Published in: Displays 35 (2014) 110–117, http://dx.doi.org/10.1016/j.displa.2014.04.001

Navigating virtual mazes: the benefits of audiovisual landmarks

Peter Werkhoven[1], Jan B.F. van Erp[2] and Tom G. Philippi[1]

[1] Utrecht University, Department of Information and Computing Sciences, Padualaan 14, 3584 CH, Utrecht, The Netherlands.
[2] TNO, Department Perceptual and Cognitive Systems, Kampweg 5, 3769 DE, Soesterberg, The Netherlands.

Corresponding Author
Peter Werkhoven
Utrecht University
P.O. Box 80.089
3508TB Utrecht
The Netherlands
Phone: +31 6 53979658
Email: peter.werkhoven@tno.nl

Abstract

It has been shown that multisensory presentation can improve perception, attention, and object memory compared with unisensory presentation. Consequently, we expect that multisensory presentation of landmarks can improve spatial memory and navigation.

In this study we tested the effect of visual, auditory and combined landmark presentations in virtual mazes on spatial memory and spatial navigation. Nineteen participants explored four different virtual mazes consisting of nodes with landmarks and corridors connecting them. Each maze was explored for 90 seconds. After each exploration, participants performed the following tasks in fixed order: 1) draw a map of the maze, 2) recall adjacent landmarks for three given landmarks, 3) place all landmarks on the map of the maze, and 4) find their way through the maze to locate five given landmarks in fixed order.

Our study shows significant effects of multisensory versus unisensory landmarks for the maze drawing task, the adjacency task, and the wayfinding task. Our results suggest that audiovisual landmark presentations improve spatial memory and spatial navigation performance in virtual environments.

1 Introduction

Spatial navigation is the process of planning and following routes to travel from the current location to a target location, and involves one of the most fundamental human cognitive functions (way finding) and motoric functions (locomotion).

A good understanding of human way finding strategies [1,2,3] and underlying perceptual and cognitive processes [1] is of particular importance to optimally support human spatial navigation in virtual environments. Virtual environments can be more complex than natural environments with respect to their scale [4,2], structural complexity and dimensionality [5]. Furthermore, virtual environments often differ from natural environments with respect to the sensory modalities involved, depending on the use of visual, auditory and vestibular displays [6,7,8,9,10,11] and interaction methods [8]. Thus, an optimal support of human spatial navigation in virtual environments relies on knowing the effects of involving multiple sensory modalities.

Finding one's way in complex environments relies on complex multisensory perceptual processes and on working memory and spatial memory. When no external representations are available (such as a map), spatial navigation is based on internal mental representations (cognitive maps) derived from sensory experiences [12,13,14,15].

Such mental representations, called cognitive maps, are constructed by perceiving spatial information from multiple sensory cues, by understanding the spatial relationships between important attributes of the environment (i.e. landmarks and their relative positions) and by encoding this (multi)sensory information into internal spatial representations in short- and long-term memory.

Cognitive maps in combination with self-motion cues are crucial to continuously maintain a sense of position and orientation and to guide navigational behavior [15]. To optimize navigation performance the various sensory inputs (e.g. visual, auditory, tactile and vestibular) in underlying perceptual, memory and cognitive processes should be optimally combined or integrated.

1.1 Multisensory perception of landmarks

In the acquisition of spatial memory contents, landmarks play an important role [16,17]. Landmarks are typically distinctive objects that stand out in the environment [18,19] and serve as reference points when we are following routes or when we need to determine where we are [4].

When navigating the real world, landmarks can be seen (e.g. buildings), heard (e.g. clock towers) and perceived in their combinations. Although our study focusses on audiovisual environments, it is of interest that multisensory navigation is likely to also involve the olfactory and tactile sensory systems. It has been known for a longer time that ants can use olfactory information in order to locate their nest entrance [20] and that pigeons use gradients of volatile organic compounds in the atmosphere for environmental odor-based navigation [21]. But it has only recently been shown that humans have a residual directional smelling ability and process spatial information in the olfactory system [22]. The involvement of the somatosensory system in navigational processes has been studied by Restat et al. [23] who showed that experiencing slanted slopes in a virtual town positively influences navigation performance. Navigation studies in real towns are lacking probably due to the difficulty of maintaining these natural environments controlled.

It has been suggested that the integration of the neural responses to multisensory stimuli into coherent and meaningful representations of landmarks can yield 'superadditivity' effects, meaning that the multisensory response is greater than the sum of its unisensory parts [24].
Research on perceptual integration indicates that multisensory stimuli are generally beneficial for perceptual task performance [25,26]. More specifically, multisensory perception can improve *reaction time* [27,28,29,30], improve *stimulus detection* [31], and reduce signal *variability* [32,26,33,34,35].

1.2 Multisensory memory of landmarks

Memory research has recently shown that multisensory experiences enhance recall and recognition of *object identity* [36,37] and *object location* [38].
A well accepted Working Memory model is that of Baddeley and colleagues [39,40,41] which postulates three components: a visuo-spatial sketchpad (for objects and spatial information), a phonological or articulatory loop (for storing auditory and verbal information) and a central executive (to coordinate the systems and bind and manipulate information). The visual-spatial sketchpad and the phonological loop are independent [42], but the central executive relies on shared (attentional) resources. This means that integrating multisensory information in working memory may come at a cost when the central executive has to compete for resources with other information processes. Later Baddeley [43,44] added a fourth component, the episodic buffer, which can be seen as the gateway to long term memory [45], amongst others to explain effects of chunking or binding.
The multisensory nature of encoding and decoding information suggests that a multisensory presentation of landmarks may improve the short as well as long term memory processes underlying navigational performance.

