Exp Brain Res (2003) 149:48–61
DOI 10.1007/s00221-002-1334-z

RESEARCH ARTICLE

Roberta L. Klatzky · Yvonne Lippa · Jack M. Loomis · Reginald G. Golledge

Encoding, learning, and spatial updating of multiple object locations specified by 3-D sound, spatial language, and vision

Received: 19 June 2002 / Accepted: 30 October 2002 / Published online: 19 December 2002

Abstract Participants standing at an origin learned the distance and azimuth of target objects that were specified by 3-D sound, spatial language, or vision. We tested whether the ensuing target representations functioned equivalently across modalities for purposes of spatial updating. In experiment 1, participants localized targets by pointing to each and verbalizing its distance, both directly from the origin and at an indirect waypoint. In experiment 2, participants localized targets by walking to each directly from the origin and via an indirect waypoint. Spatial updating bias was estimated by the spatial-coordinate difference between indirect and direct localization; noise from updating was estimated by the difference in variability of localization. Learning rate and noise favored vision over the two auditory modalities. For all modalities, bias during updating tended to move targets forward, comparably so for three and five targets and for forward and rightward indirect-walking directions. Spatial language produced additional updating bias and noise from updating. Although spatial representations formed from language afford updating, they do not function entirely equivalently to those from intrinsically spatial modalities.

Keywords Spatial updating · Memory · Language · Modality effects · Sound

R. L. Klatzky (✉)
Department of Psychology,
Carnegie-Mellon University, Pittsburgh, PA 15213 USA
e-mail: klatzky@cmu.edu
Tel.: +1-412-2683151
Fax: +1-412-2683464

Y. Lippa · J. M. Loomis
Department of Psychology, University of California,
Santa Barbara, CA 93106 USA

R. G. Golledge
Department of Geography, University of California,
Santa Barbara, CA 93106 USA

Introduction

The present research was concerned with the ability of individuals to encode and learn the layout of multiple locations in space relative to a common origin and then to spatially update their position relative to the locations while walking. Spatial updating refers to people's ability to change the representation of themselves relative to their environment, as they move. Our principal interest was in comparing encoding, learning, and updating performance across different sensory modalities by which the locations were initially presented. The modalities were two types of auditory input, descriptive spatial language versus sounds that originate from the target locations, and vision.

Numerous perceptual studies have demonstrated nonvisual updating ability, most using visual exposure (see, for example, Amorim et al. 1997; Böök and Gärling 1981; Farrell and Thomson 1999; Loomis et al. 1992, 1998; Ooi et al. 2001; Philbeck et al. 1997, 2001; Presson and Montello 1994; Rieser 1989; Rieser and Rider 1991; Thomson 1983). In one type of updating task, the participant views a target location, and then is signaled to walk to it directly with eyes closed. Performance in this task is highly accurate out to 20 m or more. Moreover, direct and indirect walking converge on the same location. These findings are taken to indicate that people: (a) perceive the target accurately and (b) execute the walk to it with negligible bias. When the same tasks are applied to perceived auditory stimuli (Loomis et al. 1998), both indirect and direct walking arrive at a common location, but one that is not at the true target location. The error occurs because the distances to far sounds are perceived as nearer than their actual locations, and updating is performed with respect to the perceptually based representation.

Encoding and updating a single object location from 3-D sound or spatial language

Loomis and colleagues (2002) conducted a task, in which participants were exposed to a single target location in space, by hearing either a 3-D sound that emanated from the location or language describing its location (for example "2 o'clock, 6 feet"). They then walked to the target either directly or indirectly (i.e., by walking forward until they were instructed to turn and walk to it). The authors proposed an "image-updating" model of the task that separated two critical processing phases, encoding and updating. During the encoding phase, a representation of the spatial location, called a spatial image, is assumed to be formed. During the updating phase, the representation is modified to accommodate the changes in position of the participant relative to the target. Either of these phases could produce systematic error (bias) or random error (noise).

The model further hypothesized that different modalities lead to differences in the process of encoding a spatial location, but that the outcome of encoding, the spatial image, could function equivalently across modalities for purposes of updating. This hypothesis does not require that the spatial images derived from different modalities have a common format, either amodal or by conversion of all inputs to a single, modality-specific code. However, for spatial updating to function effectively, the image formed from an input modality must afford processes that compute spatial location during locomotion. The image-updating model was tested by devising measures of encoding bias, updating bias, and total noise, as described below. As hypothesized, encoding bias was found to vary with modality, whereas updating bias did not. Participants evidenced an ability to effectively update target locations that were initially presented via 3-D sound or spatial language.

These results were particularly important because they indicate that language can be converted to a spatial image that is functionally equivalent to one formed from a process of spatial perception. Previous research has shown that people can construct a representation of the layouts of objects from spatial language that allows for inferences about spatial relations (see, for example, Bachmann and Perrig 1988; Bosco et al. 1996; Ferguson and Hegarty 1994; Franklin and Tversky 1990; Perrig and Kintsch 1985; Taylor and Tversky 1992; Wilson et al. 1999). However, such inferences may be based on non-spatial cognitive abstractions such as verbal mediators. Successful updating, in contrast, indicates a spatial form of representation.

Encoding, learning, and updating multiple object locations

Next consider a corresponding situation in which participants learn multiple target locations over a series of trials, rather than responding to a single target immediately after its presentation. Again, the task has multiple components. As in the single-target task, there is an *encoding* of each target as it is presented, resulting in a spatial image. Because there are several targets, however, a stable memory representation must be formed that will retain all of the spatial images. We will refer to the formation and retention of the collective layout as the *learning* phase of the task, which is potentially subject to its own form of bias and noise and may vary with modality. Differences among modalities with respect to the rate of learning a set of azimuths were found by Klatzky et al. (2002). Specifically, participants were slower to learn the azimuths associated with five objects when they were presented by spatial language, in comparison to vision or 3-D sound. Once learning has occurred, when a participant moves within the layout, *updating* occurs as in the single-target case, but now with respect to multiple targets.

Measures of bias and noise

The present studies used the multiple-target task. Participants learned a set of target locations over a series of exposures and then were tested on updating. We derived measures of bias and noise in the various phases of performance. These were based on our previous study (Loomis et al. 2002), but with some differences noted below:

A. *Encoding bias* was measured by comparing the centroid of reported locations to the actual target location. There were two types of reports in the present studies. After a series of learning trials, subjects reported the encoded locations by verbally estimating distance and pointing with a pointer. In experiment 2, encoding was also measured by the participants' stopping point when they walked a direct path to the target from the origin.[1]

B. A new measure, trials required to reach a localization criterion, was introduced to measure *learning* rate.

C. In the updating phase, participants moved from the origin to an indirect waypoint, from which they localized the targets. In experiment 1, they verbally estimated distance and pointed to the targets from the indirect waypoint, as they had after learning. In experiment 2, the waypoint was the turning point on an indirect walking path to the targets. *Updating*

[1] In contrast, Loomis et al. estimated the encoded locations by averaging the centroids of direct and indirect walking, under the assumption that people converge on a common perceived location whether by a direct or indirect route. We evaluated the direct path alone, rather than averaging it with the indirect, because the targets were learned over multiple exposures rather than having just been perceived. While there is substantial evidence that immediately after perceiving a target, people walk without vision to a common spatial image along direct and indirect paths, updating an image that is encoded into memory over time may lead to additional bias when people walk an indirect path.

