, we wished to keep the directional separation for the auditory targets several times larger than the minimum audible angle of about 7º at an azimuth of 60º (Blauert, 1997; Mills, 1972). Thus, we settled on these azimuths relative to straight ahead: −60º, −30º, 0º, 30º, and 60º, with negative values signifying the left side.

Although distance was kept constant throughout the experiment, we also wished to ensure that the auditory and visual targets were perceived to be the same distance away. In complete darkness, point light targets less than 3 m are perceived to be more distant (e.g., Ooi et al., 2001; Philbeck & Loomis, 1997), and auditory targets are generally perceived as closer than their physical distances (e.g., Loomis et al., 1998). Because adjusting the distances of the visual and auditory targets in advance presents its own difficulties, we chose to allow the observers to see the layout of visual targets and loudspeakers between trials of the working memory task, turning off all light only during the target sequence. We allowed viewing of the layout because vision can "capture" sounds when the separation between them is small (Bowen et al., 2011). Accordingly, memory for layout during the short trials would keep the perceived locations of the auditory and visual targets congruent. At the same time, presenting the visual targets in the dark meant that each visual target was perceived in the absence of concurrent visual information.

Apparatus and stimuli Observers were seated in a chair with an attached desk and chinrest. The desk was covered with a fabric, on which push-buttons could be positioned using Velcro fasteners. Arrayed in a semicircle in front of the observer were five target lights and five loudspeakers colocated on five microphone stands. The five loudspeakers (Philmore model TS36, 8.9-cm diameter) were at eye level and positioned 1.50 m in front of the eyes at the five azimuths mentioned above. The five target lights were white LEDs mounted just beneath the rims of the corresponding speakers and 2 cm more distant. A sixth LED was positioned 1 cm below the LED target at 0º as a fixation target. The lights and speakers were activated under the control of a computer using an external switchbox. When illuminated, the highly directional LEDs, which were aimed at the observer's eyes, were 6.0 log units above photopic (cone) threshold; the fixation target was continuously illuminated throughout the experiment such that it was just visible with the room lights off. With the room lights off, the laboratory appeared completely dark to the observer, except for the fixation light and the occasional appearance of one of the target LEDs. The five loudspeakers presented white noise at 73 dB at the observer's position, as measured using a sound level meter on A-weighting (Realistic Model 33-2066, Fort

Worth, TX). In all tasks of the experiment, the observer responded using push-buttons in different arrangements, according to the task. Scheduling of trials, stimulus presentation, and recording of push-button responses were controlled by a script written in the Vizard scripting language version 3.0 (WorldViz, Santa Barbara, CA) on a desktop computer.

Procedure The experiment consisted of four different tasks: a familiarization task, a localization task, a detection task, and a spatial working memory (SWM) task, in that order. The familiarization, localization, and detection tasks were run twice, first for vision (or audition) and then for audition (or vision), with the order of modalities counterbalanced over the 20 observers. These tasks were followed by the SWM task, which was of primary interest. The preceding tasks served to prepare the observer for the SWM task and to provide performance measures of secondary interest. For all tasks, the observer maintained fixation on the dim LED directly in front, with head position fixed by the chinrest.

In the familiarization task, five push-buttons were positioned from left to right in front of the observer, such that the five fingers of the preferred hand rested comfortably on them. The observer was presented with the five visual (or auditory) targets in order for 1.0 s each. Immediately after each target, the computer delivered synthetic speech specifying the target number ("1" to "5"), and the observer was to press the spatially corresponding push-button. The targets were then presented in reverse order, with similar responding by the observer. Data from this task were not retained.

The purpose of the localization task was to have the observer learn the different target locations and to obtain performance measures on how well the visual and auditory targets could be localized in terms of direction. For each modality, the localization task consisted of three blocks of 20 trials each, in which each of the five targets were presented 4 times in random order for 1.0 s each, and the observer pressed the spatially corresponding button as quickly as possible. In the first block of trials, the observers received feedback when they made errors. In the second and third blocks, no feedback was given. After each block of 20 trials, two 40-W tungsten lamps behind the observer were illuminated for 10 s to keep the observer from fully dark adapting.