1.3 Multisensory navigation

Enhanced perceptual and memory processes may further improve subsequent multisensory navigation processes. We focus on audiovisual environments.
Visual spatial knowledge acquisition and navigation has been studied extensively in natural and virtual environments [46,47,48,49]. Furthermore, case studies show that humans are also able to navigate in virtual acoustic environments based on (3D) auditory cues and beacon sounds [50,51,52,53].
However, fewer studies have investigated the effect of audiovisual presentation on navigation performance [54,55]. Gunther, Kazman, & MacGregor [54] investigated navigation in virtual environments that contained visual objects that could also produce 3D sound. They found that the addition of 3D sound to the virtual environments improved navigation, but did not increase a participant's spatial memory of the environment. However, the authors explained that the sound was audible even when participants did not look at the objects (i.e., it was audible through the walls) and that this worked as a beacon to guide participants to their destination. Because the sound was often experienced without the corresponding visual object, multisensory interactions may have played no role in this study. Another study on the effects of multisensory presentation in virtual environments was conducted by Ardito, Costabile, De Angeli and Pittarello [55]. They presented participants with different pieces of classical music in each room of a virtual museum and found that such music could benefit users' navigation and memory

performance, but only when users where informed in advance of the link between the music and the rooms. The authors suggest that for stimuli with a ‘natural’ link (semantic stimuli) the benefit on navigation and memory performance may be automatic. This suggestion is in agreement with the recent studies indicating that meaning plays a critical role in multisensory memory interactions [38,56,57,58].

In conclusion, these studies have suggested, but not conclusively shown, that multisensory object representations can indeed enhance navigation performance. However, the explicit relation between spatial memory and navigation performance has not yet been investigated.

1.4 Does audiovisual perception improve spatial memory and consequently navigation performance?

We wanted to investigate the effects of meaningful multisensory landmarks on spatial navigation performance in relation to spatial memory performance. For this purpose we constructed several virtual mazes using the game *Unreal Tournament 2004.* The mazes either contained auditory, visual or audiovisual landmarks. The landmarks were sounds and pictures of meaningful semantic objects. We experimentally investigated the effect of multisensory landmark presentation on the user’s spatial memory and on navigation performance. Our first hypothesis was that multisensory (audiovisual) landmark presentation improves spatial memory of the virtual mazes compared with unisensory presentation. Our second hypothesis was that audiovisual landmark presentation improves navigation performance compared with unisensory presentation.

2 Method

2.1 Participants

Nineteen (10 male, 9 female) students with ages between 19 and 28 (mean age = 22.2; SD = 2.14) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Ten participants did not regularly play any three-dimensional computer games. The other participants either played such games monthly (3), weekly (5) or daily (2). All participants gave their informed consent prior to participation and completed the experiment in approximately 50 minutes. Participants were paid for their participation.

2.2 Experimental setup

Participants were seated approximately 70 cm from an Iiyama 24 inch LCD monitor. The monitor displayed the first person view of a virtual maze with a resolution of 1680 (H) by 1050 (V) pixels. Participants navigated the virtual maze using a keyboard which was positioned in front of the monitor. They could press the arrow keys to move forward or backward, and to turn left or right. When participants navigated the virtual maze they wore headphones (Sennheiser HD150).

The virtual mazes were rendered in *Unreal Tournament 2004* (Atari, New York City). Participants had a field of view of approximately 90 (H) by 56 (V) degrees which was located about 1.75 m (or 88 Unreal Units [UU]) above ground-level.

Movement speed was roughly 9 m/s (or 480 UU/s) and rotation speed was 150 degrees per second.
For this experiment, four different mazes were constructed with varying geometry and topology using the level editor of *Unreal Tournament 2004*. Each maze consisted of 10 nodes connected by 13 corridors. The nodes were geometric vertical cylinders with a radius of 10 m (512 UU) and a height of 5 m (256 UU). A white cube (2.5 m; 128 UU) was positioned in the center of each node. These cubes were used for landmark presentation. The corridors were bars between nodes with a width and height of 5 m (256 UU) and lengths varying between 16 m (800 UU) and 60 m (3000 UU). The corridors were placed in such a way to prevent participants observing multiple cubes in a single screen. A topside view of each maze is presented in Figure 1. An impression of the participant's view of a maze is shown in Figure 2.

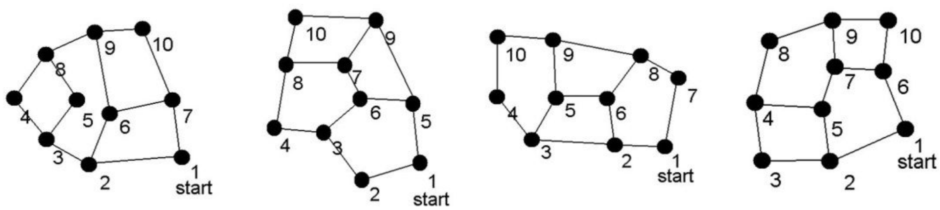

Figure 1: Top-down view of the four mazes. From left to right: maze 1, maze 2, maze 3, and maze 4. Each maze had 10 nodes and 13 corridors. The nodes are numbered 1 to 10.