*bias*was measured by the difference between the centroids of the target locations as reported by means of the indirect waypoint and those either reported after learning (experiment 1) or reached by the direct path (experiment 2). In both experiments, the updating bias measure indicates the additional systematic error resulting from a move along a path through a point that is not collinear with the origin and target location.

D. *Noise* in the response, collectively resulting from imprecision in encoding, learning, and updating, was assessed from interparticipant variability. This measure used the average distance of each participant's response from the group centroid.

A final phase in this task is *response execution*, arising from the act of reporting the target locations. When the response is walking, then updating and response execution are commingled. The studies reported above, which found highly accurate walking without vision to a viewed target, argue that walking per se produces negligible bias. When the response is verbal description or pointing, in contrast, systematic error could be produced at the response phase due to the metric imprecision of language and lack of experience with verbal description. Such response-induced bias cannot be discriminated from bias in the internal representation of the target locations due to encoding or learning. It should, however, be independent of the input modality. Response execution could produce noise as well as systematic error.

Functional equivalence across modalities

As was noted above, the Loomis et al. (2002) data supported the idea that updating, but not encoding, functions equivalently for single-target representations formed from 3-D sound and spatial language. With respect to encoding bias, an expected disadvantage in distance perception for 3-D sound was obtained. Participants consistently under-walked toward the further targets when they were specified by 3-D sound, but both modalities showed highly accurate azimuth encoding. Updating performance with 3-D sound and spatial language was remarkably similar. The measure of updating bias (difference between stopping point for direct and indirect paths) for sighted subjects showed only a small average difference in azimuth (6–10º for the two modalities), and blind participants exhibited no effect of modality on updating bias. Finally, when the noise measure was adjusted for differences in the distance walked, sighted participants showed no modality effect, although blind participants demonstrated slightly greater noise with 3-D sound.

If the same model applies to memory-based responses and multiple targets, and if it extends to vision as well as 3-D sound and spatial language, all three modalities should show equivalent updating bias, or systematic error due to responding via an indirect waypoint. The model allows other perceptual and memory-related processes to be modality-specific. This means that the learning time is likely to differ, and there may be differential effects of memory load (i.e., the number of locations that are learned). Moreover, the modalities are expected to show differential encoding bias, particularly the previously observed compression of distances when locations are presented as 3-D sounds.

Finally, differential precision across modalities could lead to different levels of unsystematic error or noise. Noise accumulates from all processes: encoding, learning, updating, and responding. An advantage for some modality, such as vision, with respect to the spatial resolution of the targets could produce an advantage with respect to noise in encoding and learning. This outcome would not be inconsistent with the idea that multiple input modalities lead to representations that function equivalently for spatial updating. Therefore, the observation of modality effects on the cumulative noise measure would not by itself disconfirm the model. However, if differential precision in encoding led to differential noise during updating, this would violate a strict interpretation of functional equivalence across modalities in the updating phase. The effect of updating on noise can be evaluated by comparing the noise measure when people respond directly from the position of learning versus via an indirect waypoint.

Experiment 1

This study compared 3-D sound (3-D) and spatial language (SL) with respect to spatial learning and updating of three or five targets. Those set sizes were chosen because they had previously shown a difference in the task of associating object names with spatial azimuths (Klatzky et al. 2002). After learning the locations of a set of named objects to a criterion level of accuracy for both distance and azimuth, participants walked forward to an indirect waypoint and indicated the new values of distance and azimuth for each object. The distances were estimated verbally, and the azimuthal judgments were made with a pointing device.

Materials and methods

This research was approved by the local ethics committee.

Participants

The participants were undergraduate students from the University of California, Santa Barbara, who received credit to fulfill a course requirement. In the three-target condition, there were ten males and six females, and the mean age was 19.1 years (range 18–23 years). In the five-target condition, there were six males and ten females, and the mean age was 18.6 years (range 17–21 years). Three additional participants in each target condition had difficulty localizing the sounds and were not tested further.

Experimental setup, stimuli, and apparatus

The experiment took place in a room with a workspace measuring 5.4 m in width and 7 m in length. In the five-target condition, there were two sets of five target locations that were arranged in a semicircle around an origin. The polar coordinates of the first set were: −90º/2.74 m, −60º/0.91 m, −30º/4.57 m, 30º/3.66 m, 60º/1.83 m. The second set was a mirror image of the first, i.e., the angular coordinate was opposite in sign. The locations in each set were assigned object labels from one of two sets: (1) *baby*, *bell*, *cat*, *horse*, *plane* or (2) *bird*, *car*, *dog*, *drum*, *phone*. The object names are listed with a word frequency of 80, 23, 42, 203, 138, 83, 393, 147, 26, and 53, respectively, in the norms of Francis and Kucera (1982). Across participants, we counterbalanced: (a) the assignment between the location sets and object label sets and (b) whether an assignment was employed in the 3-D or SL condition. The location of the indirect waypoint, to which participants walked in the spatial updating phase, was 2.5 m forward of the target. The three-target condition used a subset of the stimuli. Set 1 had locations at −90º/2.74 m, 30º/4.57 m, and 60º/0.91 m, and set 2 had locations with the reverse-signed angle.

Stimulus presentation and data collection were controlled by a Gateway E-5200 computer, using a scripting facility in the Python programming language, supplemented with a utility module written for virtual environment applications (Beall et al. 2001). To synthesize the stimuli, we used the speech synthesizer provided by AT&T (http://www.att.com/aspg/).

In the 3-D condition, the speech stimuli spoken by a male voice emanated from loudspeakers (Propart 3-inch mini speaker) that were mounted facing the participants on microphone stands at a height of 1.70 m. The duration of each object label was within an epoch of 500 ms. The intensity for the closest stimulus (0.91 m) was approximately 64 dB and for the farthest stimulus (4.57 m) approximately 58 dB (all A weighting).

In the SL condition, the speech stimuli were presented through a loudspeaker (OPTIMUS multimedia speaker 40–1401) that was mounted on a height-adjustable platform at a distance of 41.5 cm in front of the participant. A male voice specified the stimulus location and name with an intensity of approximately 63 dB. Observers first heard the digit that indicated the clock position, then the distance in feet, and then the object label spoken twice with a 500-ms interstimulus interval (for example, "10 o'clock, 6 feet, dog–dog"). The presentation time for one location was approximately 4 s.

During the spatial updating test, participants were cued by a male voice uttering the object label once, emitted from the same loudspeaker used in the SL condition. Azimuthal judgments were collected by means of a custom-fabricated pointing device that was mounted on the same platform that held the speaker for the SL condition. It consisted of a 13.5-cm-long pointer that, at its center, was mounted on a plastic box (5 cm high x 6.5 cm wide x 9 cm long) and attached to a potentiometer that was interfaced with the computer. The participant grasped the pointer and rotated it about the vertical axis. The pointer had a range from 175º through 175º. The participant pressed a button on the box to record the response angle.

Procedure

The participants' task had two phases. First was a multi-trial learning phase, in which they learned the locations of the objects. Second were two spatial-updating test trials, separated by an additional learning trial. Participants were instructed outside of the experimental room about the learning and spatial updating phases of the experiment, with the latter being demonstrated and practiced using a visual example. Thereafter, they were blindfolded and guided into the room to the origin location. The participants then completed several practice trials with the pointing device. It was particularly emphasized that they should point in relation to the device and not in relation to their body, which was approximately 10 cm distant. Participants completed the 3-D and SL condition one after the other, with the order of condition being counterbalanced.

