HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

INTERACTING WITH 3D AND 2D MOBILE MAPS: AN EXPLORATORY STUDY

Antti Oulasvirta[1], Antti Nurminen[1], Annu-Maaria Nivala [2]

[1] Helsinki Institute for Information Technology (HIIT)

[2] Finnish Geodetic Institute

April 11, 2007

HIIT TECHNICAL REPORT 2007-1

Interacting with 3D And 2D Mobile Maps: An Exploratory Study

Antti Oulasvirta[1], Antti Nurminen [1], Annu-Maaria Nivala [2]

[1] Helsinki Institute for Information Technology (HIIT)
[2] Finnish Geodetic Institute

Correspondence concerning this report should be addressed to Antti Oulasvirta, Helsinki Institute for Information Technology (HIIT), P.O. Box 9800, 02015 TKK, Finland. E-mail: antti.oulasvirta@hiit.fi

HIIT Technical Reports 2007-1

ISSN 1458-9478

INTERACTING WITH 3D AND 2D MOBILE MAPS: AN EXPLORATORY STUDY

Antti Oulasvirta[1], Antti Nurminen [1], Annu-Maaria Nivala [2]

[1] Helsinki Institute for Information Technology (HIIT)
[2] Finnish Geodetic Institute

Abstract

To study how users of mobile maps construct the *referential relationship* between points in the virtual space and in the surrounding physical space, a semi-natural field study was conducted in a built city environment, utilizing a photorealistic interactive map displayed on a PDA screen. The subjects (N=8) were shown a target (building) in the physical or in the virtual space and asked to point to the corresponding object in the other space. Two viewports, “2D” (top-down) and “3D” (street-level), were compared, both allowing the users to move freely. 3D surpassed 2D in performance (e.g., 23% faster task completion) and workload measures (e.g., 53% in NASA-TLX), particularly in non-proximal search tasks where the target lied outside the initial screen. Verbal protocols and search pattern visualizations revealed that subjects utilized varied perceptual-interactive search strategies, many of which were specific to the viewport type. These strategies are set and discussed in the terms of the theory of pragmatic and epistemic action (Kirsh & Maglio, 1994). It is argued that the purpose of such strategies was not to find the target directly by transforming either of the two spaces (pragmatic action), but to make the task processing requirements better aligned with the available cognitive resources (epistemic action). Implications to interface design are discussed with examples.

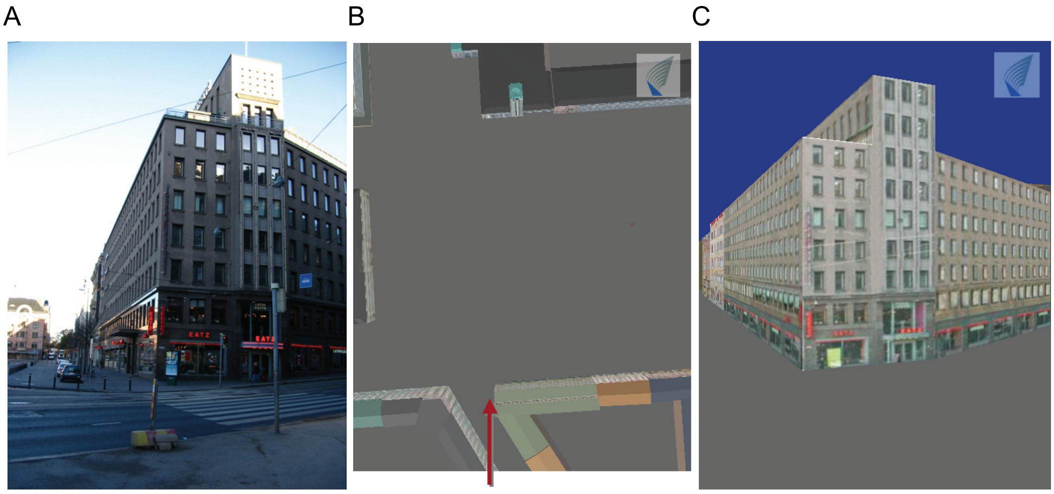

Figure 1. A) The real scene and its representation in the two view modes of m-Loma: B) 3D and C) 2D (red arrow is added to mark the building presented in figure A). In both 3D and 2D the user has 4 degrees of freedom in moving in the map: left-right rotation, up-down rotation, up-down movement, and forward-backward movement.

1 Introduction

Recently, novel interactive map applications have gained ground in the consumer market and in the research field of HCI. At first, mobile maps were designed mainly for in-car navigation and only featured pre-stored raster maps. Nowadays there are also applications utilizing vector maps for multiple purposes, such as for location-based information for tourists and urbanites, showing points-of-interests (POIs) such as restaurants and shops, public places and sights (e.g., Cheverst et al., 2000, Abowd et al., 1997). Other map-based mobile services include messaging and public forums (Heidmann, Hermann and Peissner, 2003; Rantanen et al., 2004).

Figure 1 illustrates two different presentations of a user's environment, adopted from the particular mobile map investigated in this paper. Figure 1A shows a photo of a real building in the Helsinki city center, 1B its presentation in the "2D" view of the application and 1C in the "3D" view. In contrast to the 2D view, the 3D view (or street-level view) shows the immediate surroundings of an object but not its global position or relation to distant objects (cf. Baudisch & Rosenholtz, 2003). The design of the system subscribes to the principle of pictorial realism (or veridicality, Stuart, 1996) stating that "displays are more effective when encoded graphically and can be readily identified with its meaning in the real world" (Ruddle, Payne, & Jones, 1999, p. 56).

