

Spatial Learning and Navigation Using A Virtual Verbal Display

NICHOLAS A. GIUDICE

University of Maine

JONATHAN Z. BAKDASH

University of Virginia

and

GORDON E. LEGGE and RUDRAVA ROY

University of Minnesota, Twin Cities

We report on three experiments that investigate the efficacy of a new type of interface called a virtual verbal display (VVD) for nonvisual learning and navigation of virtual environments (VEs). Although verbal information has been studied for route-guidance, little is known about the use of context-sensitive, speech-based displays (e.g., the VVD) for supporting free exploration and wayfinding behavior. During training, participants used the VVD (Experiments I and II) or a visual display (Experiment III) to search the VEs and find four hidden target locations. At test, all participants performed a route-finding task in the corresponding real environment, navigating with vision (Experiments I and III) or from verbal descriptions (Experiment II). Training performance between virtual display modes was comparable, but wayfinding in the real environment was worse after VVD learning than visual learning, regardless of the testing modality. Our results support the efficacy of the VVD for searching computer-based environments but indicate a difference in the cognitive maps built up between verbal and visual learning, perhaps due to lack of physical movement in the VVD.

Categories and Subject Descriptors: J.4 [Social and Behavioral Sciences]: Psychology; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*; H.5.2 [Information Interfaces and Presentation]: User interfaces—*Natural language, Graphical user interfaces*

General Terms: Experimentation, Performance

Additional Key Words and Phrases: Navigation, wayfinding, verbal learning, virtual environments, virtual verbal display, and human-computer interaction

ACM Reference Format:

Giudice, N. A., Bakdash, J. Z., Legge, G. E., and Roy, R. 2010. Spatial learning and navigation using a virtual verbal display. *ACM Trans. Appl. Percept.* 7, 1, Article 3 (January 2010), 22 pages.

DOI = 10.1145/1658349.1658352 <http://doi.acm.org/10.1145/1658349.1658352>

This research was supported by NIDRR grant H133A011903, NIH training grant 5T32EY07133, and NIH grant EY-02857.

Authors' addresses: Nicholas A. Giudice, Department of Spatial Information Science and Engineering, University of Maine, Orono ME, 04469; email: giudice@spatial.maine.edu; Jonathan Z. Bakdash, Department of Psychology, University of Virginia, Charlottesville, VA 22904-4400; email: jzb3e@virginia.edu; Gordon E. Legge, Department of Psychology, and Rudrava Roy, Department of Computer Science, University of Minnesota, Minneapolis, MN 55455-0344; email: legge@umn.edu and rudrava@theroyweb.com.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2010 ACM 1544-3558/2010/01-ART3 \$10.00

DOI 10.1145/1658349.1658352 <http://doi.acm.org/10.1145/1658349.1658352>

1. INTRODUCTION

Research addressing the connection between language and spatial learning can be divided into two general categories: studies investigating comprehension of spatial texts [Denis and Cocude 1989; Denis and Zimmer 1992; Ferguson and Hegarty 1994; Johnson-Laird 1983; Perrig and Kintsch 1985; Taylor and Tversky 1992] and those addressing the use of verbal directions for navigating routes in real-world environments [Allen 1997; Allen 2000; Lovelace et al. 1999; Tom and Denis 2003; Tversky 1996]. This article extends the investigation to a third domain, the use of verbal information to describe, learn, and navigate computer-based virtual environments (VEs). Most research using virtual environment technology (VET) has employed visually rendered displays. Visually rendered VEs are an excellent research tool for studying spatial cognition, as they facilitate manipulation of spatial properties or viewing perspectives that may otherwise be difficult to vary during real-world navigation [Loomis et al. 1999; Peruch et al. 2000]. The current studies are the first-known research to investigate environmental learning and wayfinding of computer-simulated layouts based on a completely nonvisual interface, called a virtual verbal display (VVD). Potential applications for the VVD include persons with low vision, situations where vision is impaired (e.g., firefighters), or for use in real-time navigation systems.

The VVD provides first-person verbal messages about a navigator's position and orientation in the environment and a description of the layout geometry (corridor structure) at their location. A sample output string is: "You are facing north, at a 3-way intersection, there are hallways ahead, right and behind." An important aspect of the VVD is that the information provided is dynamically updated, that is, based on context sensitive verbal messages that are contingent on the navigator's current position and orientation in the environment. Thus, if the participant made a 90° right rotation at the t-junction just described, the VVD would update the message to reflect that they were now facing east, with hallways extending ahead, left, and right.

The use of dynamically updated verbal descriptions differs from the static descriptions adopted by the previously mentioned research investigating text comprehension and verbal route navigation. While static descriptions are not updated with respect to movement, dynamically updated descriptions provide real-time information about changing position and orientation with navigation.

Dynamically updated auditory interfaces may be based on displays incorporating either spatial language, as is adopted in the current experiments, or spatialized sound, where an object is heard as coming from a specific location in 3D space [Loomis et al. 1990]. The research conducted with both types of interfaces has employed what Giudice [2004] dubbed a point-based display. With point-based displays, dynamically updated information is provided about the distance and direction of discrete landmarks and decision points in the environment or for giving updated route instructions, that is, as is done with speech-based in-vehicle navigation systems. The empirical research with these displays has mostly been in the context of navigation systems for the blind, where they have proven extremely effective for supporting route navigation and nonvisual guidance to a goal state [Loomis et al. 1998, 2001; Loomis et al. 2005; Marston et al. 2006; Petrie 1996; Tom and Denis 2003].

The current studies adopt a different type of dynamically updated verbal interface, which we call a geometric-based display. In contrast to point-based displays, which provide updated information about routes or specific landmarks, the geometric display conveys updated information about basic layout geometry, spatial configuration, and viable paths of travel at the user's current location. By adopting these geometric descriptions, we are able to extend the investigation of verbal spatial learning with dynamically updated displays from navigation of routes between landmarks to more complicated spatial operations such as free exploration (open search), environmental learning (configurational knowledge), and wayfinding (the ability to plan and execute routes, even if not previously traveled). With the exception of a study by Giudice et al. [2007], little is known about the capacity of dynamically updated

verbal information to support such tasks. In the previous work, Giudice et al. asked blindfolded-sighted participants to walk through real buildings and explore the layout in search of several targets. Their exploration was supported by messages in the form of verbal descriptions. The messages were automatically updated at every intersection or with any change in heading. During an initial training session, subjects were asked to use the descriptions to execute an exhaustive search, that is, traverse the entire floor (comprising approximately 500 feet of corridor extent and 12 intersections), and to seek out four hidden target locations. During transfer, they were provided with the same verbal descriptions but had to find routes between target pairs (e.g., “you are at target X, find target Y”). The results of this experiment demonstrated that people were extremely effective at using geometric verbal descriptions to perform searches, learn routes, and build up a cognitive map supporting wayfinding performance at test. Using the same verbal descriptions, the goal of the current studies was to extend our earlier work in real buildings to VEs.

One major difference between the previous studies and the current experiments is that the body-based information present during physical navigation in real environments, such as proprioceptive and kinesthetic feedback, are not available during virtual navigation in the VVD. In other tasks requiring mental transformations of space after verbal learning (e.g., reading spatial texts), evidence suggests that spatial updating and egocentric pointing performance is facilitated when introducing physical movement [Avraamides 2003; De Vega and Rodrigo 2001]. If physical body movement is a critical factor for converting the verbal code into a spatial form, we might expect difficulties when learning from the VVD, which employs imagined movement via keyboard navigation. On the other hand, keyboard navigation in the VVD likely affords greater immersion than reading a text, as the input is directly coupled to perceptual feedback about movement behavior; information which is known to improve immersive learning in a VE [Clawson 1998; Witmer et al. 1996; Ruddle and Lessels 2006]. Since the previous work in real buildings used the same training procedure and verbal descriptions as is employed in the current studies with training in simulated layouts, a similar pattern of results would suggest that there is no adverse affect from the lack of physical movement cues during virtual verbal learning. However, if differences are manifest, they could be due to (i) lack of movement in the VVD or (ii) differences in the test procedure between experiments, as the previous work used verbal descriptions during testing and the analogous study (Experiment I) uses vision at test. Theories of functional equivalence would predict that different test modalities should not affect wayfinding performance given the same training procedure [Loomis et al. 2002]. Experiment II addresses this issue by using the same verbal information at training and test, as was done in the previous Giudice [2007] study in real buildings. Thus, comparing navigation performance at test between Experiments I and II, which employ identical training procedures but different testing modalities, clarifies whether differences manifest between visual and verbal read-out operations.

The VVD used in Experiments I and II is similar to a traditional, visually rendered desktop virtual reality (VR) system, where participants have a first-person view of a computer-based environment and virtually navigate by means of the keyboard. With the VVD, however, rather than seeing the environment on a computer monitor, participants hear, through the computer’s speakers, synthetic speech messages describing their position and orientation in the layout. As with a visual display, the information provided by the verbal message changes egocentrically in relation to the user’s movement through the space. Visual displays generally provide optic flow information to specify motion and a range of 3D depth cues. With auditory displays, one way to present this information is to use sound intensity and 3D spatialized messages to indicate the changing distance and direction of landmarks, as was done in the point-based displays used by Loomis et al. [1998]. Another approach, adopted by the VVD, indicates distance and direction information about nearby hallways explicitly in the updated verbal descriptions and a tone to indicate participant movement.