Learning. Each learning trial consisted of exposure to the target locations in the set, followed by a test. During exposure, each target location occurred once (with the name repeated twice), with a stimulus-onset asynchrony of approximately 3 s. During the test, each target location was cued in turn by its label, and the participant provided first an azimuthal judgment by pointing and then a distance estimate in feet by verbal report. Order of object names during exposure and test was random. The exposures and tests alternated until the pointing error in a series of three trials, averaged across the objects, was less than 15º, and the true and indicated distances for the objects showed a rank correlation of 0.75 or higher. The rank correlation criterion required accurate relative distance judgments, and thus compensated for the fact that absolute distance judgments are expected to show errors in the 3-D condition. Feedback in the initial learning trials was limited to indicating whether the participant had met the learning criterion. If a participant had difficulty after several trials, the experimenter indicated if distance or azimuth was the problem.

Spatial updating. After the learning criterion was reached, the platform was moved and the participant was guided forward to the indirect waypoint. Immediately after, the target locations were tested as in the learning trials. To obtain another replication of spatial updating, the participant was guided back to the origin (by walking backwards), where he or she completed another single learning trial (exposure and test of the five target locations), and then was guided forward again to complete a second test at the indirect waypoint.

Results

The effects of interest were not modulated by the variables location set or order of modality. The data were therefore collapsed across these factors.

Learning

An analysis of variance (ANOVA) was conducted on the number of trials needed to achieve the learning criterion, using the within-participant variable of modality (3-D versus SL) and the between-participant variable of memory load (5 or 3). The effect of memory load did not reach significance, although there was a trend ($P<0.10$) toward more trials to learn five items (7.3 trials) than to learn three items (5.8 trials). The effect of modality also did not reach significance.

Encoding bias

Encoding bias was measured in this study by psychophysical functions relating the distance and direction of the targets, as reported from the origin, to their correct values, using the five-target condition. The functions for distance and direction in both modalities were all fit by a linear trend with $r^2 \geq 0.98$. The slopes of the *direction* functions were 1.12 for 3-D and 0.92 for SL, a significant difference, $t(30)=9.14$, $P<0.001$; the intercepts were –4.3° and –1.8°, respectively, and did not differ statistically. The *distance* function for the SL condition was irrelevant, because participants simply repeated back the distances they had been told and therefore achieved nearly perfect performance. The distance function for 3-D had a slope of 0.80 and an intercept of –0.02 m. The 95% confidence interval around the slope did not include 1.0, indicating the expected compression of distance with 3-D sound. However, the magnitude of compression was less than in our outdoor studies (typically ca 0.5; see Loomis et al. 1998), presumably because of the more reverberant indoor environment.

Spatial updating bias

A second analysis addressed whether spatial updating bias was different for the two presentation modalities. Figures 1 and 2 depict the group centroids separately for three and five targets as a function of response location (origin or indirect waypoint), modality of presentation, target location, and location set.

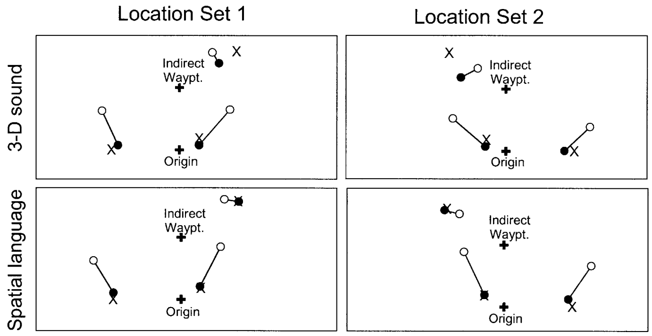

Fig. 1 Group centroids for each response location in experiment 1, with three targets, by stimulus modality (3-D sound versus SL) and location set. *Closed circles* are centroids of the locations indicated from the origin, *open circles* are centroids of locations indicated from the indirect waypoint, *crosses* are the targets. Centroids to the same target are connected by a *line*

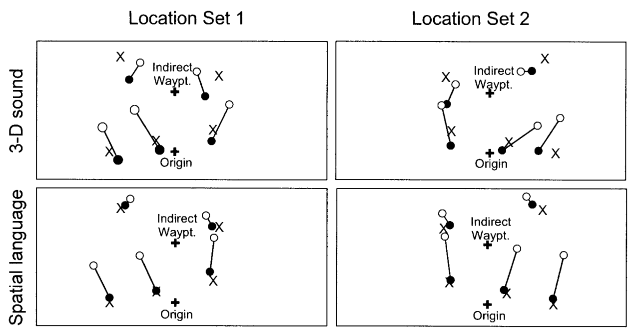

Fig. 2 Group centroids for each response location in experiment 1, with five targets, by stimulus modality and location set. Legend as in Fig. 1

As a measure of updating bias, we used the difference between the locations indicated from the origin and from the indirect waypoint, as specified by polar coordinates (vector differences). The polar coordinates from the origin were the azimuth and distance estimates provided by the participants, averaged over the last trial of the learning phase and the replication learning trial that occurred after the first spatial-updating test. For the indirect waypoint, we converted the distance and direction estimates during the spatial-updating test to polar coordinates relative to the origin, averaging the two updating tests. As there were symmetric tendencies in error relative to the sagittal axis, we signed azimuthal errors so that a negative value represents a tendency to respond closer to the midline than the correct value, regardless of whether the target is on the left or right (as in Loomis et al. 2002). Errors in these experiments and that of Loomis et al. are shown in Table 1. As can be seen in Figs. 1 and 2 and Table 1, locations indicated from the indirect waypoint tended to be displaced forward relative to the locations indicated from the origin.

The vector differences between the polar coordinates (distance; azimuth) obtained for the two response locations were submitted to a multivariate analysis of variance

Table 1 Updating bias: the mean vector difference (distance; azimuth) between the update condition and the baseline condition. Updating bias in experiment 1 is the difference between estimates made from the indirect waypoint and the origin, and in experiment 2 it is the difference between estimates made by indirect and direct walking. Data are shown by experiment and modality, along with comparable values for Loomis et al. (2002). (*SL* specifies Spatial language, *VIS* specifies visual target)

		3-D sound	SL	VIS
Distance (m)	Experiment 1: three items	0.99	0.89	
	Experiment 1: five items	0.99	0.93	
	Experiment 2: forward indirect	0.31	0.44	0.27
	Experiment 2: rightward indirect	0.20	0.37	0.20
	Loomis et al.	0.11	0.10	
Azimuth (degrees)	Experiment 1: three items	−16.8	−15.3	
	Experiment 1: five items	−21.4	−14.7	
	Experiment 2: forward indirect	−15.6	−16.2	−4.6
	Experiment 2: rightward indirect	−10.8	−17.1	−13.2
	Loomis et al.	9.9	−6.4	

(MANOVA), with target location (defined by distance from the origin, i.e., the mirror-image locations in the two sets were averaged) and modality as within-participant variables. In an initial MANOVA, memory load (5 versus 3) was also a factor; this used only the three target locations common to both experiments. The sole effect was target location, $F(4,27)=53.25$, $P<0.001$, Wilks' lambda criterion (for modality and memory load, Fs<1). Univariate tests indicated the target effect was significant for direction and distance, Ps<0.001, reflecting the fact that the displacement was larger for targets that were closer to the origin than for targets that were closer to the indirect waypoint. An additional MANOVA incorporating all five targets, with factors of modality and target, confirmed the target effect, $F(8,118)=16.90$, $P<0.001$, which was significant for both direction and distance (Ps<0.001); again modality was not significant.