2.3 Landmark stimuli
Each maze contained ten landmarks which were presented either auditorially, visually or audiovisually originating from the white cubes located in the center of each node. The identity of each landmark presented was fixed (see Table 1).
Auditory landmarks were sounds (66 dB[A]) matching the identity of the line drawing of the visual landmarks. The sounds were gathered from the internet and were all modified to a duration of 3.5 seconds. *Unreal Tournament 2004* automatically modulated the stereophonic components of the sounds to create the illusion that the sounds emanated from the white cubes (which were present with blank sides when used for auditory only landmark presentation). Sounds started playing when the corresponding white cube was present in the participant's field of view. Once started, a sound continued playing until the end of its duration. When a white cube was still (or again) in a participant's field of view after a sound finished playing, that sound was played again after an interval of 0.5 seconds.
An audiovisual landmark consisted of a line drawing and a matching sound presented simultaneously (when within the participant's visual field) and similar to the auditory condition.

Table 1 The landmarks presented in each maze				
Node	Maze 1	Maze 2	Maze 3	Maze 4
1	Cannon (AV)	Whistle	Donkey	Bird
2	Camera (A)	Saw	Train	Pistol
3	Church (AV)	Fly	Scissors	Helicopter
4	Tree (V)	Pencil	Frog	Cat
5	Umbrella (V)	Telephone	Apple	Bell
6	Pig (AV)	Clock	Bicycle	Accordion
7	Rubber Duck (A)	Cow	Airplane	Horse
8	Music Box (A)	Drum	Trumpet	Piano
9	Ball (AV)	Car	Guitar	Seal
10	Grasshopper (V)	Duck	Chicken	Toothbrush

Table 1: This table lists the landmarks presented at each node (rows) in the four mazes (columns). The node numbers correspond to those in Figure 1. Of the landmarks in the first maze, three were always presented visually (V), three auditorilly (A) and four audiovisually (AV). The landmarks in the other mazes could either be presented visually, auditorilly or audiovisually.

Visual landmarks were line drawings selected from the Snodgrass and Vanderwaart set [59]. They were presented on each vertical side of a white cube (see Figure 2 for a list of the items and the appendix for the images).

Figure 2: A screenshot from the virtual maze. In this instance, the user is at the first node in the third maze. The white cube located at this node shows the line drawing of a donkey. If the user was exploring this maze and it contained audiovisual landmarks then he or she would also hear the donkey bray when he or she looked at the cube.

2.4 Experimental design

The experiment consisted of one familiarization run and three experimental runs. In each run, participants had to explore one of the four mazes and were asked to

complete four tasks. Each maze contained 10 landmarks. All participants started with maze 1, which contained a balance of visual, auditory and audiovisual landmarks (see table 1) and was only used to familiarize participants with the experimental procedure.
Thereafter, Mazes 2 to 4 were completed in fixed order. Each of these mazes either contained ten visual, ten auditory, or ten audiovisual landmarks (see table 1). Each participant completed one maze containing only visual landmarks (the visual condition), one maze containing only auditory landmarks (the auditory condition) and one maze containing audiovisual landmarks (the audiovisual condition). The order of which condition was presented in which maze was balanced across participants.

2.5 Procedure and tasks

At the start of the experiment, participants were given written instructions explaining the experimental procedure. In each maze, participants were inserted at the first node (see Figure 1) and had 90 seconds to explore the full maze. For comparison, in each maze it took the experimenter – who was familiar navigating these mazes – about 45 seconds to visit each node and about 70 seconds to walk through each corridor at least once. The exploration time was set at 90 seconds because pilot studies indicated that when exploration time was longer than 90 seconds participants often achieved perfect scores in some of the subsequent spatial memory tasks and when exploration time was less than 90 seconds participants often did not manage to visit all the nodes in the maze.
Participants were told in advance in which sensory modalities the landmarks would be presented in that maze and were encouraged to visit each node in the maze at least once. During the exploration of each maze, participants also had to verbally repeat the letter sequence ‘a-b-c’. This articulatory suppression task was used to prevent verbal recoding of the spatial features of the maze and the identity of the landmarks [36].
After each exploration participants had to complete four tasks in fixed order. This order was chosen because in any other order participants would have to report information that was presented in a previous task. Participants had a fixed amount of time to complete each of the first three tasks. When one of these three tasks was completed early, participants had to wait until the time for the task expired.
The tasks were adopted from earlier studies exploring spatial memory and navigation [60,61,62]. A pilot study showed that they were sensitive to landmark presentation modality. The first three tasks were used to assess spatial memory, while navigation performance was measured with the fourth task.

- *Task 1*: Maze drawing task. In this first task participants were given an empty sheet of paper and asked to draw a map of the maze they had just explored. In addition to drawing the nodes and the connections they had to indicate which landmark was present at which node by writing down the name of the drawing or the sound of that landmark near each node. Participants had 90 seconds to complete this task. Task 1 reveals the spatial characteristics of the mental representation of the maze.