Table I. Overview of Experiments

| Experiment | Learning Session (virtual environment) | Transfer Session (real environment) | Experiment Goals |
|------------|--|---|---|
| | Modality | Modality | |
| I | VVD | Visual | a) Does the VVD support free exploration during learning and accurate wayfinding during transfer using vision in the real environment? |
| II | VVD | Verbal | b) Does verbal view-depth affect learning and transfer? Control for Experiment I: assessing whether same modality at learning and transfer leads to better test performance. |
| III | Visual | Visual | Assessing similarity of visual learning to verbal learning (Experiment I). |

Note. Summary table of the environment and modality for learning and test sessions by experiment. VVD denotes virtual verbal display.

Critical to the current research are findings demonstrating that visually rendered virtual displays need not have high spatial fidelity and rich detail to be effective. Indeed, learning from perceptually sparse environments, those rendered using information about geometric structure only, lead to accurate spatial learning [Aguirre et al. 1996; Giudice 2002; Stankiewicz et al. 2006; Waller et al. 2001]. Accurate environmental transfer to real-world navigation has also been demonstrated after learning with low-fidelity virtual displays [Bliss et al. 1997; Kalia et al. 2008; Ruddle et al. 1997; Schlicht 2001; Waller et al. 1998]. While there is some variability in the level of environmental transfer reported after learning in sparse VEs, the results of these studies provide clear evidence that access to layout geometry is sufficient to learn virtual spaces and that this knowledge is transferable to physical navigation of the real-world places. Since the sparse geometric information of these visually rendered VEs are similar to the information conveyed by our VVD, we hoped to show a corresponding high level of environmental learning and transfer performance in the current experiments.

Experiment I addresses the efficacy of our VVD by having subjects freely explore unfamiliar computer-simulated layouts using the VVD and then testing whether this knowledge transfers to visual navigation of the corresponding real environments. Experiment II follows the same design, but computer-based training and real-world navigation are both performed using the geometric verbal descriptions. It addresses whether the cognitive map built up from verbal learning with the VVD is easier to access when tests are carried out in the same modality as encoding or whether the verbal descriptions build up into a common spatial representation in memory [Bryant 1997; Loomis et al. 2002]. Experiment III follows the same procedure as Experiment I, but rather than verbal learning of computer-based environments, it uses visual learning with an information-matched computer display. While both experiments stand on their own, the use of a common train/test procedure provides a relative measure of navigation performance after learning with verbal and visual displays. These data speak to the similarity of the spatial representations built up between inputs. See Table I for an overview of experimental conditions and goals.

2. EXPERIMENT I: VIRTUAL VERBAL LEARNING AND ENVIRONMENTAL TRANSFER

This study investigated whether blindfolded sighted participants could use a VVD to freely explore unfamiliar computer-based training environments, search for hidden target locations, learn routes between these targets, and build up a cognitive map that supported effective navigation in the corresponding real environments. The experiment incorporated three verbal display modes. The verbal modes varied the extent of environmental detail described to the user from their position in the layout and were chosen as they represent a broad range from proximal to distal information access. These levels of environmental description are called verbal view depth (see Giudice et al. [2007] for a more detailed discussion). In summary, the three view-depth conditions range from a description of only local geometric detail at the

user's position (called Local mode), to a description that provides information about the distance and geometry of all adjacent intersections at the user's position (called Maplet mode), to a description of the environment's overall layout configuration (called Global mode).

Our previous work in real buildings employed the same three verbal view-depth conditions but no differences in learning or wayfinding performance were observed between verbal modes. Two factors lead us to expect a more pronounced effect of view-depth with the VVD in the current experiments. First, spatial updating is known to be more cognitively effortful in VEs than in real spaces [Richardson et al. 1999; Wilson et al. 1997]. Second, there is increased cognitive cost associated with the indirect processing of spatial language into a spatial representation, as compared to the spatial senses [Klatzky et al. 2003]. The greater cognitive demands associated with VE learning, coupled with the cognitive resources required for converting the symbolic verbal messages, indicate that we will see a view-depth effect with the VVD. If integrating many local samples of the environment into a global understanding of layout configuration is difficult, we would expect deficits in the local condition, especially in tasks requiring accessing one's cognitive map during the planning and updating of novel routes during the transfer test in the real environment. On the other hand, if decoding the longer messages of the Maplet and Global descriptions proves challenging, then these conditions will yield the worst performance, even though they afford access to greater environmental detail.

The principal questions addressed by this experiment are the following.

- (1) Does use of a VVD support free exploration and wayfinding of novel computer based environments?
- (2) Does learning with the VVD in simulated layouts transfer to accurate visually based navigation in the corresponding real environment?
- (3) Does the amount of layout geometry conveyed by individual messages from the verbal display affect training behavior or environmental transfer performance?

2.1 Method

2.1.1 Participants. Nineteen sighted participants, nine male and ten female, between the ages of 17 and 24 (mean = 20.9) ran in the experiment. They received course credit for their time. All subjects were blindfolded during the training session of the experiment.

2.1.2 Environments. The environments consisted of portions of three computer-based simulations of floors of the Psychology department building at the University of Minnesota and the corresponding three real floors. The floors were of similar size and complexity but differed in layout topology, reducing the chance of transfer of learning between environments. The environments averaged 560 feet of corridor extent and 13.6 decision points (intersections). Four targets, presented as verbal messages (e.g., dog, cat) were placed in each layout (see Figure 1 for an illustration of the three layouts).

The VEs were rendered to be perceptually sparse. That is, only information about layout geometry (e.g., corridor connectivity) and metric information (e.g., distance between intersections) was described. All of the simulated layouts were made to fit a Cartesian grid by breaking them into corridor segments separated by nodes (each segment approximated 15 feet in the real space).

2.1.3 Movement Behavior. During training, blindfolded participants virtually explored the computer-based environments using the keyboard to simulate movement. To navigate, they moved forward by pressing the up arrow and turned by pressing the left and right arrows. Each forward key press translated the participant one segment along the virtual corridor and each press of the left and right arrow rotated them 90 degrees in place. Speech descriptions were generated automatically upon reaching an intersection and an updated message about user heading was given with any rotation. A beep was generated with each forward key press when navigating a hallway between intersections and

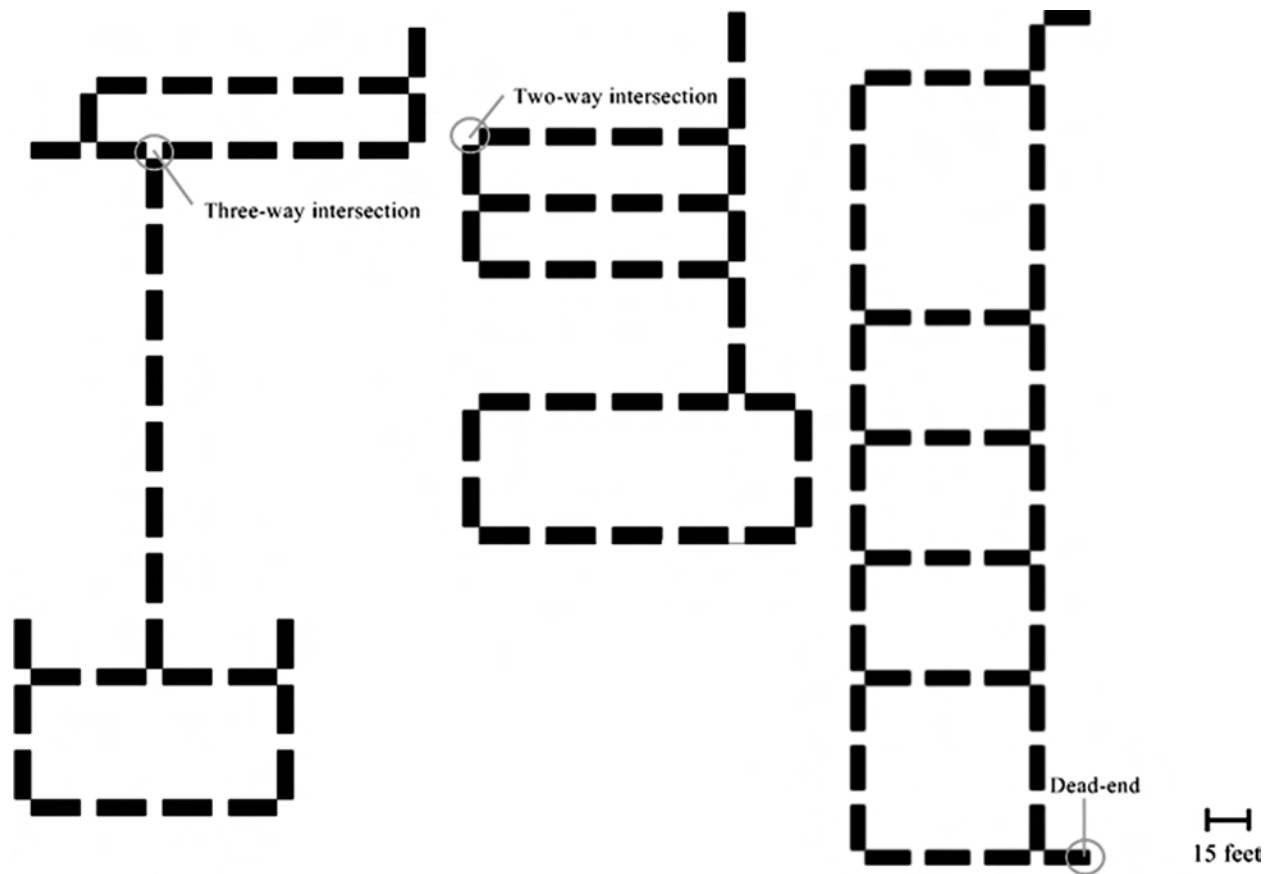


Fig. 1. Three experimental layouts with intersection types denoted. Floor 3, 4a, and 4b (left to right).

an error tone was sounded if subjects tried to move forward while facing a wall. Movement transitions took approximately 1 second, a “speed limit,” which was found to optimize the balance between moving too fast to hear the messages or too slow, which caused an unnatural interstimulus lag. All verbal descriptions could be repeated by pressing the down arrow key.