To confirm that the updating bias was primarily in the forward direction, we performed additional, separate analyses on the x (rightward) and y (forward) components of the indicated locations. The analyzed variable was the difference between the coordinate as indicated from the two response locations, averaging over target locations. The analysis of the y-coordinate showed that the mean indicated location was further forward (on average, by 0.94 m) when reported from the indirect waypoint than from the origin; the 95% confidence interval around the mean y–coordinate difference excluded zero within each modality and number of targets. The 95% confidence interval around the mean x-coordinate difference included zero for each condition. For both x and y differences, an ANOVA on modality and number of targets showed no significant effects. Thus the updating bias lay primarily forward, regardless of memory load or modality.

Noise analysis

As in Loomis et al. (2002), we used the between-participant variability (i.e., the mean distance between the centroid of each participant's indicated locations and the group centroids, averaging the two replications) as a measure of noise. As shown in Fig. 3, the variability measure was regressed on the distance from the origin to the group centroid of the indicated locations and on number of target items, modality, and response location (origin or indirect waypoint). There were significant effects of distance, response location, and modality, (Ps<0.05), but not number of targets. Variability increased with increasing distance (β=0.55), was greater for responses from the indirect waypoint than from the origin, (β=0.40), and was greater for the 3-D sound than the SL condition (β=0.26). Note that the effect of modality is relatively small and is in the reverse direction of the small

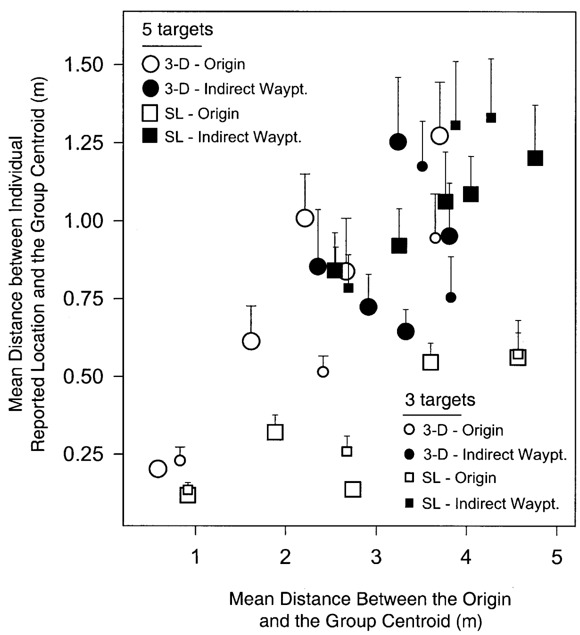

Fig. 3 Noise in experiment 1, as measured by the distance between the individual reported location and the group centroid, as a function of distance from origin to group centroid, by number of targets, response location (origin versus indirect waypoint), and stimulus modality

effect found for blind participants by Loomis et al. (2002).

Further analyses were performed to determine whether the additional noise due to updating varied by modality. Within each modality, the noise measure was regressed on a single predictor, response location (origin or indirect waypoint), with the observations representing each combination of target location and memory load. The effect was significant for SL (β=0.89, P<0.01) but not for 3-D sound (β=0.32), indicating that updating on the indirect path added noise when targets were learned through spatial language.

Discussion

The present results indicate that when multiple targets are presented through spatial language, locomotion introduces no more systematic error in spatial updating than if the targets were exposed through 3-D sound. Updating bias was in the forward direction and was virtually identical across modality and memory load. However, functional equivalence across modalities was not complete: updating produced additional noise (between-subject variability) relative to reports from the origin when targets were learned from spatial language, but not from 3-D sound. This effect contradicts the image-updating model.

As did Loomis et al. (2002), we found a modality effect in encoding, in that people underestimated distances to targets specified by 3-D sound. Modality did not affect learning, in contrast to the deficit for azimuth learning via language found by Klatzky et al. (2002); however, spatial language may have had a compensatory advantage here in that distance was reported verbally. A larger memory load tended to produce longer learning but not greater noise.

Experiment 2

The principal purpose for experiment 2 was to further evaluate the image-updating model in the context of multiple targets, but using an action-based measure of updating and with vision as an additional modality. Walking to a target is arguably the best indication of its subjective location. In contrast, pointing with a pointer held in the hand can add to systematic error (Haber et al. 1993; Montello et al. 1999), and verbal distance estimation has been found less accurate than a walking response (Loomis et al. 1998). Accordingly, experiment 2 used the same measures of target localization that had been used by Loomis et al. (2002). The combination of direct and indirect walking allowed use of the metric developed by Loomis et al. to evaluate whether there is a systematic difference among modalities with respect to updating bias, namely, the difference between the direct and indirect centroids. It also allowed the effect of updating on noise to be evaluated by comparing variability under the two walking conditions. Finally, the level of error with multiple targets could be directly compared to that of Loomis et al. with single targets, within a common response.

The nature of the response may explain why, relative to the data of Loomis et al. (2002; see Table 1), experiment 1 led to a considerable increase in systematic error or bias after updating. That is, experiment 1 showed a substantial discrepancy between the indicated locations at the end of initial learning and after walking to the indirect waypoint. However, it is not obvious from the present analysis why the nature of the response should affect bias that accrues while walking to the response location. That is, our model does not explain a priori how differences in the response could lead to differences in updating bias. One possibility is that updating is influenced by participants' expectancy about the nature of the ultimate response; in particular, by whether it is action-based (as in Loomis et al. 2002) or a stationary report (as in the present experiment 1). Loomis et al. (1998) suggested, for example, that when anticipating a verbal response, people may direct less attention to updating.

The pattern of spatial updating bias observed in experiment 1 was characterized by people misreporting targets from the indirect waypoint as further forward than they actually were. A possible locus for the bias is that participants miscalibrated their self-motion during updating, by underestimating their walking speed along the indirect path (although this would contradict the evidence cited above that there is little bias in walking per se). For example, a target location that had actually passed behind the participant while walking might be perceived as still ahead, so that the subject would give an erroneous angle estimate that pointed forward. To test this hypothesis, in experiment 2 participants moved either forward or rightward to the indirect waypoint. If the updating bias reflects underestimation of velocity during updating, targets approached from an indirect waypoint in the rightward direction should be displaced to the right rather than forward.

To provide a baseline of optimal updating performance for general comparison with the auditory conditions, vision was added as a third modality. We used the perceived distance in the 3-D sound condition as derived from the distance judgments in previous experiments to select matching values to use in the spatial language and vision condition. To our knowledge, the current experiment is the first to compare updating after visual preview to updating after auditory exposure to spatial targets, when differences in the encoded locations are controlled for. Thus, to the extent that updating bias and noise depend exclusively on perceived location, performance should be unaffected by the differential encoding bias across modalities. However, vision could still have greater encoding precision, that is, the variability in the perceived location of the target could be less after visual exposure, which could also affect the precision of subsequent processes. These collective effects should then emerge as a modality effect on our measure of noise.

There could also be an advantage for vision in learning due to the simultaneous exposure of targets.

Materials and methods

Participants

Seventy-two undergraduate students at the University of California, Santa Barbara, served as participants, either to receive credit to fulfill a course requirement or for pay. There were 28 males and 44 females, and the mean age was 19.7 years (range from 18 to 38 years; SD=2.6 years). Two additional participants were eliminated because of difficulty in sound localization, and data from two others were eliminated because they confused object-designated locations in the spatial updating phase. Direction of indirect walking and modality were between-participant variables.

Apparatus and stimuli

The experimental space and data collection in the learning phase were the same as in experiment 1. Position data in the spatial updating phase were collected using a tracking system. Two cameras tracked an LED, which was mounted on the top of the plastic headband worn by the participant. The camera outputs were digitized, and the computer calculated the coordinates of the LED with an accuracy of about 1 cm.