The purpose of the detection task was to confirm that the visual and auditory targets were detected in approximately the same amount of time as indicated by similar detection response times. In addition, any slight systematic differences observed in the detection response times for vision and hearing could be taken as differences in neural transmission times for the two types of targets, differences that could be used to explain similar differences in response time in the working memory task. In the detection task, a single push-button was used by the index finger of the observer's preferred hand. This task consisted of three blocks of 20 trials each, in which each of the five targets were presented 4 times in random order for 1 s each, and the observer was to press the button as quickly as possible following signal onset. To prevent anticipation, the interstimulus interval ranged between 0.5 and 4 s. As in the localization task, the two lamps were illuminated for 10 s after every block of 20 trials.

On a given trial in the SWM task, the observer was presented with a sequence of targets, all visual, all auditory, or a mixture of the two. The sequence lengths were evenly distributed among values of 2, 3, and 4, with each visual or auditory target lasting 0.5 s, and with successive targets being separated by 1.0 s, thus resulting in sequence durations ranging from 2 to 5 s. Sequence lengths were randomly interspersed. On two thirds of the trials, one of the target locations was repeated (making it the last in the sequence), and observers were to press a single push-button with the index finger of the preferred hand as soon as a repetition was detected. Pressing the button in the absence of a repetition was counted as a false alarm. On a no-repetition trial, the absence of a push-button response counted as a correct rejection, and failing to press when a repetition had occurred counted as a miss. Verbal feedback was given using synthetic speech. "Correct" was issued when either a correct repetition was detected (a hit) or the observer made a correct rejection; otherwise, "incorrect" was issued.

The SWM task consisted of six blocks of 27 trials each. Each block consisted of three subblocks of 9 trials, with one subblock comprising trials of sequences with visual targets only, another block comprising trials of sequences with auditory targets only, and the remaining block comprising trials of sequences with both visual and auditory targets. The order of the three subblocks was randomized within each block of 27 trials. For mixture trials, the different proportions of auditory and visual targets were equally probable. For all trials involving a repetition of a location, the first target of the repeating pair occurred with equal probability in each of the sequence positions preceding the last target; thus, for a sequence of three targets ending with a repetition, the first of the repeating pair could be in the first position in the sequence or in the second position, with equal probability. After every 9 trials, the two 40-W lamps were illuminated for 10 s to prevent complete dark adaptation. In addition, after every trial, the computer display was fully illuminated with white light for 3 s. It provided enough light for the observer to dimly see the layout of speakers and lamps, thus ensuring that observers would perceive the visual and auditory targets as equally far away during the subsequent target sequence that was conducted in complete darkness. After the display went dark, a blank interval of 1.0 s preceded initiation of the next trial.

Results

Detection task The mean response latencies for blocks 2 and 3 in the detection task, filtered of a few extreme outliers (0.3 %), are given in the lower part of Fig. 1 as a function of modality and target azimuth. The mean auditory response time was 26.3 ms longer than the mean visual response time. Although a small difference, a two-way (modality × azimuth) repeated measures ANOVA revealed that auditory response latencies were significantly longer, $F(1, 19) = 16.81$, $p = .001$, $\eta_p^2 = .469$. The ANOVA also indicated that there was neither a main effect of target azimuth nor a reliable modality × azimuth interaction (ps > .25 in both cases). Thus, it can be concluded that the two types of targets were readily detected, but with a small difference in response latency.

Localization task In the localization task, error rates in blocks 2 and 3 were very low: two errors total out of 800 trials for vision and five errors total out of 800 trials for audition. The low error rates indicate that observers were able to discriminate the locations in both modalities with very high accuracy. We expected this to be the case given that the 30 ° angular separation between targets was over 10 times the angular discrimination threshold for vision (Anderson, Mullen, & Hess, 1991) and over 4 times that for audition (Blauert, 1997; Mills, 1972) throughout the angular range of targets used. The mean response latencies for both correct and incorrect trials, filtered of a few extreme outliers (0.3 %), are given in the upper part of Fig. 1 as a function of modality and target azimuth. The mean auditory response time was 117 ms longer than the mean visual response time, but because of large variability in the difference scores across observers, this difference did not attain statistical significance; a two-way (modality × azimuth) repeated measures ANOVA resulted in $F(1, 19) = 3.597$, $p > .05$. In addition, there was a main effect of target azimuth, $F(4, 76) = 2.854$, $p = .029$, but no reliable modality × azimuth interaction ($p > .2$). Both the similar error rates and the response times that were not reliably different indicate that the visual and auditory targets were comparable in localizability.