The experiment was run on a Dell Precision 610 desktop computer with a 650MHz Pentium III Xeon processor. The environments were rendered using custom software written in the Python programming language utilizing Win32 API GDI functions for visual rendering and the MS Speech SDK (SAPI 5) for verbal rendering. The synthetic speech descriptions were delivered in a clear, intelligible female voice at a user-adjustable volume through the computer’s onboard sound card and a pair of stereo speakers (Harmon Kardon HK195), positioned approximately 60 cm in front of the participant at approximately 15 degrees off midline.

2.1.4 Verbal Modes. The experiment employed three VVD modes, each providing the user with a different level of verbal view-depth information about corridor structure (see Figure 2).

- (1) Local verbal mode describes layout geometry at the user’s current position. “Facing east, at a two-way intersection, ahead is a hallway, to the left is a hallway.”
- (2) Maplet verbal mode includes the local information and adds a description of the distance and

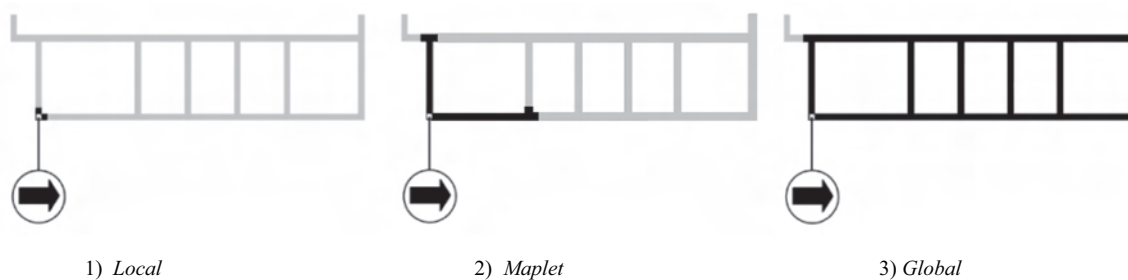


Fig. 2. The circle and arrow represent the user's location and orientation in the layout. The black lines are the areas described by the verbal description, and the gray lines symbolize the entire floor.

geometry for all adjacent intersections. Note that the information about adjacent intersections is given in first-person, as if walking down the hallway being described. "Facing east, at a two-way intersection, ahead is a 60 foot hallway extending through a 3-way intersection to the left, to the left is a 45 foot hallway ending at a 3-way intersection."

- (3) Global verbal mode includes the Maplet information and adds a general description of the overall geometric structure of the layout. "This floor can be thought of as a 195 by 45 foot east-west rectangle intersected by four north-south hallways..." The global message was immediately followed by a Maplet description. To reduce verbosity, the full global description was only spoken three times, at the beginning, one-quarter, and three-quarters of the way through the training session. The maplet description was given the rest of the time.

2.1.5 Design and Procedure. The study employed a within-subjects design with participants training on three unfamiliar floors, one for each of the verbal conditions. The order of floor by verbal mode was counterbalanced as fully as possible given the number of participants using a Latin Square design. The experiment took approximately 5 hours per subject and was conducted in two sessions. It was broken into three phases: a practice session, a training session, and a transfer session for testing.

The practice session provided an explanation of the three verbal modes, movement behavior, and test procedures. Participants were visually presented with a sample map and shown what would be heard from each verbal mode. Before proceeding, they had to accurately describe an example of each intersection using the terminology of the three verbal modes. Participants were then blindfolded, presented with a practice VE, and run through the entire experimental procedure.

During the training session, participants explored three VEs using a VVD based on each of the three verbal modes (Local, Maplet, and Global). To perform the task, they were blindfolded, seated in front of a computer running the VVD, and directed to the navigation (arrow) keys. The training session began from a random starting position in the environment. Participants were instructed to use the navigation keys and the verbal information provided by the VVD to search the entire floor and find four target locations. Although no specific information was given about routes, they were encouraged to use a search strategy that would facilitate route finding between all targets. The targets consisted of high imagery words, such as dog, cat, horse, and pig, and were always situated on a wall at an intersection. When encountered, their name and position, both in relative and absolute terminology, was announced, for example, "Facing east, target dog is to your left." Participants were asked to turn and face each target, imagining it as a picture on the wall. Training continued until they traversed three times the total number of segments comprising the environment. In this way, training was normalized across participants, irrespective of view-depth condition, movement speed, or layout size. Each forward

Table II. Experiment I: Training Measures by Verbal View-Depth Condition

| View-depth condition | Floor coverage percentage (%) | Percentage of unique targets encountered (%) | Number of shortest paths traversed | Entropy |
|----------------------|---|--|---|--|
| Local | M = 92.61 SE = 2.19 | M = 98.68 SE = 1.32 | M = 9.68 SE = 0.78 | M = 4.98 SE = 0.05 |
| Maplet | M = 96.37 SE = 1.40 | M = 97.37 SE = 2.63 | M = 9.79 SE = 0.68 | M = 4.97 SE = 0.06 |
| Global | M = 96.41 SE = 1.15 | M = 100 SE = 0% | M = 10.74 SE = 1.23 | M = 4.85 SE = 0.06 |
| Overall | M = 95.13 SE = 1.58 $\eta^2 = 0.03$ | M = 98.68 SE = 1.32 $\eta^2 = 0.02$ | M = 10.07 SE = 0.89 $\eta^2 = 0.06$ | M = 4.94 SE = 0.03 $\eta^2 = 0.06$ |

Note. Each cell represents the mean and standard error from 19 participants. Overall values are based on the average of the three view-depth conditions and include the effect size for each repeated measures ANOVA.

key press (15-foot segment) was considered a move, and they were alerted when 50% and 75% of their moves had been used.

Immediately following the training session, participants engaged in an environmental transfer test requiring navigation of the physical floor corresponding to the simulated training environment. They were led, blindfolded, along a circuitous route to ensure initial disorientation and started at one of the target locations on the real floor. The blindfold was then removed and they were asked to walk the most direct route to a second target location. Routes between four unique target pairs were requested. The last route always returned to the first target, making the test trajectory a complete loop. During these transfer tests, participants navigated using vision; no verbal descriptions about the environment or target locations were provided. Participants indicated that they had reached the appropriate target location by giving a verbal confirmation of the target's name and an estimate of its cardinal direction (e.g., "I have reached target dog and am facing north"). To reduce the accumulation of errors during route finding trials, participants were brought to the correct target location for incorrectly localized targets before proceeding to the next trial. Participants found routes between four target pairs, the order of which were counterbalanced.

2.2 Results and Discussion

2.2.1 Training Session: Measures of Search Behavior. Four measures were obtained from participants' training trajectories:

- (1) floor coverage percentage: amount of total floor covered during training;
- (2) unique targets encountered percentage: out of four in total;
- (3) number of shortest paths traversed: a shortest path equals the route between target locations with the minimum number of intervening nodes;
- (4) entropy: used to describe the distribution of moves during the search. High entropy indicates that participants are distributing their movement equally across the entire environment and low entropy indicates that they are concentrating their search to specific regions of the layout [Schlicht 2001]. Entropy is expressed by the equation: $H(e) = -\sum_x p(x) \log_2[p(x)]$, where e is the environment and x is an individual node. The probability the subject visited an individual node is $p(x)$; calculated from the number of times node x was reached divided by the total number of forward moves executed during training.

The results of these four training measures are shown in Table II. Across all view-depth conditions, participants traveled over 95% of the nodes on each floor and found over 98% of the target locations during the training session. Although no explicit information was provided about routes, participants

traveled more than 10 shortest paths between targets. Given the same number of moves during training and 100% floor coverage, the theoretical maximum number of shortest paths that could be traversed is 21.3, averaged across floors.

Consider two different search strategies: With an environmental learning strategy, we would expect subjects to learn the environment by broadly distributing their moves throughout the layout and traveling as many unique routes between target locations as possible in order to learn the complete connectivity matrix. In contrast, we would expect subjects using a search strategy aimed at traversing as many routes as possible to adopt a less-distributed pattern of movement behavior, which concentrates on back-and-forth travel of the same paths between targets. The Entropy and shortest path data from all of our subjects provides evidence that they were using an environmental search strategy. The overall entropy for participants in Experiment I was 4.94, demonstrating that they were broadly distributing their moves throughout the layout. Subject performance approached the theoretical maximum entropy score of 5.21, representing the most equally distributed training session possible across environments. This broadly distributed exploration was coupled with travel of 74% (7.4 of 10) unique shortest paths during the average search, with no single route traveled more than three times. As shown in Figure 3, the search trajectory for the max route strategy yields a very different pattern of results. Moves are concentrated to a particular region of the floor, yielding a low entropy score of 3.47. Likewise, only 19% (4 of 21) of the shortest paths traveled are unique, with the majority of the search encompassing back-and-forth movement along the same routes with the least number of nodes between targets (e.g., pig to sheep and sheep to pig).

In summary, although no information about specific routes was given, participants movement patterns suggested that they adopted a highly efficient environmental learning strategy based on a broadly distributed search that facilitated travel of previously untraveled paths. Results from all four of the training measures almost perfectly replicate the results of the same training measures from our study with verbal learning in real buildings [Giudice et al. 2007]. This similarity suggests that the absence of real movement information in the VVD does not adversely affect search performance and provides clear evidence that effective virtual exploration is possible using dynamically updated geometric descriptions in a verbally rendered computer display.