- *Task 2*: Landmark adjacency task. In this second task participants were given a sheet of paper containing a list of three landmarks from the maze they had explored. For each landmark, they had to write down all the landmarks which were directly connected with it (i.e. through a single corridor). For example, the landmark *tree* is directly connected with the landmarks *music box* and *church* (see Figure 1 and Table 1). The landmarks on the lists were not directly connected with each other. They were *tree*, *camera*, and *rubber duck* for maze 1; *saw*, *cow*, and *telephone*, for maze 2; *train*, *apple*, and *frog* for maze 3; and *bell*, *accordion*, and *piano* for maze 4. Participants were given 60 seconds to complete this task. Task 2 tests if participants have memorized information on the adjacency of nodes even when they have no spatial representation of the maze (e.g. by memorizing 'lists of connections between pairs of landmarks').
- *Task 3*: Landmark placement task. In this third task participants were given a sheet of paper with the actual top-down view of the maze (as in Figure 1) and a list of all the landmarks encountered in that maze. Here, they had to place each landmark on the correct node of the maze. Participants had 90 seconds for this task. Task 3 tests object location memory.
- *Task 4*: Way finding task. In this fourth task participants had to find their way in the virtual maze. They were inserted in the maze and got the instruction to navigate to another node with the shortest possible route in the fastest possible time. On insertion, the experiment leader announced to which node the participant had to go. The landmarks they visited were, in fixed order: *pencil* (starting point), *duck*, *clock, whistle, fly,* and *car* for maze 2; *chicken* (starting point) *scissors*, *trumpet, donkey, bicycle* and *frog* for maze 3; and *piano* (starting point), *helicopter*, *horse*, *bird*, *toothbrush,* and *cat* for maze 4. Once the participant arrived at the appropriate node, the procedure was repeated until the participant had located a total of five nodes. In each maze, participants were dropped opposite of the starting node of the exploration phase. Navigating to the target node required them to traverse at least one or two other nodes. The paths they had to walk in each maze all had about similar patterns and distances. There was no time limit for this task. Task 4 tests navigation performance (route length and travel time).

2.6 Statistical analyses

The effect of the experimental condition on the performance during exploration and in all four tasks was investigated by employing separate repeated measures ANOVAs on the data for each task. Post-hoc Fisher LSD tests were conducted where necessary. The data acquired in the training condition were not analyzed.

3 Results

3.1 Exploration phase

Each landmark was observed (i.e. looked at and/or heard) at least once in nearly all explorations. On average, participants encountered a landmark 23.2 times (SD = 2.8) during each exploration. In 3.3 (SD = 1.9) of these encounters on average participants saw and/or heard the same landmark twice or more without seeing and/or hearing another landmark in between. Furthermore, they visited a landmark (i.e. they entered the cylindrical node area of each landmark) 15.0 times (SD = 2.2) during each exploration. So, when a participant saw and/or heard a landmark, he or she did not always move into that landmark's node area.

The number of landmarks observed, the number of times a participant observed a landmark twice in a row, and the number of times they visited a node did not differ between experimental conditions ($F_{2,36} = 1.3$; $p = .28$, $F_{2,36} = 0.2$; $p = .82$, and $F_{2,36} = 2.79$; $p = .07$, respectively).

3.2 Maze drawing (Task 1)

The maze drawing task was scored for two variables:

- *Number of landmarks*: When a landmark's name was written on a node on the map it was considered to be correctly recalled, even when it was the wrong node (in contrast to Task 3).
- *Number of corridors between the nodes of the recalled landmarks*: A corridor in the drawing of the maze was considered correct if 1) this corridor connected two recalled landmarks, and 2) these landmarks were connected by a corridor in the maze explored.

In other words, the relative location of the landmarks and corridors drawn was not taken into account for scoring. Furthermore, no penalties were applied when a landmark or corridor drawn was not present in the maze explored.

The scores for the three experimental conditions are displayed in Figure 3. We found a significant effect of Experimental Condition (3) on the number of recalled landmarks ($F_{2,36} = 6.4$; $p < .001$) and on the number of drawn corridors ($F_{2,36} = 8.2$; $p < .001$). Both the number of recalled landmarks and the number of drawn corridors differed between the audiovisual (AV) condition and the auditory (A) and visual (V) conditions (all at least $p < .05$), but not between the A and the V condition ($p = .83$ and $p = .94$ respectively).

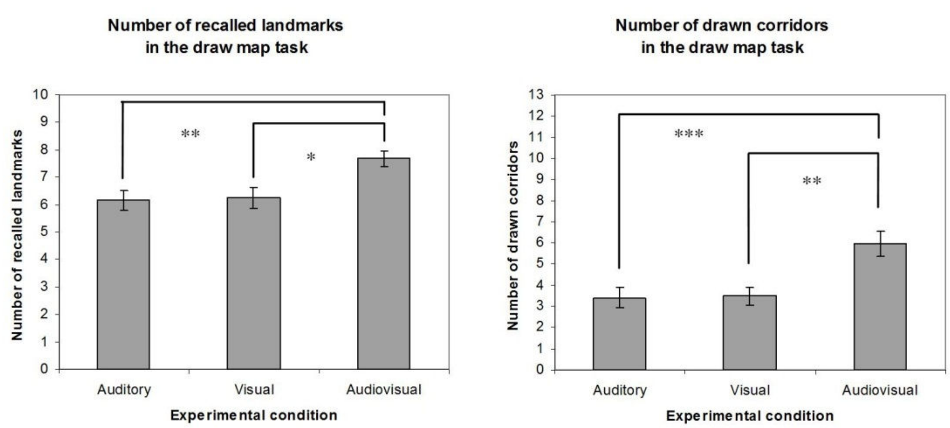

Figure 3: Performance in the maze drawing task (Task 1) for the three experimental conditions. Each panel displays performance for the auditory, visual, and audiovisual conditions. In the left panel, performance was scored for the number of recalled landmarks. In the right panel, performance was scored for the number of drawn corridors. Significant differences are indicated by

asterisks; one, two and three asterisks denote a significance level of respectively $p < .05$, $p < .01$ and $p < .001$.

3.3 Landmark adjacency (Task 2)

The number of recalled adjacent landmarks for each condition is displayed in Figure 4. We found a significant effect of Experimental Condition (3) ($F_{2,36} = 4.2$; $p < .05$). The number of adjacent landmarks differed only for the AV and A ($p < .01$) condition and not for the AV and V ($p = .13$) and the A and V conditions ($p = .19$).