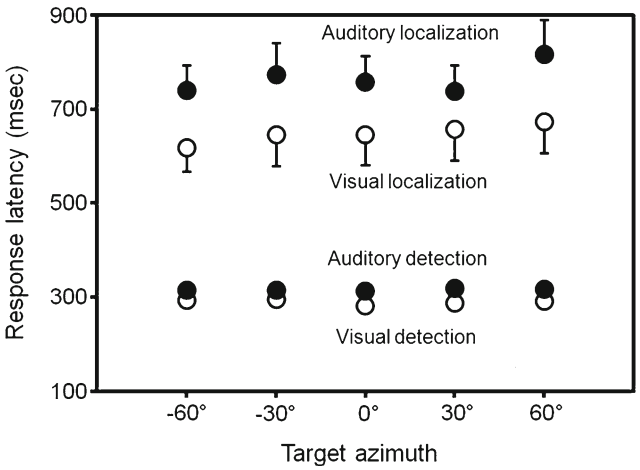

Fig. 1 Mean response latencies for the detection and localization tasks as a function of target azimuth. The error bars shown for localization are standard errors of the means. The error bars for detection fall within the symbols. Because the error bars reflect between-observer variability, they are somewhat misleading with respect to statistical significance, which depends only on the variability of the within-observer difference scores

SWM task Of greatest interest are the results of the SWM task. Prior to the analysis, 22 out of 3,240 trials were discarded because observers told the experimenter at the time that they had mistakenly made a buttonpress. For the remaining valid trials, the task was performed with very high accuracy (98.7 %, 98.4 %, and 97.9 % correct for audition, vision, and mixed modalities, respectively). The miss rates for audition, vision, and mixed trials were 1.7 %, 1.1 %, and 2.3 %, respectively; the corresponding false alarm rates were 0.6 %, 2.5 %, and 1.7 %.

The subsequent analysis of response latencies is based only on trials on which there was a correct detection of a repeat in the target location. Because there were five possible target locations and a maximum of four targets on a trial, observers could not anticipate that a repeated location would occur on the last target presentation, as they would be able to for a trial of five targets. A one-way repeated measures ANOVA of the SWM response latencies showed a significant effect of trial type, $F(2, 38) = 16.66$, $p < .001$, $\eta_p^2 = .467$, with mean latencies of 627, 601, and 671 ms, respectively, for trials of audition, vision, and mixed modalities. A subsequent analysis broke the mixed trials down into those ending with an auditory target and those ending with a visual target, since it seemed likely that the 26.3-ms difference in auditory and visual response times in the detection task might reflect a difference in neural transmission times for the visual and auditory targets. The mean response time for all SWM trials ending in an auditory target (for both mixed and pure trials) was 646 ms, and that for all trials ending in a visual target (mixed and pure trials) was 620 ms, for a difference of 25.8 ms, which is in close agreement with the 26.3-ms value for the detection task. The fact that the auditory response times were slightly larger than the visual response times poses no difficulty for the subsequent analysis. This variation is compensated for by contrasting the response times for trials of a single modality (e.g., vision) with mixture trials ending with the same modality (viz., vision).

The response latencies for the SWM task are given in Fig. 2. The solid symbols in each panel are the mean latencies for each of the 20 observers. The left panel plots values for the mixed trials ending with a visual target against the values for all-vision trials, and the right panel gives the