A one-way repeated measures ANOVA comparing the three levels of verbal view-depth (Local, Maplet, and Global) was performed for each of the training measures, but no statistically reliable differences in training performance by view-depth were found, $ps > 0.05$. These data indicate that even with the increased cognitive demands associated with navigating VEs [Richardson et al. 1999; Wilson et al. 1997], a minimalist Local description is all that is needed to support efficient search behavior with the VVD.

2.2.2 Transfer Session. The next analyses addressed how well training with the VVD transferred to real-world test performance. Three measures were analyzed: target localization accuracy (percent of target locations correctly found), target estimated direction accuracy (percent of targets that were said to be oriented in the correct cardinal direction), and route efficiency (the length of the actual route between correct target localization trials divided by the length of the route executed). The results of these measures are shown in Table III.

The ability of subjects to correctly localize 51.3% of the targets at test was significantly above chance performance of ~3%, defined as 1 divided by 37 possible target locations (e.g., a target can be located at any of the 37 segments comprising the environment, $t(56) = 10.89$, $p < 0.001$). By comparison, participants using verbal descriptions to perform the same task in our study in real buildings yielded 85% target localization accuracy [Giudice et al. 2007]. Formal comparisons cannot be made between the previous and current experiments as they differed both in training environment (i.e., real vs. virtual) and

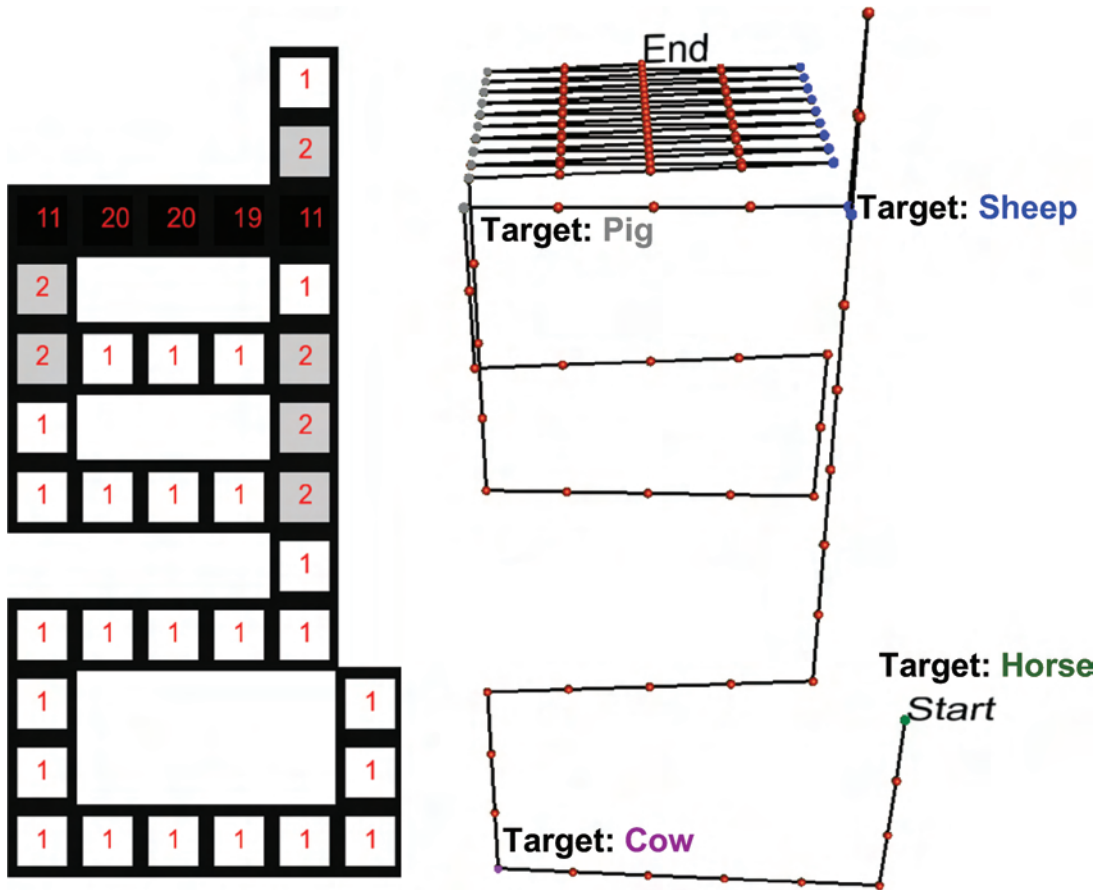


Fig. 3. Hypothetical example of the theoretical maximum shortest paths (21) for floor 4a. Frequency of visits to locations is indicated by red numbers inside each square (left). Trajectory and target locations are also shown (right). In this example, entropy = 3.47, note the high concentration of moves between the targets “Pig” and “Sheep.”

Table III. Experiment I: Transfer Test Measures by Verbal View-Depth Condition

| View-depth condition | Target localization accuracy (%) | Target estimated direction accuracy (%) | Route Efficiency (%) |
|----------------------|---|---|---|
| Local | M = 46.05 SE = 8.61 | M = 68.42 SE = 5.99 | M = 98.32 SE = 1.18 |
| Maplet | M = 53.95 SE = 7.22 | M = 61.84 SE = 9.42 | M = 97.18 SE = 1.54 |
| Global | M = 96.41 SE = 1.15 | M = 100 SE = 0.00 | M = 95.37 SE = 2.79 |
| Overall | M = 65.47 SE = 5.66 $\eta^2 = 0.04$ | M = 76.75 SE = 5.14 $\eta^2 = 0.02$ | M = 96.96 SE = 1.84 $\eta^2 = 0.05$ |

Note. Each cell for target accuracy represents the mean and standard error of 4 trials for each of 19 participants. Route efficiency is based only on correct trials from the same Ss. Overall is based on an average of the three view-depth conditions and includes the effect size for each repeated measures ANOVA.

testing modalities (i.e., navigating with verbal descriptions vs. vision respectively). If only one parameter varied, comparisons could be made, as is done in Experiment II of this article. Even so, the more than 50% difference of the current study represents a dramatic performance hit. Since performance on the training measures was almost identical between learning in real environments and VEs, the decline in test performance cannot be attributed to difficulties in using the verbal information to support the task or to errors in interpreting the output of the VVD. For these scenarios to be true, we would have either (i) observed poor training performance in both our previous experiment in real buildings and the current study in virtual layouts or (ii) found reliably worse performance on the training measures with the VVD in the current experiment. Since neither outcome is supported by the data, the results of Experiment I suggest that the lower target localization accuracy in the current study stems from the development of a deficient spatial representation when training with the VVD compared to training in real building layouts.

The finding that subjects correctly estimated absolute target direction 64.7% of the time suggests that first-person verbal descriptions (e.g., the left-right relative messages used by the VVD), also benefit from the use of absolute terminology (e.g., North-south heading indicators) findings that support the use of multiple reference frames for verbal learning [Tversky 1996].

One-way repeated measures ANOVAs revealed no significant differences between the three levels of verbal view-depth used during training and any of the test measures, $ps > 0.05$. Given the small effect sizes (see Table III), the differences between view-depth conditions for the test measures do not appear to be meaningful. The lack of a view-depth affect follows from the results of the training measures and is also in agreement with the findings from our work on verbal learning in real buildings. These data indicate that a local verbal description is sufficient to describe an environment with the VVD and that increasing access to geometric detail does not result in better transfer performance. The next experiment addresses possible issues that could have affected cognitive map development and environmental transfer.

3. EXPERIMENT II: USE OF VERBAL INFORMATION DURING TRAINING AND TEST

The goal of Experiment II was to investigate whether the nonverbal testing paradigm used during the transfer session of Experiment I could account for the lower target localization accuracy observed after learning with the VVD versus learning in real buildings. In our previous work with verbal learning in real buildings [Giudice et al. 2007], both exploration of the training environments and subsequent wayfinding tests were performed with blindfolded sighted subjects using verbal spatial descriptions. In contrast, Experiment I only included verbal descriptions and blindfolded subjects during training with the VVD. Real navigation at test was performed under normal viewing conditions (i.e., with vision). This design was chosen to facilitate a common testing procedure with the subjects who learned visually in Experiment III (see Experiment III Methods). For test performance in Experiment I to be accurate, the spatial representation built up from verbal learning with the VVD must be effectively accessed during vision-based navigation in the real building. The underlying assumption was that once the spatial representation was developed, it could be accessed and read out equivalently from other modalities (e.g., as from a common “spatial image”), [Loomis et al. 2002].

An alternative explanation is that the verbal messages received during training develop into a modality-specific spatial representation in memory. From this perspective, performance of the last experiment would be expected to be poor, relative to experiments that employed a common train/test modality, as participants were required to access a spatial representation at test via a different input from learning. This alternative hypothesis is tested in the current experiment, as both training and testing occur by means of verbal descriptions. If the results of this experiment are reliably better than the first, then we have evidence for the development of modality-specific representations. If there are no

differences between experiments, then we can retain our previous conclusion that the representation is accessible to multiple modalities. This finding would indicate that the poor test performance is not a read-out problem related to route finding during transfer but due to some other factor, such as lack of physical movement during learning with the VVD, which is limiting the development of an accurate cognitive map.

3.1 Method

3.1.1 Participants. Nine participants, five male and four female, between the ages of 18 and 41 (mean = 20.3) ran in the study. The experiment took approximately 1.5 hours and participants received monetary compensation for their time.