Because the memory load effect had been non-significant in experiment 1, we used only the larger set size of five items. The target locations were indicated by 3-D sound (3-D) or spatial language (SL) or were specified visually (VIS), with modality being a between-participant variable. Only one object label set was used: *baby*, *bell*, *cat*, *horse*, *plane*. There was one location set, matched for perceived distance across modalities. In the 3-D condition, the locations were at −90°/2.74, −60°/1.22 m, −30°/3.66 m, 30°/4.57 m, and 60°/1.83 m. To match distances in the 3-D condition to the corresponding distances in the SL and VIS conditions (to the closest foot), we used a psychophysical function for perceived 3-D sound that was derived from averaging responses in two previous experiments (slope=0.763; intercept=0.04 m). In the SL and VIS conditions, the nominal locations were therefore at −90°/2.13 m, −60°/0.91 m, −30°/2.74 m, 30°/3.66 m, and 60°/1.52 m. In the VIS condition, the first location actually was at −90°/1.91 m, because a light fixture interfered with placement of the stimulus. There were 12 random assignments between the object labels and locations, with participants in each condition being yoked. The turn location that was used in the forward indirect walking condition was located at 0°, 2.5 m, and the turn location used in the rightward indirect walking condition was located at 2.5 m, 90°.

In the 3-D condition, the speech stimuli emanated from loudspeakers (Radio Shack, 3.5 inch, full-range speakers) that were suspended by rods from the ceiling. The center of a loudspeaker was 1.9 m above the floor and faced the participants. In the VIS condition, the loudspeakers were covered up with sheets of cardboard (25.5×12.5 cm), on which the object labels were printed in black on a white background. In the SL condition, the speech stimuli were presented as in previous experiments.

Procedure

Participants were instructed outside of the experimental room and wore a blindfold during the entire experiment. They were informed about the learning part as well as the walking part of the experiment. The difference between direct and indirect walking was demonstrated and practiced using a visual example.

Learning. The learning procedure replicated that of experiment 1, except that the criterion in the learning phase had to be reached six times. In the VIS condition, participants lifted the blindfold for a presentation time of 23 s (matched to the total presentation time for all five locations in the 3-D sound condition) and were allowed to look freely around. There were no additional learning trials once the spatial updating phase began.

Spatial updating. In the spatial updating phase, one experimenter initiated the cue presentations, while another one monitored the participant's safety. Participants first completed a block of indirect walking trials and then a block of direct walking trials. We chose this blocked order of walking trials to prevent participants from learning about the locations by directly walking to them. Each block consisted of ten trials, two replications for each of the five targets. In an indirect walking trial, the experimenter guided the participant directly to the turn point; the subject walked forward or sidestepped rightward without change of heading. Then the cue was presented and participants walked the rest of the way to the target location, which was recorded by the experimenter. Then participants were led back to the origin while facing forward, and a new trial started. A direct walking trial followed the same procedure, except that the participant walked directly to the stimulus location from the origin.

Results

Learning

An ANOVA was performed on the number of trials needed to achieve the learning criterion, using the between-participant variables of modality and indirect walking direction. There was a significant effect of

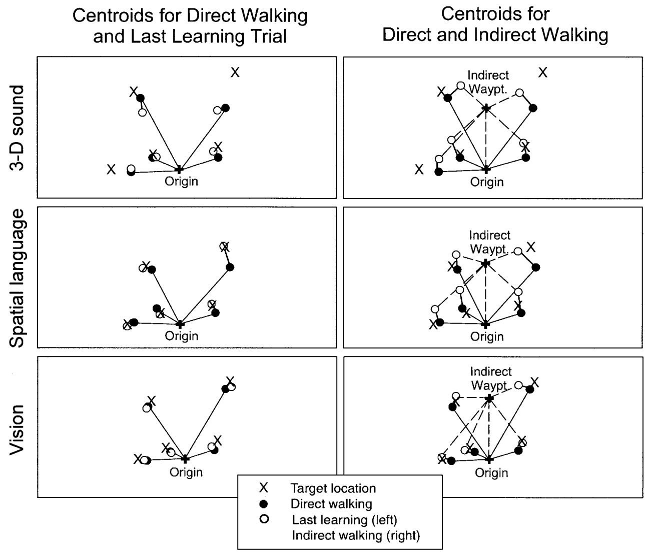

Fig. 4 Biases for forward indirect-walking condition in experiment 2, by modality (*rows*). *Left panels* Difference between the two measures of encoding bias, as indicated by group centroids for the last learning trial (*open circle*) and direct walking (*dark circle*). Targets are at *crosses*. *Solid lines* connect centroids corresponding to the same target and also indicate the direct path to the target from the origin. *Right panels* Updating bias as indicated by group centroids for direct (*dark circle*) and indirect (*open circle*) walking to each target location (*crosses*). *Solid lines* are as in left panel, *dashed lines* indicate the path from the indirect stopping point

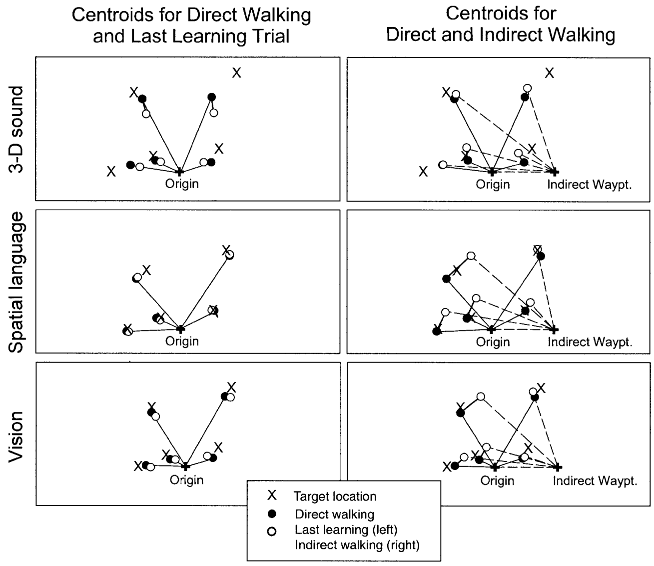

Fig. 5 Biases for rightward indirect-walking condition in experiment 2, by modality (*rows*). Legend as in Fig. 4

modality, $F(2,66)=9.04$, MSE=3.10, $P<0.10$. The VIS group needed fewer trials to learn the object array (6.9 trials) than the 3-D group (9.0 trials, Scheffé test, $P<0.01$) and fewer than the SL group (8.4 trials, Scheffé test, $P=0.05$). The 3-D and SL conditions did not differ from each other.

Encoding bias

There were two measures of encoding bias, the average reported values at the last learning trial and the centroid of direct walking. The reported locations from these measures are compared in Figs. 4 and 5 *left panels*. The psychophysical functions relating reported distance and direction to the correct values were fit well with linear trends for both measures, as summarized in Table 2.

Considering first the *direction* functions, both measures of encoding bias indicated that direction encoding was highly accurate. The slopes were near 1.0 and intercepts near 0 (see Table 2), and both slopes and intercepts were statistically equivalent across modalities by each measure.