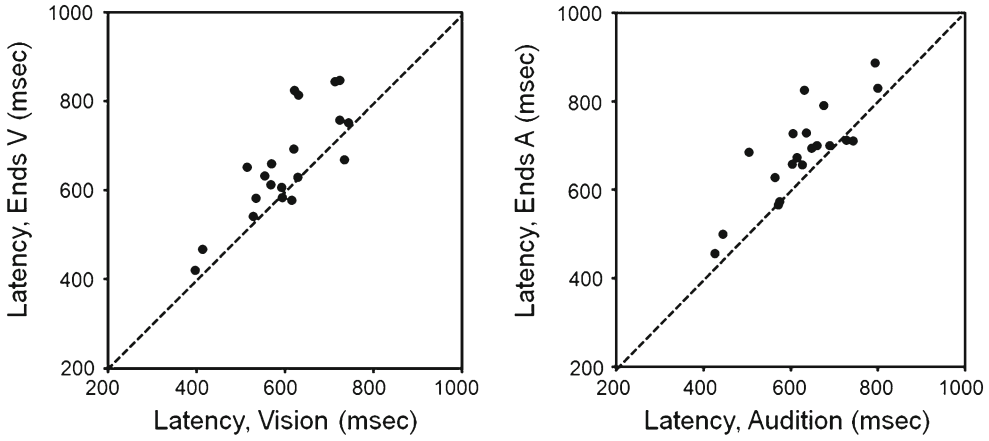

Fig. 2 Response latencies in the spatial working memory task. The panel on the left plots the mean response latencies for the 20 observers in the mixed trials ending with a visual target (Ends V) against the mean latencies for the all-visual trials. The panel on the right plots the mean response latencies for the 20 observers in the mixed trials ending with an auditory target (Ends A) against the mean latencies for the all-auditory trials

corresponding results for audition. It is evident that despite the large variability in response times (over 400 ms), mixed trials have reliably longer response times than do pure trials for both vision and audition (vision, 57.2 ms; audition, 56.3 ms). A two-way repeated measures ANOVA (modality of last target on a trial vs. mixed or pure trials) revealed a statistically significant main effect of modality, $F(1, 19) = 6.85$, $p = .02$, $\eta_p^2 = .265$, a significant effect of trial type, $F(1, 19) = 21.99$, $p < .001$, $\eta_p^2 = .54$, and no interaction, $F(1, 19) < 1$.

An even more detailed analysis provides some insight into the reason for the performance decrement produced by mixing modalities within a trial. This analysis focused on the modalities of the targets falling at the same location only for trials within the mixed-modalities condition. We classified the trials by the modalities of the repeated targets and by the lag between the first presentation of the location and its repeated presentation, where a lag of 1 signifies immediate succession. Because target sequences of lengths 2, 3, and 4 were equiprobable in our experiment, a lag of 1 occurred most frequently, followed by a lag of 2 and then by a lag of 3. This means that the mean latencies were based on declining *N* values as lag increased. Figure 3 shows the mean response latencies by lag and by the modalities of the targets for the repeated location (e.g., AA vs. AV). Here, we were interested in whether, for each lag, there was a performance decrement for nonmatching pairs relative to matching pairs. Matched-sample *t*-tests showed that there were reliable performance decrements only for a lag of 1. For the comparison of AA versus VA, $t(19) = 1.74$, $p < .05$, one-tailed, and for the comparison of VV versus AV, $t(19) = 3.71$, $p < .001$, one-tailed. The associated switching costs are 85 and 63 ms for sequences ending with vision and audition, respectively. The *p* values for the remaining four *t*-tests were all greater than .20.

Interestingly, the values for a lag of 2, which were more reliable than those for a lag of 3, showed very little performance decrement for nonmatching pairs relative to matching pairs; for target pairs ending with vision, the performance decrement was only 5 ms, and for target pairs ending with audition, the decrement was only 12 ms. Importantly, the nearly equal latencies for the lag of 2 supports the assumption that the visual and auditory targets were perceived as having the same direction and distance, as we had desired, adding to the evidence from the earlier analysis showing very low error rates. Had the visual and auditory targets been at different perceived locations, we would have expected a performance decrement in the response latencies for all values of lag, as well as a higher error rate in the mixed-modality condition relative to the pure vision and pure audition conditions.