3.1.2 Environments and Procedure. The practice, training, and transfer test procedures and environments were the same as Experiment I. Given the lack of a view-depth affect in the previous experiment, participants in this study only trained using the Local verbal mode. Training with the VVD and transfer testing in the real building occurred under blindfold, with verbal descriptions serving as the primary source of environmental information for both. For real building navigation during the transfer test, participants had to find requested target locations, as was done in Experiment I, but this time, they were blindfolded and guided by the experimenter. Their movement was logged in real-time on a laptop computer (Dell Latitude C600, 750MHz) loaded with a virtual rendering of the floor and the VVD. At each intersection, participants heard, via the laptop's speakers, the same verbal information from the VVD that was given during training. After hearing the description, they instructed the experimenter where to walk in order to reach the requested target location.

3.2 Results and Discussion

One-way between subjects ANOVAs, using Type III sums of squares, were conducted to compare the training and test measures for Experiment I (Local) and Experiment II.¹ Note that comparing only the local condition of Experiment I with the single local condition of this study is justified given the identical training procedure and lack of any uncontrolled order effects between the three learning conditions of Experiment I. No reliable differences were observed between the experiments for any of the training measures (Figure 4) or transfer tests (Figure 5), $ps > 0.05$, with the exception of route efficiency, which was significantly higher in Experiment I, $F(1, 19) = 12.84$, $p = 0.002$, $\eta^2 = 0.37$. Since participants in Experiment I had unrestricted viewing of the environment during the transfer test session, the near ceiling route efficiency performance indicates that access to distal visual cues served as an aid during route planning and execution. As Experiment II only afforded proximal verbal access to the environment, subjects could not use this “look-ahead” information to help disambiguate their position. Thus, Experiment II performance likely reflects that subjects had to orient themselves before moving to the destination, which led to travel of slightly less-efficient routes.

The most important finding from this experiment was the similarity of wayfinding performance at test based on verbal descriptions compared to that observed in Experiment I employing vision to perform the same task. These results provide good evidence that the spatial representation built up from learning with the VVD in computer-based environments was equally accessible during transfer to real-world navigation. Since learning with the VVD was common to both experiments and only the testing conditions differed, the finding of poor transfer performance observed in both experiments suggests that the problem was due to deficiencies in the spatial representation built up from verbal learning

¹Note that random assignment was violated in all cross-experiment comparisons. However, we do not believe this was a major problem, since all participants were drawn from a homogeneous population (students taking Introduction to Psychology) and were matched closely for age, gender, and education.

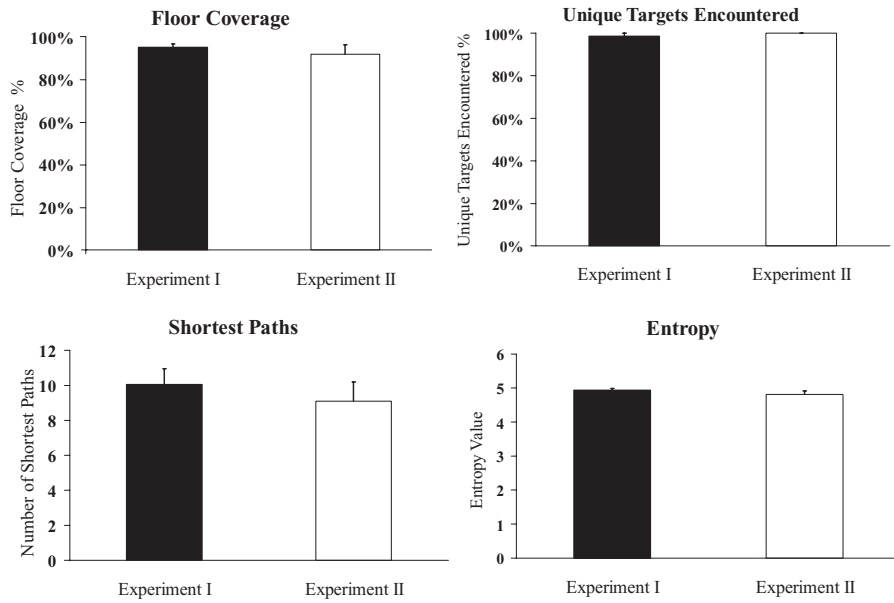


Fig. 4. Comparison of Experiment I and II training measures, collapsed across view-depth. Error bars represent one standard error of the mean.

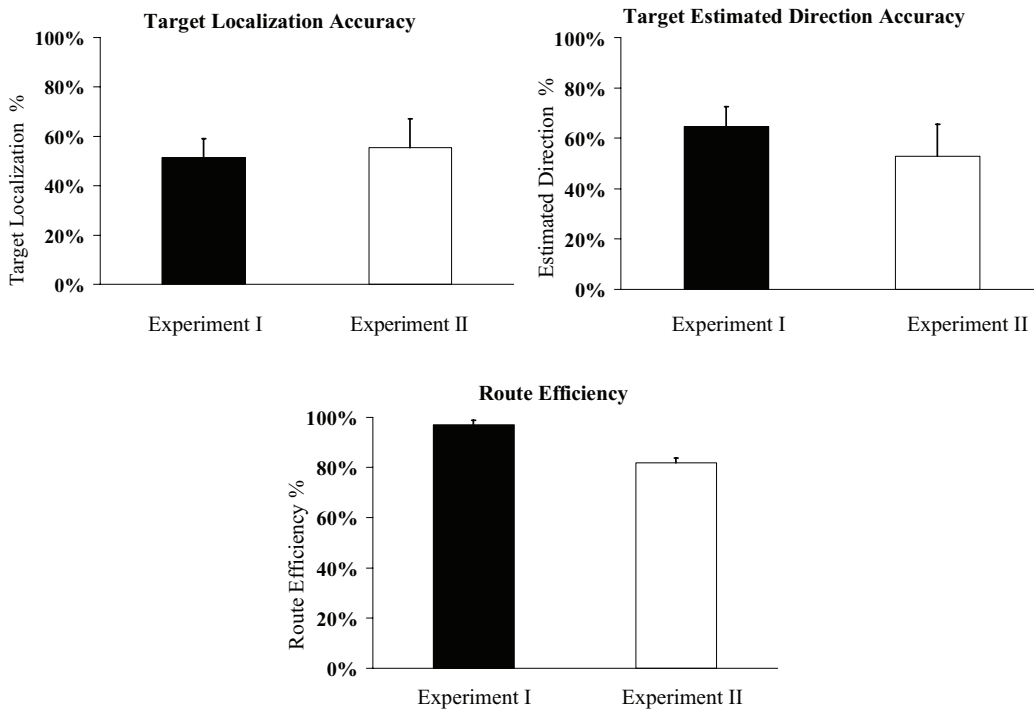


Fig. 5. Comparison of Experiment I and II transfer test measures, collapsed across view-depth. Error bars represent one standard error of the mean.

with the VVD, not with accessing the representation from memory during environmental transfer. The next experiment investigates whether a similar level of transfer performance found after training with the VVD also occurs after training with visually rendered layouts.

4. EXPERIMENT III: VISUAL LEARNING AND ENVIRONMENTAL TRANSFER

This study employs the same environments, view-depth conditions, and procedures as Experiment I. However, rather than learning from a VVD, training occurred in computer-based layouts using an information-matched visually rendered display. By comparing performance between Experiment I (verbal learning) and this experiment (visual learning), we can assess whether training with the VVD yields a similar pattern of search behavior and transfer performance, as is obtained from training in visually rendered VEs.

Although direct comparisons have not been made between verbal and visual learning during wayfinding tasks in computer-based environments, our previous work in real buildings revealed a similar pattern of training and test performance between these learning modalities [Giudice et al. 2007]. These findings are in agreement with a series of studies that found near-equivalent performance within participants on updating target locations during blindfolded walking after learning with spatial language, 3D spatialized sound and vision [Avraamides et al. 2004; Klatzky et al. 2003; Loomis et al. 2002]. Similar transfer performance between the verbal and visual experiments in this article is not considered strong evidence for functional equivalence, as only between subjects comparisons are made, but it provides good evidence that the spatial representations built up from learning in both encoding modalities can be used during wayfinding tasks. A finding of similar transfer results would also suggest that the mediocre target localization performance observed during transfer in the verbal experiments was due to task difficulty, virtual movement, or insufficient learning and not a function of the training modality. Conversely, reliably better transfer performance after visual training would argue against the development of comparable spatial representations and indicate that the VVD, at least in its current incarnation, is not an effective tool for learning VEs.

The geometric detail conveyed by the three visual view-depth conditions used in this study was analogous to the information heard from the three verbal modes used in the VVD. Therefore, a secondary goal of this experiment was to investigate the effect of information availability on exploration and wayfinding behavior of computer-based environments using a visual display. Although few studies have manipulated visual view-depth, there is some evidence to suggest visual learning may be more sensitive than verbal learning to increased spatial integration demands. In a study using similar computer-based corridor layouts to those used here, results showed that reducing the visual view depth from seeing a global view to seeing only a local view of the layout significantly impaired environmental learning time and map reproduction performance [Giudice 2002]. Studies comparing searching of sparse desktop VEs for hidden targets when augmented with a global map or local map have also shown that use of the local map led to poorer development of survey knowledge [Ruddle et al. 1999; Ruddle and Peruch 2004]. These findings suggest that training with the Local visual condition in this study will yield poorer performance compared to the other visual view-depth conditions.

4.1 Method

4.1.1 Participants. Eighteen fully sighted participants, 10 female and 8 male, between the ages of 18 and 40 (mean = 23) ran in the study. The experiment took 4 to 5 hours per subject and was conducted in two sessions.