The measures also converged with respect to the *distance* functions, in that distance encoding in the 3-D condition showed compression of the longer distances. This distance encoding bias was anticipated; it was in fact the reason for adjusting the actual distance values for SL and VIS to match perceived locations in the 3-D condition (see Materials and methods). With the last learning trial as the measure of bias, the distance responses for the SL condition were not considered, because participants simply reported back the verbal distance they had been told. When the distance function from the last learning trial for 3-D was compared to VIS, the modalities differed in slopes and intercepts, $t(46)=6.10$ and 5.49, $Ps<0.001$. With the direct-walking measure of encoding bias, an ANOVA on the slopes of the distance functions across all three modalities showed a significant effect, $F(2,69)=7.95$, MSE=0.06, $P<0.01$; *post hoc* Scheffé tests indicated that VIS had a higher slope than 3-D ($P<0.01$). A

Table 2 Slopes, intercepts, and goodness of fit of linear functions relating reported distance and direction to correct values in experiment 2, by reported parameter, modality, and measure–learning versus direct walking

	Direction (degrees)			Distance (m)		
	3-D	SL	VIS	3-D	SL	VIS
Slope: last learning trial	1.02	1.02	1.04	0.60	1.00[a]	1.07
Slope: direct walking	1.10	1.07	1.09	0.72	0.84	1.00
Intercept: last learning trial	2.13	–2.56	0.84	0.23	0.00[a]	–0.51
Intercept: direct walking	3.44	2.92	2.22	0.29	0.33	–0.28
r^2: last learning trial	1.00	1.00	1.00	0.96	1.00[a]	0.99
r^2: direct walking	1.00	1.00	0.99	0.90	0.97	0.96

[a] Participants achieved perfect performance in the verbal response, given verbally designated targets

corresponding ANOVA on the distance intercepts was also significant $F(2,69)=13.12$, MSE=0.21, $P<0.001$; VIS differed from both 3-D and SL (Scheffé test, Ps<0.001).

Spatial updating bias

The updating biases were generally higher than those observed for single targets by Loomis et al. (2002; Table 1).[2] A MANOVA on modality, indirect-walking direction, and target location was applied to the difference in polar coordinates between the locations designated by direct and indirect walking, i.e., updating bias, shown in Figs. 4 and 5 (*right panels*). No effect involving modality reached significance. The bias varied with target location, $F(8,59)=10.92$, $P<0.001$, which interacted with indirect-walking direction, $F(8,59)=3.51$, $P<0.01$. Univariate analyses indicated these effects were significant for both distance and angle, Ps<0.05.

To further investigate whether the updating biases lay along the direction of indirect walking, we performed additional analyses comparing the indirect and direct stopping points with respect to the *x*- (rightward) and *y*- (forward) coordinates. The analysis of the *y*-coordinate difference showed that the mean stopping point was further forward for the indirect path than the direct path; the 95% confidence interval on the mean difference excluded zero within each modality for each walking direction. An ANOVA by modality and direction on the *y*-coordinate difference showed that the magnitudes of forward updating bias for forward indirect paths (mean=0.50 m) and rightward indirect paths (mean= 0.42 m) did not differ significantly, $F<1$, but there was a modality effect, $F(2,66)=4.46$, $P<0.05$. The forward bias averaged 0.66, 0.44, and 0.29 m for SL, 3-D, and VIS, respectively, and the SL and VIS means differed by Scheffé test. The interaction was not significant.

The analysis of the *x*-coordinate difference showed that there was an overall trend for the indirect stopping point to be rightward of the direct stopping point after rightward walking (by 0.21 m on average, $P<0.05$), but not after forward walking (which showed an average leftward tendency, –0.17 m, $P>0.05$). An ANOVA on the *x*-coordinate difference by modality and indirect-walking direction showed a significant effect of direction, $F(2,66)=16.34$, $P<0.001$. Although the modality × direction effect was marginal, $P=0.07$, an analysis by modality showed that only the SL condition manifested significant rightward updating bias after rightward indirect walking (0.35 m); the mean direction of bias was actually reversed in the 3-D sound condition. None of the pairwise modality comparisons reached significance by Scheffé test.

[2] Latencies for the walking response (i.e., the time from onset of the cue to the onset of walking) were also recorded. The only significant effect was that of walking condition. The time to initiate walking from the origin directly to a target was shorter (5.2 s) than the time to start walking from the turn point, in the indirect condition, to the target (7.3 s). The average time was 5.8, 6.0, and 7.1 s for VIS, 3-D, and SL, but differences were not significant.

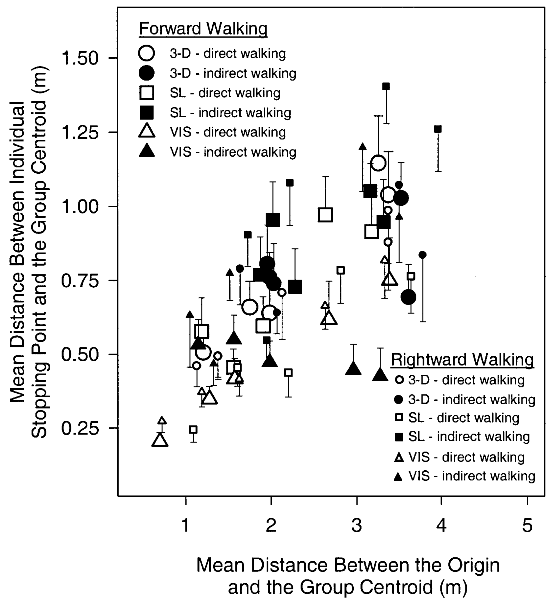

Fig. 6 Noise in experiment 2, as measured by the distance between individual stopping points and the group centroid, as a function of distance from origin to group centroid, by direct and indirect walking, indirect-walking direction, and stimulus modality

Noise analysis

Again between-participant variability was the measure of noise, now using the distance between the individual's stopping point and the group centroid (see Fig. 6). The previous response-location variable (origin versus indirect waypoint) was replaced with a path-type variable (direct versus indirect walking) as a predictor in the regression. The other predictors were distance from the origin to the group centroid of stopping points, indirect-walking direction, and modality. Three such regressions were run, one for each pairing of modalities. A significant modality effect was found for the comparisons of VIS to 3-D and to SL (both βs=0.28), such that noise was lower for VIS than either SL or 3-D. In contrast, the regression comparing 3-D versus SL produced no modality effect ($P>0.50$). Distance was significant in all three analyses (range of $\beta=0.57$–0.70). Path type (direct versus indirect) was significant when comparing 3-D versus SL ($\beta=0.25$), similar to the effect in experiment 1 when reports from the origin were compared to those from the indirect waypoint. The path-type effect was also significant when VIS was compared to SL ($\beta=0.34$), but not when VIS was compared to 3-D ($P>0.25$).

Analogously to experiment 1, the noise measure was regressed on path type within each modality, with the observations representing the combinations of target location and indirect-walking direction. The effect was significant for SL ($\beta=0.60$, $P<0.01$) but not for 3-D or VIS ($\beta=0.09$ and 0.34, respectively). Thus, as in the previous study, updating on the indirect path augmented noise only when targets were learned through spatial language.

Discussion

Forward updating bias was observed for all modalities, and the magnitude of the forward bias was equivalent for forward and rightward walking. Although updating bias did not differ significantly by modality when the polar-coordinate vector was analyzed, analyses separating the forward and rightward components revealed differences. Specifically, SL showed greater forward updating bias, and only SL exhibited systematic rightward bias after rightward indirect walking. Moreover, as before, there was an increase in noise due to updating on the indirect path only in the SL condition. These effects violate the hypothesis that the images formed from different perceptual inputs function equivalently for purposes of updating.