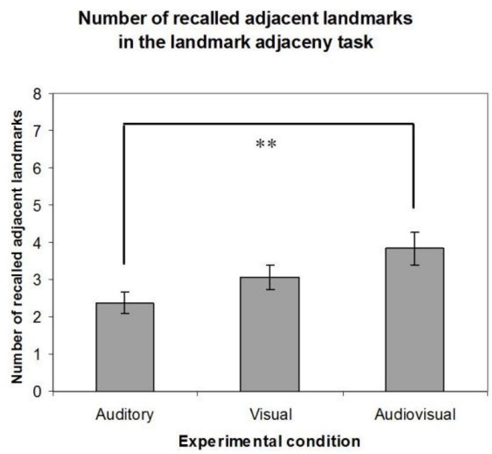

Figure 4: Performance for the landmark adjacency task (Task 2) as a function of experimental condition. Two asterisks denote significant differences of $p < .01$.

3.4 Landmark placement (Task 3)

Participants correctly placed 4.78 (SD = 0.39), 5.42 (SD = 0.50), and 5.05 landmarks (SD = 0.56) on the correct node in the A, V, and AV conditions, respectively. There was no significant effect of Experimental Condition ($F_{2,36} = 0.4$; $p = .66$).

3.5 Way finding (Task 4)

In the navigation task participants had to find their way to five landmarks in the maze. The optimal route to visit those five landmarks required a participant to pass (or visit) twelve nodes. With a map, it took the experimenter approximately 60 s to navigate this route. For one or more experimental conditions, three participants required more than the average time plus three times the standard deviation to find their way to the five landmarks. These participants were considered outliers and all data from these three participants were excluded from this analysis. The time it took the remainder of the participants to find the five target landmarks and the number of nodes visited by participants is displayed in Figure 5. We found a significant effect of Experimental Condition (3) for the number of visited landmarks

($F_{2,30} = 7.5$; $p < .01$) and the time taken ($F_{2,30} = 9.1$; $p < .001$). For the number of visited landmarks, AV differed from A ($p < .01$) and V ($p < .05$), but A did not differ from V ($p = .09$). For the time taken, AV differed from A ($p <.001$) and V ($p <.01$), but A did not differ from V ($p = .11$).
A correlation analysis between the average time taken and the average number of visited landmarks per participant revealed a strong correlation between these measures (*r = 0.91; p < .001*).

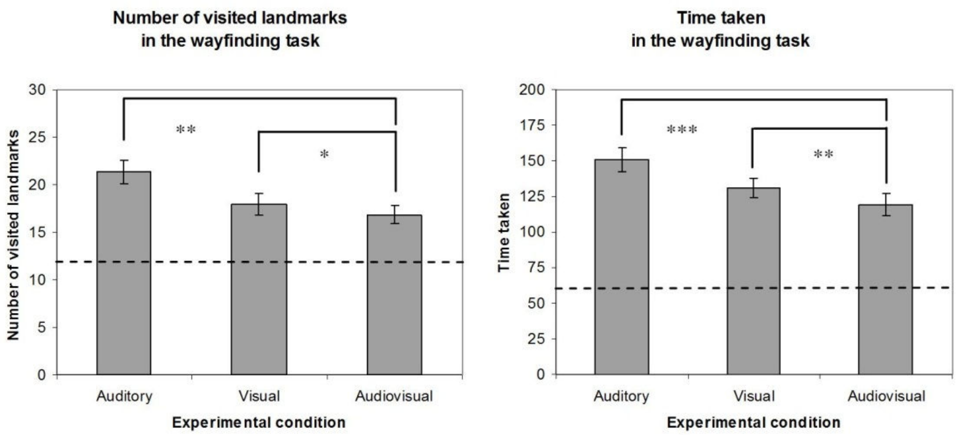

Figure 5: Way finding performance (Task 4) as a function of experimental condition. Performance was scored for the number of landmarks visited (left panel) and the time taken (right panel) to find all the five targets. Significant differences are indicated by asterisks: a single asterisk denotes significance level of $p < .05$, two asterisks denote a significance level of $p < .01$, and three asterisks of $p < .001$. The dashed lines indicate optimal performance.

4 Discussion

We investigated the effects on spatial memory and navigation performance of presenting visual and/or auditory landmark information in virtual environments. Spatial memory performance was measured with a maze drawing task, an adjacency task and a landmark placement task while navigation performance was measured with a way finding task. In two of the three spatial memory tasks as well as in the way finding task we find significant benefits of audiovisual landmark presentation versus visual or auditory landmarks.

Multisensory effects on spatial memory performance
Our first hypothesis was that an audiovisual presentation of landmarks improves user's spatial memory performance compared with either auditory or visual presentations. Spatial memory was measured in three different tasks.
The maze drawing task was scored for the number of reported landmarks and for the number of drawn corridors. The number of recalled landmarks in the maze drawing task is in agreement with Thompson & Paivio [36]. In this study sounds, pictures or sound-picture pairs were presented and it was found that participants

recalled about six sounds, six pictures or nine sound-picture pairs. In our study, these numbers were about six, six and eight, respectively. The minor difference in the audiovisual scores may be explained by a ceiling effect: we presented only ten landmarks whereas the number of sound, pictures, or sound-picture pairs presented by Thompson & Paivio [36] was twenty or more. The number of drawn corridors was low. On average, participants drew only 4 out of 13 corridors on their map. However, the number of drawn corridors in the audiovisual condition increased by approximately 70% compared with the visual and auditory conditions. This increase confirms our hypothesis that multisensory presentation can improve spatial memory in virtual environments. Interestingly, the multisensory benefit for the number of drawn corridors is more than twice as large as the multisensory benefit for the number of recalled landmarks. This is especially surprising considering that only the landmarks, but not the corridors, were presented multisensorially. This may be explained by the ceiling effect in the number of recalled landmarks. Because the number of recalled landmarks was close to a ceiling, it was theoretically impossible to obtain a multisensory improvement equal in size to the multisensory improvement in the number of drawn corridors. The fact that more corridors have been reported when landmarks were represented multisensorially suggests that corridors are not encoded as independent entities but by the landmarks that define them.