4.1.2 Environments and Procedure. The environments and three view-depth conditions were identical to those used in Experiment I. However, rather than hearing verbal descriptions of corridor structure,

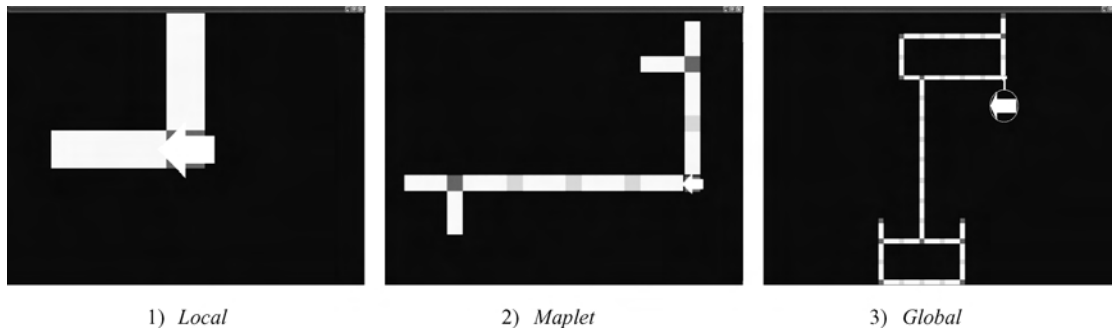


Fig. 6. Examples of visual displays used during the training session of Experiment III. The user's position and orientation in the environment is represented by the white arrow (enlarged in Global for illustrative purposes).

the sparse geometric information was visually displayed on a 17-inch computer monitor. In order to make valid comparisons between experiments, the amount of geometric detail seen from each of the three visual view-depth conditions was matched with what was heard from the three verbal view-depths in the VVD. Equating the spatial information between visual and verbal display modes was challenging, as the VVD provided a verbal description of hallways in all directions at the user's position (affording a panoramic view). However, it is not possible to represent multiple views from a single first-person vantage point in a traditional visually rendered VE (e.g., simultaneously depicting hallways ahead and behind the observer). Our visual display, called a *dynamic digital map*, adopts a modified bird's-eye viewing perspective of the environment, which provides dynamically updated egocentric information about user heading on a stationary north-up map. This type of display has also been dubbed a dynamic you-are-here map [Ruddle et al. 1999]. As with the verbal experiments, the map remained fixed but a visual arrow, representing the observer's first-person ego point, was updated with every translation and rotation. Indicators of cardinal direction (N/S/E/W) were affixed to the monitor to provide a global frame of reference. Coupling information about first-person and environmental reference frames in a visual display has been shown to benefit navigation as it harnesses the advantages of both presentation modes [Aretz 1991]. Figure 6 depicts an example of what was seen by the participant for each view-depth condition.

In the Local condition, participants could only see the corridor segments of the hallway or intersection at their current position. The Maplet condition presented the current intersection and all adjacent intersections. The Global view presented the user with a bird's-eye map of the layout. The Local and Maplet views were displayed in the center of the monitor and were updated with each forward translation (Local) or whenever the user reached an intersection (Maplet). For the global view, the arrow was updated on the map with each key press. As shown in Figure 6, there are changes in the scale of the environments between conditions. This is due to differences in information content between view depths and is not considered problematic. Participants were trained with many examples during the practice session, and pilot data from the first author's dissertation demonstrated that these scale variations did not affect learning. Likewise, the world in miniature (WIM) method [Stoakley and Paush 1995], which shrinks the scale in VEs to allow faster exploration of large worlds, has proven effective for learning the mapping between a miniaturized world and a standard size VE. When WIM was applied with dynamic scaling of VEs, travel performance and orientation did not decline [Wingrave et al. 2006]. However, the effect of scale was not specifically investigated in this study.

The training procedure was identical to the verbal experiments except that the environments were perceived through visual rather than verbal input. All visual training conditions were followed by

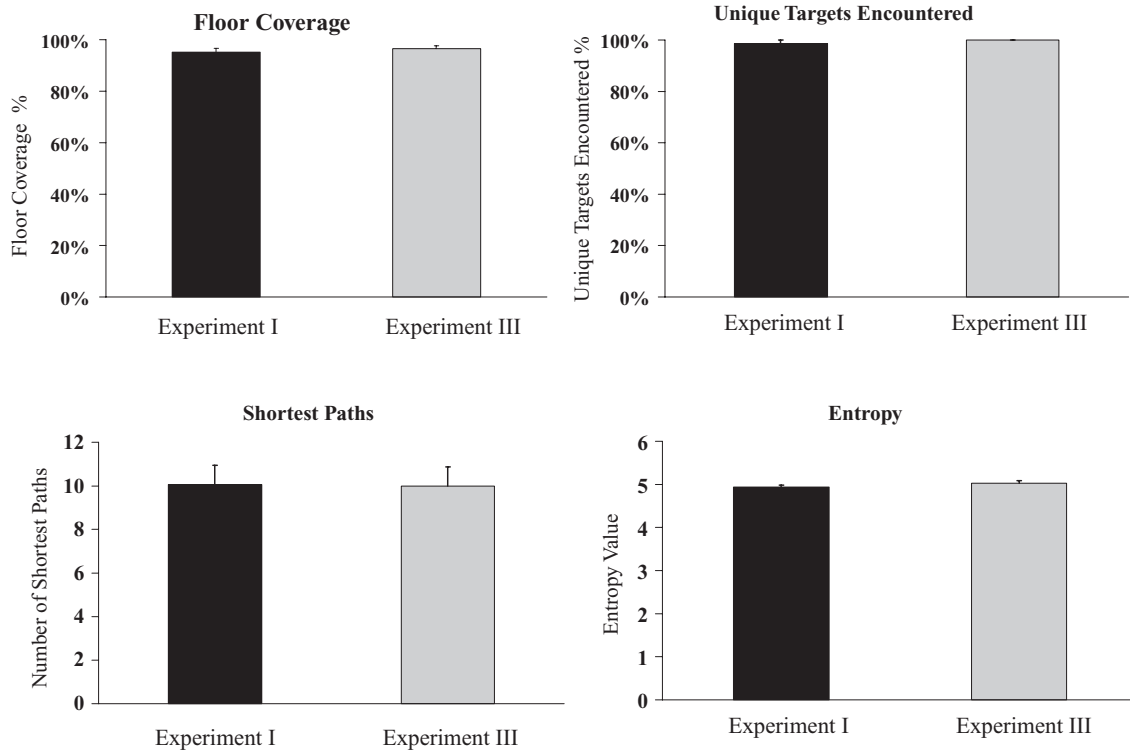


Fig. 7. Comparison of Experiment I and III training measures collapsed across view-depth. Error bars represent one standard error of the mean.

the same environmental transfer tests in the corresponding real building as were used in the verbal experiments.

4.2 Results and Discussion

One-way repeated measures ANOVAs comparing the three visual view-depth conditions for each of the three training measures revealed no reliable differences between conditions, $ps > 0.05$, with all effect sizes, $\eta^2s < 0.08$. This lack of a view-depth effect is in agreement with the findings with verbal learning in Experiment I. Indeed, the overall pattern of visual training performance was very similar to the training results observed with the VVD (see Table III), indicating that a similar search strategy was adopted when learning with both verbal and visual display modes. See Figure 7 for comparison of the training measures between Experiments I and III.

4.2.1 Transfer Session. For this task, participants had to find routes between targets by walking through the real building corresponding to the computer-based training environment. As with Experiment I, this wayfinding task was done with unrestricted vision. One-way repeated measures ANOVAs comparing the three visual view-depth conditions were conducted for each of the three transfer test measures of target localization accuracy, estimated target direction, and route efficiency. In agreement with the results after verbal learning in Experiment I, no significant differences were observed between conditions for any of the test measures, $ps > 0.05$. Contrary to our prediction, these results do

Table IV. Experiments I and III: Contingency Table of Training Experience versus Test Target Localization

| Training experience/target accuracy performance | Correctly localized | Incorrectly localized | Total count training experience |
|---|---------------------|-----------------------|---------------------------------|
| Route traversed at training: | | | |
| <i>Experiment I</i> | 28 | 27 | 55 (44.0%) |
| Experiment III | 66 | 16 | 82 (50.6%) |
| Route not traversed at training: | | | |
| <i>Experiment I</i> | 30 | 40 | 70 (56.0%) |
| Experiment III | 63 | 17 | 80 (49.4%) |
| Total count for target accuracy: | | | |
| <i>Experiment I</i> | 58 (46.4%) | 67 (53.6%) | 125 (100%) |
| Experiment III | 129 (79.6%) | 33 (20.4%) | 162 (100%) |

Note. Data is summed across all participants. The count for routes traversed during training represents only exact matches, the shortest path in the same direction, with the test routes. Only routes with a single shortest path were analyzed.

not support the notion that reduced view-depth will lead to increased errors in visual learning, although training with the Local condition did yield the lowest numeric results and the greatest variability.

Of our three transfer tests, the target localization measure represents the most theoretically interesting test of wayfinding behavior, as it requires planning and executing of novel routes. Thus, we only address this measure when comparing performance after visual (Experiment III) and verbal (Experiment I) learning. A one-way ANOVA was used to compare target localization accuracy performance between Experiments I and III, collapsing across view-depth. Reliable differences were found for target localization accuracy by experiment, $F(1, 35) = 12.92, p < 0.001, \eta^2 = 0.27$, with Experiment I ($M = 65.47\%$, $SE = 5.66\%$) having significantly lower performance than Experiment III ($M = 79.63\%$, $SE = 4.12\%$). These results indicate that the spatial representations built up from verbal and visual learning were not comparable, supporting our earlier interpretation that exploration with the VVD does not lead to a robust cognitive map of the virtual training environments.