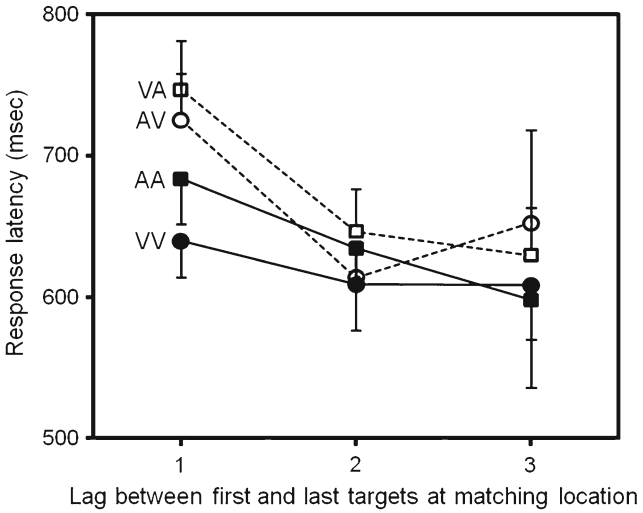

Fig. 3 Mean latencies in the spatial working memory task as a function of lag and the modalities of the repeated target location. The first letter of the label (A [auditory] or V [visual]) represents the modality of the first target, and the second letter represents the modality of the second target at the same location. Error bars are standard errors of the mean. For purposes of clarity, only the lower ones are shown for the matching modalities, and only the upper bars are shown for the nonmatching modalities. Because the error bars reflect between-observer variability, they are somewhat misleading with respect to statistical significance, which depends only on the variability of the within-observer difference scores

Discussion

In the introduction, we mentioned the study by Giudice et al. (2009) involving the integration of visual and haptic targets within spatial working memory, the results of which favored the amodality hypothesis. However, the primary evidence for this conclusion might have suffered from reduced statistical power because of a multistage response with long latency. The present SWM task optimized the conditions for detecting a small performance cost associated with the mixed condition by using a binary judgment involving a simple buttonpress. Indeed, the response latencies averaged only 633 ms, which is approximately one seventh of the mean response time in the Giudice et al. (2009) study.

To ensure the validity of our SWM task, we took care to guarantee that our visual and auditory targets were well above detection threshold, highly localizable in direction, and perceived at the same distance. The similar response latencies of the detection task indicated that the two modalities were well matched in terms of stimulus detectability, although there was a small but reliable difference of 26 ms, with vision being faster. The very low and nearly identical error rates in the localization task, along with response latencies that were not reliably different, indicate that observers were able to identify the five target locations with comparable levels of performance. In addition, the very high accuracy of performance in the SWM task showed that observers performed the task equally well in the auditory and visual conditions (98.7 % and 98.4 %, respectively), providing strong evidence that the auditory and visual targets were highly localizable in direction. Just as important, the fact that the error rate of 97.9 % in the mixed-modalities condition was virtually the same as in the other two conditions indicates that observers did not perceive discrepancies in the directions of the visual and auditory targets. Finally, the analysis of response latencies for a lag of 2 in the most fine-grained analysis of the SWM task indicates that the perceived distances of the visual and auditory targets were not noticeably different. From all this evidence, we conclude that the visual and auditory targets were well suited for comparing performance based on the SWM task.

The primary focus of this study was on the SWM task as a test of the amodality hypothesis. Trials ending with an auditory target resulted in response latencies about 26 ms longer than those ending with a visual target, mirroring the results seen in the detection task. This difference was of no consequence, for in the analysis of greatest interest, we compensated for it by contrasting the response times for trials of a single modality (e.g., vision) with those for mixture trials ending with the same modality (viz., vision). The results speak against the amodality hypothesis (that spatial images have no trace of their modal origins) because the analysis of the response latencies showed reliable effects of mixing modalities within a trial. The comparison of mixed trials ending in visual or auditory targets with the corresponding all-vision and all-audition trials showed performance costs of 57 and 56 ms, respectively. While these mean difference scores were small in comparison with the between-observer variability (of over 400 ms), they were nonetheless highly reliable statistically.