In the adjacency task, participants reported more adjacent landmarks when they had explored a maze containing audiovisual landmarks than containing just auditory landmarks. However, in contrast with the results for the maze drawing task, we do not find a significant difference between the visual and audiovisual condition. This is a surprising result considering that one would expect that if a participant draws more corridors in the maze drawing task he or she is expected to be able to recall more adjacent landmarks. However, participants were asked about less landmarks in the adjacency task (8 adjacent landmarks per maze) than in the maze drawing task (basically all 10 landmarks), yielding less statistical power than when all landmarks would have been tested in both tasks.

For the Landmark placement task, it did not matter whether participants had explored a world containing auditory, visual or audiovisual landmarks. In other words, we did not find a multisensory benefit. This absence of a multisensory benefit may be the result of proportional decay of memory (please note that the tests were completed in a fixed order) which decreases absolute performance differences over time [63].

An additional, more speculative, explanation is that the alignment of the internal representation that participants constructed during exploration of the mazes may not have matched the alignment of the map they were provided with [64]. Internal representations may have been scaled, rotated, or (non)linearly deformed relative to the maze itself. This so-called ‘misalignment’ can considerably affect performance [65,66,67]. Matching the internal representation to the map may have required complex operations (such as mental rotation and rescaling) which may have blurred differences in the quality of the internal representation [68,69].

Multisensory effects on navigation performance

Our second hypothesis was that audiovisual landmarks improve navigation performance. Navigation performance was measured with a way finding task (Task 4). Results show that audiovisual landmarks versus auditory or visual landmarks help participants to select shorter routes and travel faster to the target landmarks. A very strong correlation between the length of the travelled route and the travel time indicates that participants travelled faster because they selected shorter routes. These results confirm the hypothesis that audiovisual landmark presentation improves navigation performance in virtual environments.

Previous research on navigation shows that information from our spatial memory enables us to find our way in our environment [16]. This is supported by strong correlations between the quality of maps drawn and navigation performance in real and virtual environments [62,70]. Our findings show that way finding performance is significantly better for conditions with multisensory landmarks than without. For multisensory conditions we also find a coinciding improved performance for maze drawing (recalled landmarks and corridors) and landmark adjacency reporting, all associated with a more correct mental representation, the 'database' underlying way finding. Therefore we speculate that the multisensory encoding and/or retrieval of environmental information in spatial memory improve navigation performance. Further research will be required to disentangle the effects on the separate encoding and retrieval processes.

Interestingly, our findings also show no significant differences in navigation performance for the unisensory visual and auditory conditions, suggesting that auditory landmarks can be as effective as visual landmarks.

How representative is our setup for real world situations?

Several aspects of our navigation task in virtual mazes make our task less representative for real world navigation. First of all, the landmarks used in this experiment were depicted as 'posters on a wall' and not real 3D animals and objects. Second, landmarks could only be perceived at relatively short distances, and only one landmark at once, unlike our experiences in real worlds. Third, in our experiment, participants had to perform a distracter task when they explored the virtual landmarks and corridors that they had to remember. This distracter task was primarily used because it was shown that multisensory memory effects were larger when such a task was employed [19]. When navigating the real world the benefits of multisensory presentation would be lower if such distracter tasks do not exist. However, in real and virtual worlds the main task of navigators usually has to do with communication and/or problem solving tasks interfering with the navigation task. Thus, a distracter task may well reflect real world situations.

This study reveals that spatial memory as well as way finding performance in virtual worlds can significantly benefit from audiovisual landmark presentation versus visual or auditory landmarks. This may yield important functional specifications for the design of multimodal content of virtual worlds and even for the multimodal displays representing the virtual world, in particular when spatial congruency of multisensory representations is important.

References

[1] T. Wolbers, M Hegarty. What determines our navigational abilities? Trends Cogn.Sci.(Regul.Ed.). 14 (2010) 138-146.

[2] N. Etchamendy, VD Bohbot. Spontaneous navigational strategies and performance in the virtual town. Hippocampus. 17 (2007) 595–599.

[3] C. Hölscher. Adaptivity of wayfinding strategies in a multibuilding ensemble: the effects of spatial structure, task requirements and metric information. J. Environ. Psychol. 29 (2009) 208–219.

[4] G. Janzen. Memory for object location and route direction in virtual large-scale space, Q.J.Exp.Psychol. 59 (2006) 493-508.

[5] M. D'Zmura, P Colantoni, G Seyranian. Virtual environments with four or more spatial dimensions. Presence. 9 (2001) 616-631.

[6] N.H. Bakker, PJ Werkhoven, PO Passenier. The effects of proprioceptive and visual feedback on geographical orientation in virtual environments. Presence. 8 (1999) 36-53.

[7] N.H. Bakker, PJ Werkhoven, PO Passenier. Calibrating Visual Path Integration in VE. Presence. 10 (2001) 216-224.

[8] N.H. Bakker, PJ Werkhoven, PO Passenier. The Effects of Head-Slaved Navigation and the use of Teleports on Spatial Orientation in Virtual Environments (VE). Human Factors. 45 (2003), 160-169.