Walking to the targets, rather than pointing and verbalizing distance as in experiment 1, reduced the updating bias in the 3-D sound and spatial language conditions, suggesting that some amount of the previous systematic error was attributable to the verbal response. The similarity of the direct-walking responses to those made at the last learning trial indicated that there was little updating bias attributable to the direct path. However, the magnitude of updating bias measured by comparing direct to indirect walking was still found to be greater than the level found with single targets by Loomis et al. (2002). The requirement to update multiple targets appears to introduce systematic error when participants walk along an indirect path.

Where does this additional bias come from? If memory load per se underlies the difference, the effect must be primarily in the move from single to multiple targets, as the number of multiple targets (3 versus 5) did not affect the results of experiment 1 (similar to findings of Rieser and Rider 1991). Another possibility is the extended learning process. In the Loomis et al. study, targets were responded to immediately after exposure, whereas, in the present study the targets were learned over multiple trials. The precise mechanism by which a longer learning stage would lead to increased updating bias is not known, but a relatively lower strength of the internal spatial representation could make it more susceptible to bias tendencies. The present data rule out miscalibration of self-motion as the sole basis for the updating bias, as regardless of indirect-walking direction, people moved the targets forward relative to the origin when they walked via the indirect waypoint. However, miscalibration could contribute to the direction-specific bias effects found for the SL condition. Systematic differences in calibration would not, however, explain the differential noise across modalities due to updating.

Other effects of this experiment were as expected. The 3-D condition produced an encoding bias that compresses distance responses. An advantage for vision with respect to measures of learning and noise was obtained, which can be attributed to the simultaneity of presentation that vision affords and its spatial precision. Confirming previous results, the 3-D sound and SL conditions did not differ with respect to learning and overall noise.

General discussion

The main purpose of the present research was to compare 3-D sound, spatial language, and vision as a means of conveying multiple object locations that allowed a person to update self-position while walking. As an extension of Loomis et al. (2002), we tested a model that assumed the modalities would differ in encoding the target locations into memory, but that the resulting spatial image would function equivalently for spatial updating across modalities. Overall, our results indicate that language can produce spatial images that support updating. However, in some respects language-based images do not function equivalently to those obtained via intrinsically spatial modalities, vision, and 3-D sound. In the following sections, we will discuss the implications of this work for spatial processes and for the hypothesis of functional equivalence across modalities.

Updating bias

The magnitude of updating bias observed here for multiple targets was greater than that observed for immediate responses to single targets by Loomis et al. (2002), although there was no effect of three versus five target locations. Moreover, in both experiments the updating bias primarily took the form of moving the target locations forward relative to their true values. This updating-induced error appears to reflect distortions in multiple remembered locations from the spatial perspective that occurred during learning.

The present data showed equivalent updating bias between the modalities of vision and 3-D sound, and in experiment 1, between 3-D sound and spatial language. However, the prediction of the image-updating model with regard to updating was not unequivocally supported, as there was evidence in experiment 2 of greater updating bias when targets were originally learned through spatial language. Specifically, the SL condition showed somewhat greater forward bias after indirect walking and was the only condition to produce significant rightward bias after rightward indirect walking.

These bias differences notwithstanding, participants who learned the target locations from spatial language, like those using 3-D sound and vision, evidenced substantial ability to update. It appears that some disadvantage accrues for updating of multiple targets that were learned through language, but updating nonetheless occurs. Below, we will consider the implications of these results for the issue of functional equivalence across modalities.

Experiment 1, which used pointing plus a verbal distance estimate, showed greater updating bias (i.e., less convergence of reports from the origin versus the indirect waypoint), than experiment 2, based on walking. Sizable discrepancies in verbally reported locations from an origin versus an indirect waypoint were similarly obtained by Böök and Garling (1981) and Loomis et al. (1998;

experiment 2). As was suggested above, the nature of the anticipated response, verbal versus action-based, may affect the process of updating en route to the response location. These findings further indicate that the measure of updating error should be based on action, such as the comparison of indirect to direct walking.

Noise in the memory representation

Noise in the memory representation could arise from multiple sources: at encoding of individual locations, loss of precision while multiple locations are learned and retained in memory, or effects of updating. That updating increases noise is indicated by the finding that the noise measure (based on interparticipant variability) increased with walked distance and was greater for the indirect path. Thus we find that a self-directed translation in space introduces loss of precision in localization of remembered landmarks. Others have found that the internal consistency of spatial relationships among multiple targets can be reduced by arbitrary motion of the body (Wang 1999; Wang and Spelke 2000).

When the effect of an indirect waypoint was evaluated by modality, only SL was found to produce a significant increment in noise due to the indirect response and the updating it required. This finding is consistent with other disadvantages for spatial language, namely, the greater updating bias and the slower learning of azimuths presented by spatial language that was found by Klatzky et al. (2002).

Noise was also affected overall by input modality. In particular, vision produced lower noise than spatial language or 3-D sound, which have found to be equivalent or to differ unreliably. The spatial precision of visual perception presumably reduces noise in the encoded spatial image, although we cannot rule out effects at later stages as well.

Encoding bias

Experiment 2 was unique in providing two measures of encoding bias. Participants pointed to the target direction and verbally estimated its distance as they learned, and after learning, they walked directly to the target locations. An important result was that these measures converged on common locations, indicating that there is little bias induced by direct walking.

The present studies measure bias in the encoded representations of locations that are learned over a series of presentations, as opposed to the immediate products of perception. One would expect, of course, that biases which are introduced by perception would be maintained in memory encoding, and that was true of the most substantial bias–the compression of distances that were encoded from 3-D sound. A further question is whether memory introduces biases of its own. In this regard, there was a tendency for overall under-responding to visually presented distances, as indicated by a negative zero intercept on the psychophysical function.

Learning rate for multiple object locations

Given our criteria of absolute direction accuracy and relative distance ranking, participants learned the locations of up to five target objects within nine exposures. Learning rates did not differ significantly between the auditory modalities. However, a deficit for spatial language in learning azimuths might have been masked in the present study by the advantage of using a verbal report to assess distance learning.

The present data did show faster learning for the visual modality than for the auditory conditions. The source of this advantage cannot be specified from our data, but it is likely to reflect the simultaneous exposure of the visual targets and the flexibility it affords for rehearsal. Klatzky et al. (2002) found that there was no advantage for vision over audition when the object labels that marked azimuths were exposed successively with limited field of view.

Implications for functional equivalence across modalities

Each of the measures we examined in these studies showed modality effects: encoding, learning, updating bias, and noise. Most critical for the image-updating model, and its hypothesis of functional equivalence of spatial images, are the findings showing modality differences in updating of self-position while walking. Functional equivalence appears to apply to 3-D sound and vision, but spatial language produces additional error, both systematic and random.

When modalities are found to function equivalently in updating, this does not specify the nature of the spatial images that are utilized. Images might be modality specific, but functional equivalence could still arise by coincidence, if the various representations have the same bias and noise tendencies. Alternatively, equivalent performance might result from encoding images into a common format that is supramodal, i.e., at a level of abstraction not tied to any one modality. The posterior parietal cortex has been implicated as one locus for such a representation, as it combines inputs from multiple modalities to form a unified representation of space (see, for example, Andersen 1999; Farah et al. 1989).

Another potential basis for functional equivalence is a common representational format that is modality specific, such as a visual image. A pathway from 3-D auditory cues to a visual image might involve cortical neurons that code both auditory and visual spatial inputs, as have been found in the ventral premotor cortex of the monkey (Graziano and Gross 1998; Graziano et al. 1999). Spatial language too might ultimately be represented through visual coding, as verbal descriptions are often used to induce visual images, and brain imaging studies have shown that early visual processing areas are activated as a result (reviewed in Thompson and Kosslyn 1999). However, an argument against a visual format for the spatial images from 3-D sound and spatial language is the finding of Loomis et al. (2002) that congenitally blind participants were able to perform updating these inputs; presumably they did not recode the stimuli visually.