In the introduction, we considered two alternatives to the amodality hypothesis: the separate-but-equal hypothesis and the transient modality-specific tags hypothesis. Our most fine-grained analysis, represented by Fig. 3, came out in favor of the second alternative. When the matching location was specified by the penultimate and final targets, there were very reliable switching costs of 85 and 63 ms for sequences ending with vision and audition, respectively. However, when the lag between the targets representing the same location was either 2 or 3, there was no reliable performance cost associated with nonmatching pairs, as compared with matching pairs, and the more precise latencies for a lag of 2 showed virtually no difference. This indicates that even for the short durations of the target sequences used here (2–5 s), the results are consistent with amodality for lags greater than 1, thus ruling out the separate-but-equal hypothesis and supporting the transient modality-specific tags hypothesis. This conclusion is consistent with the result of Giudice et al. (2009), which support amodality for spatial images held in working memory for longer periods of time. More research needs to be done to confirm and extend these results. Neuroimaging research will undoubtedly provide further elucidation as to the role of modal and amodal storage. It is interesting in this regard to note that a PET imaging study of visual and auditory localization that made use of both pointing and delayed matching-to-sample tasks found that the superior parietal lobe was modality specific and that the inferior parietal lobe was amodal (Bushara et al., 1999).

Author Note This work was supported by NSF Grant BCS-0745328 and NIH Grant R01-EY016817. The authors thank two anonymous reviewers and the editor, Adriane Seiffert, for very helpful comments, Jerome Tietz for technical assistance, and Ted Hsu for assistance in conducting the experiment.

References

Anderson, S. J., Mullen, K. T., & Hess, R. F. (1991). Human peripheral spatial resolution for achromatic and chromatic stimuli: Limits imposed by optical and retinal factors. *The Journal of Physiology, 442,* 47–64.

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

Avraamides, M., 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,* 801–814.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences, 5,* 119–126.

Baddeley, A. D., & Lieberman, K. (1980). Spatial working memory. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 521–539). Hillsdale, NJ: Erlbaum.

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. Cambridge, MA: MIT Press.

Bowen, A. L., Ramachandran, R., Muday, J. A., & Schirillo, J. A. (2011). Visual signals bias auditory targets in azimuth and depth. *Experimental Brain Research, 214,* 403–414.

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

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends in Cognitive Sciences, 10,* 551–557.

Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., & Hallett, M. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience, 2,* 759–766.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review, 114,* 340–375.

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Easton, R. D., & Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 483–500.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433.

Giudice, N. A., Betty, M. R., & Loomis, J. M. (2011). Equivalence of spatial images from touch and vision: Evidence from spatial updating in blind and sighted individuals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 621–634.

Giudice, N. A., Klatzky, R. L., Bennett,C. R., & Loomis, J. M. (in press). Combining locations from working memory and long-term memory into a common spatial image. *Spatital Cognition and Computation*.

Giudice, N. A., Klatzky, R. L., & Loomis, J. M. (2009). Evidence for amodal representations after bimodal learning: Integration of haptic-visual layouts into a common spatial image. *Spatial Cognition and Computation, 9,* 287–304.

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

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature, 363,* 623–625.

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,* 48–61.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hillsdale: Erlbaum.

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology. Human Perception and Performance, 18,* 906–921.

Loomis, J. M., Klatzky, R. L., & Giudice, N. A. (in press). Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language. In S. Lacey & R. Lawson (Eds.), *Multisensory imagery: Theory and applications*. 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,* 966–980.

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

McCarthy, G., Blamire, A. M., Puce, A., Nobre, A. N., Bloch, G., Hyder, F., Goldman-Raikic, P. & Shulman, R. G. (1994). Functional magnetic resonance imaging of human prefrontal cortex activation during a spatial working memory task. *Proceedings of the National Academy of Sciences, 91,* 8690–8694.

Mills, A. W. (1972). Auditory localization. In V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2, pp. 301–342). New York: Academic Press.

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 142–157.

Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature, 414,* 197–200.

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

Philbeck, J. W. & Loomis, J. M. (1997) Comparison of two indicators of visually perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 72–85.

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1157–1165.

Rieser, J. J., Garing, A. E., & Young, M. F. (1994). Imagery, action, and your children's spatial orientation: It's not being there that counts, it's what one has in mind. *Child Development, 65,* 1262–1278.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M.-P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature, 380,* 526–528.

Sathian, K., & Lacey, S. (2007). Journeying beyond classical somatosensory cortex. *Canadian Journal of Experimental Psychology, 61,* 254–264.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology. General, 125,* 4–27.

Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9,* 255–266.

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 867–882.

Wang, R. F. (2004). Between reality and imagination: When is spatial updating automatic? *Perception & Psychophysics, 66,* 68–76.