[9] J.B.F. van Erp, P Werkhoven. Validation of Principles for Tactile Navigation Displays. Proceedings of the 50th annual Human Factors and Ergonomics Society Conference (HFES 2006), October 2006, San Francisco, USA, 1687-1691.

[10] K.N. de Winkel, J Weesie, P Werkhoven, EL Groen. Integration of Visual and Inertial Cues in Perceived Heading of Self-motion. Journal of Vision. 10(2010) 1–10.

[11] K.N. de Winkel, F Soyka, M Barnett-Cowan, HH Bülthoff, EL Groen, P Werkhoven. Integration of Visual and Inertial Cues in the Perception of Angular Self-Motion. Experimental Brain Research. 231 (2013) 209-218.

[12] C.E. Tolman. Cognitive maps in rats and man. Psychological Review. 55 (1948) 189–208.

[13] P.W. Thorndyke, B Hayes-Roth. Differences in spatial knowledge acquired from maps and navigation. Cognitive. Psychology. 14 (1982) 560–589.

[14] R.A. Epstein, JS Higgins, SL Thompson-Schill. Learning places from views: variation in scene processing as a function of experience and navigational ability. J. Cognitive Neuroscience 17 (2005) 73–83.

[15] M. Hegarty, DR Montello, AE Richardson, T Ishikawa, K Lovelace. Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. Intelligence 34 (2006) 151–176.

[16] A.W. Siegel, SH White. The development of spatial representations of large-scale environments, Adv.Child Dev.Behav. 10 (1975) 9-55.

[17] D.R. Montello, A new framework for understanding the acquisition of spatial knowledge in large-scale environments, in: Egenhofer JM, Golledge GR (Eds.), Spatial and temporal reasoning in geographic information systems, Oxford University Press, New York, 1998, pp. 143-154.

[18] C.C. Presson, DR Montello. Points of reference in spatial cognition: Stalking the elusive landmark, Br.J.Develop.Psychol. 6 (2011) 378-381.

[19] D. Caduff, S Timpf. A framework for assessing the salience of landmarks for wayfinding tasks, Cognitive processing. 7 (2006) 23-23.

[20] K. Steck, BS Hansson, M Knaden. Smells like home: Desert ants, Cataglyphis fortis, use olfactory landmarks to pinpoint the nest. Frontiers in Zoology. 6 (2009).

[21] A. Gagliardo. Forty years of olfactory navigation in birds. Journal of Experimental Biology. 216 (2013) 2165-2171.

[22] C. Moessnang, A Finkelmeyer, A Vossen, F Schneider, U Habel. Assessing implicit odor localization in humans using a cross-modal spatial cueing paradigm. PLoS ONE. 6 (2011).

[23] J.D. Restat, SD Steck, HF Mochnatzki, H.A. Mallot. Geographical slant facilitates navigation and orientation in virtual environments. Perception. 33 (2004) 667-87.

[24] N.P. Holmes, C Spence. Multisensory integration: space, time and superadditivity, Current Biology. 15 (2005) R762-R764.

[25] BE Stein, MA Meredith, The merging of the senses., The MIT Press 1993.

[26] M.O. Ernst, HH Bülthoff. Merging the senses into a robust percept, Trends Cogn.Sci.(Regul.Ed.). 8 (2004) 162-169.