4.2.2 Comparing Route Traversal between Training and Test. This final analysis was performed to better characterize the structure of the cognitive map built-up from training with the two computer-based displays. If the spatial representation was route-based, higher target accuracy at test would be predicted for familiar (previously traveled) versus unfamiliar (never experienced) routes. In contrast, if a map-like representation of layout configuration was developed, target accuracy at test should be independent of whether or not the route had been previously traveled. For this analysis, the four test routes traveled between targets for each subject were compared against all routes traveled during their training session. In this way, we could evaluate if the test routes were novel or if they had been previously experienced during training, see Table IV.

Observations of route traversal and target accuracy were not independent within each participant. Thus, we evaluated if there was a relationship between these two variables using a Cochran-Mantel-Haenszel (CMH) chi-square test, stratified by participant. The CMH test revealed no significant general association between route traversal during training and subsequent target localization accuracy for either verbal learning in Experiment I, $\chi^2(1) = 0.06, p = 0.81$, or visual learning in Experiment III, $\chi^2(1) = 0.17, p = 0.68$.

These data demonstrate that after both verbal and visual learning, route-finding ability at test is not contingent on prior experience with the same route. In other words, participants were not simply remembering a sequence of previously executed actions to navigate routes between targets at test but were able to plan and infer novel paths, optimally in most instances, from their cognitive map. These results are particularly noteworthy with respect to Experiment I, as they demonstrate that effective

wayfinding behavior is possible on the basis of virtual verbal learning; despite the inaccuracies of the ensuing spatial representation built up from training with the VVD.

5. GENERAL DISCUSSION

This article described three experiments addressing verbal and visual learning in large-scale VEs. Our first goal was to establish whether virtual exploration of computer-based layouts could be accomplished completely nonvisually using an interface called a virtual verbal display. The VVD proved extremely effective in supporting these tasks. Performance on the training measures with the VVD in Experiments I and II was both highly accurate and remarkably similar to the training performance observed when learning from the visual display in Experiment III. The near-equivalent training behavior between experiments is of theoretical interest, as it demonstrates that exhaustive searching of unfamiliar VEs can be as accurate when hearing dynamically updated verbal descriptions from the VVD, as when seeing the same information visually rendered on a computer monitor.

In addition, training performance with the VVD was almost identical to the results from our previous work, which used the same training measures and verbal modes but with learning in real environments rather than virtual layouts [Giudice et al. 2007]. In their previous study, Giudice et al. also showed that when given an equal number of moves during training in the same environments, a Monte Carlo simulation of a random walk model exhibits reliably lower performance on all training measures than human performance. Thus, we interpret the high level of training performance exhibited with the current verbal experiments as representing efficient search and route-finding behavior, which could not be accounted for by chance decision-making.

The similar pattern of results between the virtual verbal training demonstrated here and verbal learning of real buildings in our previous experiment is also of interest as it suggests that the absence of physical movement with the VVD does not impair search behavior.

A second goal of this research was to investigate whether learning with the VVD transferred to efficient navigation in the physical environment. Transfer performance from the VVD to real-world navigation (Experiment I) was found to be lower than performance on the same tasks in Experiment III after learning with the visual display (averaging 51% vs. 80% target localization accuracy, respectively). The lower wayfinding performance of Experiment I could be attributed to the spatial representation being more difficult to access after verbal than visual training. However, this explanation seems untenable given our previous results with identical verbal descriptions in real buildings, showing that participants averaged 85% accuracy on the same wayfinding task and that this performance did not significantly differ from what was observed after visual learning [Giudice et al. 2007].

Experiment II of the current work investigated whether the development of a sensory-specific spatial representation of the training environment could explain the relatively poor transfer performance of Experiment I. The results demonstrated this was not the case as route-finding at test did not reliably differ between Experiment I (which employed no verbal information during transfer) and Experiment II (which employed identical verbal information during the training and transfer tests). These results indicate the spatial representation built up from verbal learning (i) can be equally accessed by verbal and visual read-out operations at test and (ii) that deficits observed after verbal learning arise from the process of forming an accurate representation from training with the VVD, rather than problems in transferring VE learning to real-world navigation.

A possible explanation relates to the movement behavior with the VVD. Movement with the VVD occurs via the keyboard and thus lacks the proprioceptive and vestibular information associated with real walking. Although the simulated movement behavior did not adversely affect training performance with the VVD or training or transfer performance with the visual display, it may account for the poor transfer performance after virtual verbal learning. The accurate use of verbal descriptions during

real building navigation provides support for the view that physical movement is necessary for the conversion of verbal information into a spatial representation. From this perspective, the absence of body-based information in the VVD during learning may have impaired cognitive map development.

There is some evidence to support this movement hypothesis. In the third experiment of Avraamides et al. [2004], target locations were learned equally well with spatial language and vision. However, mental updating of allocentric target locations learned with spatial language was impaired without physical translation. Updating object locations learned from a text description is also better when the reader physically rotates to adopt the perspective described by De Vega and Rodrigo [2001], with egocentric direction judgments made faster and more accurately after physical, rather than imaginal rotation [Avraamides 2003]. In addition, without appropriate vestibular and proprioceptive cues, people are poor at updating changes in heading, even when optic flow information specifies the rotation [Chance et al. 1998; Klatzky et al. 1998]. Interestingly, recent work by Giudice et al. using a VVD based on the same information as in the Local condition of the current study found no improvement in environmental transfer performance with the addition of real body rotation (Giudice and Tietz 2008). However, significant improvements were found by adding spatialized audio in the VVD, for example, a hallway on the left side was heard in the left ear. These findings, in conjunction with the current results showing highly similar performance between view-depth conditions, suggest that spatialized local information may represent the optimal trade-off between information content and performance. These results are relevant to designers of navigation systems employing dynamically updated geometric-based verbal displays, such as the VVD studied here for indoor navigation, or Sendero Group's accessible GPS for outdoor navigation (see www.senderogroup.com). In addition to the obvious application of this technology to virtual learning and navigation by the blind [Giudice 2004], the VVD could also be employed in real-time navigation systems requiring travel in conditions where visual information is limited or not available (e.g., firefighters operating in smoke-filled environments or for covert nighttime military operations).

The finding that transfer test performance was reliably lower after training with the VVD compared with the visual display indicates that the spatial representations built up from the two learning modalities were not identical. It is possible that increasing the training session with the VVD would improve learning, and the development of a more accurate spatial representation. Several studies have demonstrated that building up functionally equivalent spatial representations requires more trials to reach a learning criterion when learning locations by language versus vision (Avraamides et al. 2004; Klatzky et al. 2002; 2003). However, since the same amount of verbal training was given for learning real building layouts in our previous study, with performance found to be nearly equivalent to visual learning, it seems more likely that deficits stem from problems with the VVD (i.e., lack of physical movement) rather than from insufficient training. Although the current results do not allow us to assess whether a common representation could be built up given better learning parameters, they provide some interesting clues about the underlying spatial representations. For instance, the target localization performance showing that the majority of correctly executed routes were novel and had not been previously experienced in their entirety during training, suggests the development of survey knowledge, even if incomplete. These findings contrast with the traditional landmark, route, survey model of spatial knowledge development [Siegel and White 1975], which posits that routes are learned before a map-like representation is developed. The results also demonstrate that verbal spatial learning is possible in a much broader context than the traditional domain of direction giving and route navigation (e.g., free exploration and wayfinding).

Finally, these studies addressed how manipulating view-depths, or the availability of geometric detail from a given vantage point, effected training and transfer performance. Both verbal and visual exploration was predicted to be most difficult from the minimalist "Local" condition, as it provided the

least information about the environment of all the views and required the most spatial integration. While performance in the Local condition tended to be slightly lower and more variable than the other view-depth conditions, there were no reliable differences between the three view-depths for any of the training or transfer test measures with either modality. The finding that visual transfer performance was reliably better than was observed after training with the VVD for all view-depths speaks to the differences in the representations built up from training, not to a view-depth effect. Indeed, the differences between the three views followed a similar trend for both experiments. Although with much larger sample sizes, we may have found differences in target localization performance by view-depth. Given the effect sizes for Experiment I, $\eta^2 = 0.04$, and Experiment III, $\eta^2 = 0.03$, the contribution of manipulating geometric detail was quite small, indicating that view-depth had a minimal effect on performance. These findings imply that providing local verbal information about layout geometry is sufficient and that the marginal improvement observed by increasing description of geometric detail does not warrant the additional complexity of the expanded (maplet or global) verbal messages. These results are in agreement with the verbal training data from our study in real buildings [Giudice et al. 2007] as well as a study employing restricted visual viewing during training in sparsely rendered VEs [Stankiewicz et al. 2006].

Taken together, our findings clearly demonstrate that VEs can be explored as effectively with a verbal display as with a visual display. The difference in real-world navigation performance between the verbal and visual experiments suggests that the spatial representation built up from virtual verbal learning is less accurate than from visual learning. However, we believe that increased training or the inclusion of spatialized audio in the VVD may well narrow this performance gap.