Just as functional equivalence does not guarantee a common type of representation in memory, lack of equivalence does not rule it out. There are two ways to interpret modality differences in updating. One is that the nature of the spatial *representation* is fundamentally different. The other is that properties of the image such as memory strength or precision differ, leading to *processing* differences that affect performance.

This latter account may explain the present finding that SL updating performance was highly similar to VIS and 3-D, but generated additional bias and noise. The fundamental difference between language and 3-D sound or vision is that the latter modalities offer spatial cues directly through perceptual processing, whereas non-spatial verbal descriptions must be converted to a representation that affords spatial processing. Klatzky et al. (2002) found that this conversion adds to learning time. A possible result is that the spatial images formed from language have weaker strength and require more processing capacity to update. For example, capacity demands could influence calibration of self-motion, which would produce bias specific to the direction of travel. The finding that updating bias is higher with multiple remembered targets than with a single perceived target supports the idea that memory demands can affect updating.

Issues for future research

The present experiments make significant advances in our understanding of non-visually guided navigation. They demonstrate that across variations in input modality, encoding of multiple spatial locations results in memory representations that can be updated during locomotion, although with some differential vulnerability to updating bias and noise. At the same time, the studies raise a number of important topics to be addressed in future research. One is the nature of the spatial image that is formed from various modalities. This might be investigated, for example, by imposing modality-specific interference during indirect walking. Another issue is how memory demands affect updating, which could be addressed by manipulations such as varying the retention interval between learning and updating. Another question is the basis for the specific pattern of updating bias that was attained. Experiment 2 ruled out miscalibration of velocity as the sole cause of bias, but did not point to the mechanism that would cause apparent target locations to recede from the origin. This might be addressed with an array of locations that surrounded the participant, to determine whether targets move relative to the facing direction or relative to the origin. The neural mechanisms that underlie cross-modal spatial images are also, of course, an exciting topic for further research.

Acknowledgements This research was supported by National Eye Institute grant EY09740. A preliminary report of this work was reported at the 2001 Annual Meeting of the Psychonomic Society in Orlando, Florida. We thank Jerome Tietz for constructing the set up in experiment 2, Mike Kinsella for running participants in experiment 1, and Eyal Aharoni and Max Krasnow for helping to run participants in experiment 2. We also thank the reviewers.

References

Amorim MA, Glasauer S, Corpinot K, Berthoz A (1997) Updating an object's orientation and location during nonvisual navigation: a comparison between two processing modes. Percept Psychophys 59:404–418

Andersen RA (1999) Multimodal integration for the representation of space in the posterior parietal cortex. In: Burgess N, Jeffery KJ, O'Keefe J (eds) The hippocampal and parietal foundations of spatial cognition. Oxford University Press, New York, pp 90–103

Bachmann E, Perrig WJ (1988) Die mentale Repräsentation von verbalen Ortsbeschreibungen. Schweizerische. Z Psychol 47:25–36

Beall AC, Blascovich J, Loomis JM (2001) PC-based immersive virtual environment technology as a behavioral science research tool. Unpublished manuscript, University of California, Santa Barbara

Böök A, Gärling T (1981) Maintenance of orientation during locomotion in unfamiliar environments. J Exp Psychol Hum Percept Perform 7:995–1006

Bosco A, Filomena S, Sardone L, Scalisi TG, Longoni AM (1996) Spatial models derived from verbal descriptions of fictitious environments: the influence of study time and the individual differences in visuo-spatial ability. Psychol Beitr 38:451–464

Farah MJ, Wong AG, Monheit MA, Morrow LA (1989) Parietal lobe mechanisms of spatial attention: modality-specific or supramodal? Neuropsychologia 27:461–470

Farrell MJ, Thomson JA (1999) On-line updating of spatial information during locomotion without vision. J Mot Behav 3:39–53

Ferguson EL, Hegarty M (1994) Properties of cognitive maps constructed from texts. Mem Cognit 22:455–473

Francis WN, Kucera H (1982) Frequency analysis of English usage: lexicon and grammar. Houghton Mifflin, Boston

Franklin N, Tversky B (1990) Searching imagined environments. J Exp Psychol Gen 119:63–76

Graziano MS, Gross CG (1998) Spatial maps for the control of movement. Curr Opin Neurobiol 8:195–201

Graziano MS, Reiss LA, Gross CG (1999) A neuronal representation of the location of nearby sounds. Nature 397:428–430

Haber L, Haber RN, Penningroth S, Novak K, Radgowski H (1993). Comparison of nine methods of indicating the direction to objects: data from blind adults. Perception 22:35–47

Klatzky RL, Lippa Y, Loomis JM, Golledge R G (2002) Learning directions of objects specified by vision, spatial audition, or auditory spatial language. Learn Mem (in press)

Loomis JM, Da Silva JA, Fujita N, Fukusima SS (1992) Visual space perception and visually directed action. J Exp Psychol Hum Percept Perform 18:906–921

Loomis JM, Klatzky RL, Philbeck JW, Golledge RG (1998) Assessing auditory distance perception using perceptually directed action. Percept Psychophys 60:966–980

Loomis JM, Lippa Y, Klatzky RL, Golledge RG (2002) Spatial updating of locations specified by 3-D sound and spatial language. J Exp Psychol Learn Mem Cogn 28:335–345

Montello DR, Richardson AE, Hegarty M Provenza M (1999) A comparison of methods for estimating directions in egocentric space. Perception 28:981–1000
Ooi TL, Wu B, He ZJ (2001) Distance determined by the angular declination below the horizon. Nature 4:97–200
Perrig W, Kintsch W (1985) Propositional and situational representations of text. J Mem Lang 24:503–518
Philbeck JW, Loomis JM, Beall AC (1997) Visually perceived location is an invariant in the control of action. Percept Psychophys 59:601–612
Philbeck JW, Klatzky RL, Behrmann M, Loomis JM, Goodridge J (2001) Active control of locomotion facilitates nonvisual navigation. J Exp Psychol Hum Percept Perform 27:141–153
Presson CC, Montello DR (1994) Updating after rotational and translational body movements: coordinate structure of perspective space. Perception 23:1447–1455
Rieser JJ (1989) Access to knowledge of spatial structure at novel points of observation. J Exp Psychol Learn Mem Cogn 15:1157–1165
Rieser JJ, Rider EA (1991) Young children's spatial orientation with respect to multiple targets when walking without vision. Dev Psychol 27:97–107
Taylor HA, Tversky B (1992) Descriptions and depictions of environments. Mem Cogn 20:483–496
Thomson JA (1983) Is continuous visual monitoring necessary in visually guided locomotion? J Exp Psychol Hum Percept Perform 9:427–443
Thompson WL, Kosslyn SM (1999) Neural systems activated during visual mental imagery. In: Toga AW, Mazziotta JC (eds) Brain mapping: the systems. Academic, San Diego, pp 535–560
Wang RF (1999) Representing a stable environment by egocentric updating and invariant representations. Spat Cogn Comput 1:431–445
Wang RF, Spelke ES (2000) Updating egocentric representations in human navigation. Cognition 77:215–250
Wilson PN, Tlauka M, Wildbur D (1999) Orientation specificity occurs in both small- and large-scale imagined routes presented as verbal descriptions. J Exp Psychol Learn Mem Cogn 25:664–679