[27] I.H. Bernstein, MH Clark, BA Edelstein. Effects of an auditory signal on visual reaction time. J.Exp.Psychol. 80 (1969) 567-569.
[28] M. Hershenson. Reaction time as a measure of intersensory facilitation. J.Exp.Psychol. 63 (1962) 289-293.
[29] R.S. Nickerson. Intersensory facilitation of reaction time: Energy summation or preparation enhancement? Psychol.Rev. 80 (1973) 489-509.
[30] J.B.F. van Erp, L Eriksson, B Levin, O Carlander, J Veltman, WK Vos. Tactile cueing effects on performance in simulated aerial combat with high acceleration, Aviat.Space Environ.Med. 78 (2007) 1128-1134.
[31] J. Vroomen, B Gelder. Sound enhances visual perception: cross-modal effects of auditory organization on vision. J.Exp.Psychol: Hum.Percep.Perf. 26 (2000) 1583-1590.
[32] M.O. Ernst, MS Banks. Humans integrate visual and haptic information in a statistically optimal fashion, Nature. 415 (2002) 429-433.
[33] L. Shams, WJ Ma, U Beierholm. Sound-induced flash illusion as an optimal percept, Neuroreport. 16 (2005) 1923-1927.
[34] T.G. Philippi, JBF van Erp, PJ Werkhoven. Multisensory temporal numerosity judgment, Brain Res. 1242 (2008) 116-125.
[35] P.J. Werkhoven, JBF van Erp, TG Philippi. Counting visual and tactile events: The effect of attention on multisensory integration, Atten, Perc,Psychophysics. 71 (2009) 1854-1861.
[36] V.A. Thompson, A Paivio. Memory for pictures and sounds: Independence of auditory and visual codes. Can.J.Exp.Psychol. 48 (1994) 380-398.
[37] F. Delogu, A Raffone, MO Belardinelli. Semantic encoding in working memory: Is there a (multi) modality effect? Memory. 17 (2009) 655-663.
[38] J.B.F. van Erp, TG Philippi, PJ Werkhoven. Multisensory Presentation Benefits Memory for Semantic but not for Non-semantic Items. J of Cognitive Psychology. Under review.
[39] A.D. Baddeley, G Hitch, G Bower, The psychology of learning and motivation, Advances in research and theory, Academic Press, New York, 1974, pp. 47-89.
[40] A.D. Baddeley. Working memory, Oxford University Press, New York, 1986.
[41] R.H. Logie. Visuo-spatial processing in working memory, Q.J.Exp.Psychol. 38 (1986) 229-247.
[42] A. Baddeley, S Grant, E Wight, N Thomson, Imagery and visual working memory, in: Rabbitts PMA, Dornic S (Eds.), Attention and performance V, Academic Press, London, 1975, pp. 205-217.
[43] A. Baddeley. The episodic buffer: a new component of working memory? Trends Cogn.Sci.(Regul.Ed.). 4 (2000) 417-423.
[44] A. Baddeley, BA Wilson. Prose recall and amnesia: Implications for the structure of working memory, Neuropsychologia. 40 (2002) 1737-1743.
[45] B. van Geldorp, HC Bergmann, J Robertson, AJ Wester, RPC Kessels. The interaction of working memory performance and episodic memory formation in patients with Korsakoff's amnesia, Brain Res. 1433 (2011) 98-103.
[46] D.H. Uttal. Seeing the big picture: map use and the development of spatial cognition. Dev. Sci. 3 (2000) 247–286.
[47] J.F. Norman, CE Crabtree, AM Vlayton, H.F. Norman. The perception of distances and spatial relationships in natural outdoor environments. Perception. 34 (2005) 1315–1324.
[48] T. Ishikawa, DR Montello. Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. Cogn. Psychol. 52 (2006) 93–129.
[49] J.W. Kelly, TP McNamara, B Bodenheimer, T.H. Carr, J.J. Rieser. Individual differences in using geometric and featural cues to maintain spatial orientation: cue quantity and cue ambiguity are more important than cue type. Psychon. Bull. Rev. 16 (2009) 176–181.
[50] P. Rutherford. Virtual Acoustic Technology: Its Role in the Development of an Auditory Navigation Beacon for Building Evacuation. Proc. 4th UK Virtual Reality SIG Conf., R. Bowden, ed., United Kingdom Virtual Reality Special Interest Group, 1997.
[51] T. Lokki, M Gröhn, L Savioja, T Takala. A Case Study of Auditory Navigation in Virtual Acoustic Environments. Proceedings of International Conf. on Auditory Display (ICAD2000) 145-150.
[52] B. Walker, J Lindsay. Effect of Beacon Sounds on Navigation Performance in a Virtual Reality Environment. Proceedings of International Conf. on Auditory Display (ICAD2003), 204-207.
[53] T. Lokki, M Gröhn (2005). Navigation with auditory cues in a virtual environment. Multimedia IEEE. 12 (2005) 599-614.
[54] R. Gunther, R Kazman, C MacGregor. Using 3D sound as a navigational aid in virtual environments, Behav & Inform Techn. 23 (2004) 435-446.

[55] C. Ardito, M Costabile, A De Angeli, F Pittarello. Navigation help in 3D worlds: some empirical evidences on use of sound, Multimedia Tools Appl. 33 (2007) 201-216.
[56] P.J. Laurenti, RA Kraft, JA Maldjian, JH Burdette, MT Wallace. Semantic congruence is a critical factor in multisensory behavioral performance, Exp. Brain Res. 158 (2004) 405-414.
[57] S. Yuval-Greenberg, LY Deouell. What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition, J.Neuroscience. 27 (2007) 1090-1096.
[58] S. Yuval-Greenberg, LY Deouell. The dog's meow: asymmetrical interaction in cross-modal object recognition, Exp.Brain Res. 193 (2009) 603-614.
[59] J.G. Snodgrass, M Vanderwaart. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J.Exp.Psychol: Hum.Learn.Mem. 6 (1980) 174-215.
[60] R.P. Darken, JL Sibert. Navigating large virtual spaces, Int. J. Human-Computer Interaction. 8 (1996) 49-71.
[61] S.E. Goldin, PW Thorndyke. Simulating navigation for spatial knowledge acquisition, Human Factors. 24 (1982) 457-471.
[62] M.J. Rovine, GD Weisman. Sketch-map variables as predictors of way-finding performance, J.Environ.Psychol. 9 (1989) 217-232.
[63] C.V. Buhusi, WH Meck. Interval timing with gaps and distracters: evaluation of the ambiguity, switch, and time-sharing hypotheses. J.Exp.Psychol.: Anim.Behav.Processes. 32 (2006) 329.
[64] A.L. Shelton, TP McNamara. Orientation and perspective dependence in route and survey learning. J.Exp.Psychol: Learn.Mem.Cognit. 30 (2004) 158-170.
[65] M. Palij, M Levine, T Kahan. The orientation of cognitive maps. Bull.Psychon.Soc. 22 (1984) 105-108.
[66] C.C. Presson, N DeLange, MD Hazelrigg. Orientation specificity in spatial memory: What makes a path different from a map of the path? J.Exp.Psych: Learn.Mem.Cognit. 15 (1989) 887-897.
[67] A.E. Richardson, DR Montello, M. Hegarty. Spatial knowledge acquisition from maps and from navigation in real and virtual environments, Mem.Cognit. 27 (1999) 741-750.
[68] M.A. Just, PA Carpenter. Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. Psychol.Rev. 92 (1985) 137-172.
[69] P. Péruch, A Savoyant. Conflicting spatial frames of reference in a locating task, in: Logie RH, Denis M (Eds.), Mental images in human cognition, North Holland, Amsterdam (1991) 47-55.
[70] S. Murakoshi, M Kawai. Use of knowledge and heuristics for wayfinding in an artificial environment, Environ.Behav. 32 (2000) 756-774.