REFERENCES

- AGUIRRE, G. K., DETRE, J. A., AND ESPOSITO, D. 1996. The parahippocampus sub-serves topographical learning in man. *Cerebral Cortex* 6, 823–829.
- ALLEN, G. L. 1997. From knowledge to words to wayfinding: Issues in the production and comprehension of route directions. Lecture Notes in Computer Science, vol. 1329, S. C. Hirtle and A. U. Frank, Eds. Springer, Berlin, 363–372.
- ALLEN, G. L. 2000. Principles and practices for communicating route knowledge. *Appl. Cognitive Psych.* 14, 333–359.
- ARETZ, A. J. 1991. The design of electronic map displays. *Hum. Factors*, 33, 1, 85–101.
- AVRAAMIDES, M. N. 2003. Spatial updating of environments described in texts. *Cognitive Psych.* 47, 4, 402–431.
- AVRAAMIDES, M. N., LOOMIS, J. M., KLATZKY, R. L., AND GOLLEDGE, R. G. 2004. Functional equivalence of spatial representations derived from vision and language: evidence from allocentric judgments. *J. Exp. Psych.* 30, 4, 804–814.
- BLISS, J. P., TIDWELL, P. D., AND GUEST, M. A. 1997. The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence* 6, 1, 73–86.
- BRYANT, K. J. 1997. Representing space in language and perception. *Mind Lang.* 12, 3, 239–264.
- CHANCE, S. S., GAUNET, F., BEAL, A. C., AND LOOMIS, J. 1998. Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence* 7, 2, 168–178.
- CLAWSON, D. M., MILLER, M. S., KNOTT, B. A., AND SEBRECHTS, M. M. 1998. Navigational training in virtual and real buildings. In *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society*. Springer, Berlin, 136–166.
- DE VEGA, M. AND RODRIGO, M. J. 2001. Updating spatial layouts mediated by pointing and labeling under physical and imaginary rotation. *Euro. J. Cognitive Psych.* 13, 369–393.
- DENIS, M. AND COUDE, M. 1989. Scanning visual images generated from verbal descriptions. *Euro. J. Cognitive Psych.* 1, 293–307.
- DENIS, M. AND ZIMMER, H. D. 1992. Analog properties of cognitive maps constructed from verbal descriptions. *Psych. Res.* 54, 286–298.
- FERGUSON, E. L. AND HEGARTY, M. 1994. Properties of cognitive maps constructed from texts. *Memory Cognition* 22, 4, 455–473.
- GIUDICE, N. 2004. Navigating novel environments: A comparison of verbal and visual learning. Unpublished Dissertation, University of Minnesota, MN.
- GIUDICE, N. A., BAKDASH, J. Z., AND LEGGE, G. E. 2007. Wayfinding with words: spatial learning and navigation using dynamically updated verbal descriptions. *Psych. Res.* 71, 347–358.

- GIUDICE, N. A., MASON, S., AND LEGGE, G. E. 2002. The relation of vision and touch: Spatial learning of small-scale layouts. *J. Vision* 2, 7, 522a.
- GIUDICE, N. A., AND TIETZ, J. 2008. Learning with virtual verbal displays: effects of interface fidelity on cognitive map development. In C. Freksa, N. Newcombe, P. Gärdenfors, S. Wöfl Eds., *Spatial Cognition VI*, Lecture Notes in Artificial Intelligence, Springer, Berlin.
- JOHNSON-LAIRD, P. N. 1983. *Mental Models*. Harvard University Press, Cambridge, MA.
- KALIA, A., LEGGE, G. E., AND GIUDICE, N. A. 2008. Learning building layouts with non-geometric visual information: The effects of visual impairment and age. *Perception* 33, 11, 1677–1699.
- KLATZKY, R. L., LIPPA, Y., LOOMIS, J. M., AND GOLLEDGE, R. G. 2002. Learning directions of objects specified by vision, spatial audition, or auditory spatial language. *Learn. Memory* 9, 364–367.
- KLATZKY, R. L., LIPPA, Y., LOOMIS, J. M., AND GOLLEDGE, R. G. 2003. Encoding, learning, and spatial updating of multiple object locations specified by 3D sound, spatial language, and vision. *Exp. Brain Res.* 149, 1, 48–61.
- KLATZKY, R. L., LOOMIS, J. M., BEALL, A. C., CHANCE, S. S., AND GOLLEDGE, R. G. 1998. Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psych. Sci.* 9, 4, 293–299.
- LOOMIS, J. M., BLASCOVICH, J. J., AND BEAL, A. C. 1999. Immersive virtual environment technology as a basic research tool in psychology. *Behav. Res. Methods Instrum. Comput.* 31, 4, 557–564.
- LOOMIS, J. M., HERBERT, C., AND CINCINELLI, J. G. 1990. Active localization of virtual sounds. *J. Acoust. Soc. Am.* 88, 4, 1757–1764.
- LOOMIS, J. M., GOLLEDGE, R. G., AND KLATZKY, R. L. 1998. Navigation system for the blind: Auditory display modes and guidance. *Presence* 7, 193–203.
- LOOMIS, J. M., GOLLEDGE, R. G., AND KLATZKY, R. L. 2001. GPS-based navigation systems for the visually impaired. In *Fundamentals of Wearable Computers and Augmented Reality*, W. Barfield and T. CAUDELL, Eds. Erlbaum, Mahwah, NJ, 429–446
- LOOMIS, J. M., LIPPA, Y., GOLLEDGE, R. G., AND KLATZKY, R. L. 2002. Spatial updating of locations specified by 3D sound and spatial language. *J. Exp. Psych.* 28, 2, 335–345.
- LOOMIS, J. M., MARSTON, J. R., GOLLEDGE, R. G., AND KLATZKY, R. L. 2005. Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. *J. Visual Impairment Blindness* 99, 219–232.
- LOVELACE, K., HEGARTY, M., AND MONTELLO, D. 1999. Elements of good route directions in familiar and unfamiliar environments. In *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science*, C. Freksa and D. M. Mark, Eds., Springer, Berlin, 65–82.
- MARSTON, J. R., LOOMIS, J. M., KLATZKY, R. L., GOLLEDGE, R. G., AND SMITH, E. L. 2006. Evaluation of spatial displays for navigation without sight. *ACM Trans. Appl. Percept.* 3, 2, 110–124.
- PERRIG, W. AND KINTSCH, W. 1985. Propositional and situational representations of text. *J. Memory Lang.* 24, 503–518.
- PÉRUCH, P., GAUNET, F., THINUS-BLANC, C., AND LOOMIS, J. 2000. Understanding and learning virtual spaces. In *Cognitive Mapping: Past, Present and Future*, R. Kitchin and S. Freundschuh, Eds., Routledge, London, 108–124.
- PETRIE, H., JOHNSON, V., STROTHOTTE, T., RAAB, A., FRITZ, S., AND MICHEL, R. 1996. Mobic: designing a travel aid for blind and elderly people. *J. Navig.* 49, 45–52.
- RICHARDSON, A. E., MONTELLO, D. R., AND HEGARTY, M. 1999. Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory Cognition* 27, 4, 741–750.
- RUDDLE, R. A. AND LESSELS, S. 2006. For efficient navigational search, humans require full physical movement, but not a rich visual scene. *Psych. Sci.* 17, 6, 460–465.
- RUDDLE, R. A., PAYNE, S. J., AND JONES, D. M. 1997. Navigating buildings in “desk-top” virtual environments: Experimental investigations using extended navigational experience. *J. Exp. Psych.* 3, 2, 143–159.
- RUDDLE, R. A., PAYNE, S. J., AND JONES, D. M. 1999. The effects of maps on navigation and search strategies in very-large-scale virtual environments. *J. Exp. Psych.* 5, 1, 54–75.
- RUDDLE, R. A. AND PÉRUCH, P. 2004. Effects of proprioceptive feedback and environmental characteristics on spatial learning in virtual environments. *Int. J. Hum. Comput. Stud.* 60, 299–326.
- SCHLICHT, E. J., LEGGE, G. E., STANKIEWICZ, B. J., AND GIUDICE, N. A. 2001. Are visual landmarks necessary for effective transfer of navigational knowledge between real and virtual buildings. In *Proceedings of the Annual Meeting of the Association of Research in Vision and Ophthalmology*.
- SIEGEL, A. AND WHITE, S. 1975. The development of spatial representation of large scale environments. In *Advances in Child Development and Behavior*, H. Reese, Ed., Academic Press, New York.
- STANKIEWICZ, B., LEGGE, G. E., MANSFIELD, J. S., AND SCHLICHT, E. J. 2006. Lost in virtual space: Studies in human and ideal spatial navigation. *J. Exp. Psych. Human Percept. Perform.* 32, 688–704.

- STOAKLEY, C. AND PAUSCH, R. 1995. Virtual Reality on a WIM: Interactive worlds in miniature. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, 265–272.
- TAYLOR, H. A. AND TVERSKY, B. 1992. Spatial mental models derived from survey and route descriptions. *J. Memory Lang.* 31, 261–292.
- TOM, A. AND DENIS, M. 2003. Referring to landmark or street information in route directions: what difference does it make? In *Spatial Information Theory: Foundations of Geographic Information Science*, W. Kuhn, M. F. Worboys, and S. Timpf, Eds., Springer, Berlin, 384–397.
- TVERSKY, B. 1996. Spatial perspective in descriptions. In *Language, Speech, and Communication*, P. Bloom, M. A. Peterson, L. Nadel and M. F. Garrett, Eds., MIT Press, Cambridge, MA.
- WALLER, D., HUNT, E., AND KNAPP, D. 1998. The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperators Virtual Environ.* 7, 129–143.
- WALLER, D., KNAPP, D., AND HUNT, E. 2001. Spatial representations of virtual mazes: The role of visual fidelity and individual differences. *Hum. Factors* 43, 1, 147–158.
- WILSON, P. N., FOREMAN, N., GILLET, R., AND STANTON, D. 1997. Active versus passive processing of spatial information in a computer-simulated environment. *Ecol. Psych.* 9, 3, 207–222.
- WINGRAVE, C. A., HACHIAHMETOGLU, Y., AND BOWMAN, D. 2006. Overcoming world in miniature limitations by a scaled and scrolling WIM. In *Proceedings of the IEEE Symposium on 3D User Interfaces*. IEEE, Los Alamitos, CA, 11–16.
- WITMER, B. G., BAILEY, J. H., AND KNERR, B. W. 1996. Virtual spaces and real world places: Transfer of route knowledge. *Int. J. Hum. Comput. Stud.* 45, 413–428.

Received November 2008; accepted November 2008