In general, what makes mobile maps distinct and *different* from (typical) virtual environments (VEs)— such as virtual reality and desktop-based navigation systems—is that users are actually *physically embedded* in the world that the virtual model represents. When interacting with a physical environment (PE), the roles of eye, head and body movements for acquiring information are emphasized, whereas VEs are typically associated with decreased field of view and low fidelity of landmarks and non-visual cues. This difference is crucial, as it means that users have to *mentally construct the referential relationship* between the virtual and physical information spaces. By contrast, in augmented reality, the two spaces, or at least parts of them, are superimposed on the display (Milgram & Colquhoun Jr., 1999; Trevisan, Vanderdonckt, & Macq, 2004). This mentally performed task, called henceforth the *mapping task*, is illustrated in Figure 2. Solving the mapping underlies more complex tasks carried out on mobile maps—such as locating, investigating, or exploring locations, objects, or areas, estimating spatial relationship (e.g., distance) between them, finding routes etc.

What makes mobile maps further different from VEs and map artifacts is the strong influence of the *keyhole property* to interaction (Woods & Watts, 1997). The keyhole property means that "the number of potential views is much greater than the physical size of the available viewports" (p. 619). Since direct recognition is not often possible due to this property, unless the target happens to be on the display, users have to *interact* with the space(s) to achieve a position in which the target can be matched. In addition, and in contrast to VEs, mobile maps allow the user to interact with *two* spaces instead of one: physical and virtual movement can be used to search for the target. Previous models of VE interaction (e.g., Jul & Furnas, 1997) do not account for acting in two spaces at the same time.

Thus far, no studies have reported how users solve the mapping problem with mobile maps, given these special characteristics. The present paper provides the first empirical study investigating this question, particularly comparing 2D vs. 3D. This comparison is of both practical and theoretical interest. The results are examined with the purpose of constructing a theoretical model of mobile map interaction. The paper concludes by discussing implications to design. Before turning to the method, we review previous studies on mobile maps and orientation to identify relevant open questions.

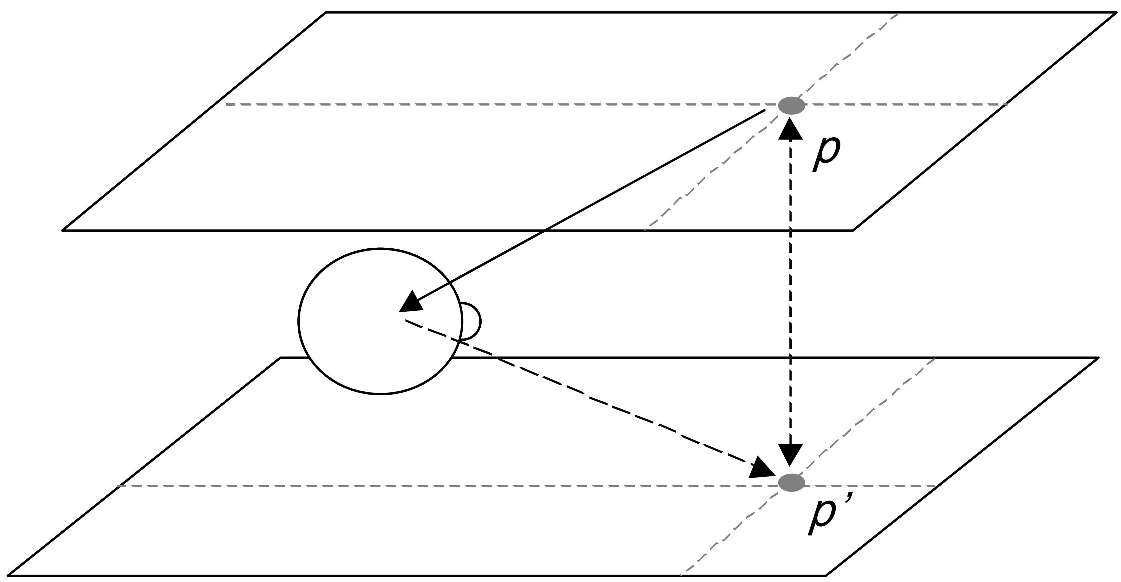

Figure 2. The virtual–physical mapping problem: the user has to mentally construct a correspondence between points p and p' in the virtual and physical spaces.

1.1 Previous Field Studies

One of the problems of the human-eye (3D) view is that users have to integrate information from many different positions to learn the layout of large-scale environments by navigation (Ruddle, Payne, & Jones, 1999). On the other hand, users might be more familiar with the interaction styles of 2D maps, so it is not easy to predict which map type is better. Therefore, several developers have empirically compared 2D maps (paper maps or mobile maps) with 3D.

Laakso (2002) found that a 2D map was faster to use in orientation and wayfinding, but that use of 3D maps was “more fun”. Burigat and Chittaro (2005) described a system where POIs can be queried by pointing in a 3D representation. Their field test allowed users to move freely in a certain city square. They concluded that users had “no problems” in matching the real world objects with the 3D view. Rakkolainen and Vainio’s (2001) user interface (UI) featured a 2D map and a 3D model view simultaneously shown on the screen. The results of a pilot study indicated that users preferred to use a combined view rather than either 2D or 3D alone, corroborating findings with a VE system (Ruddle, Payne, & Jones, 1999). Rakkolainen and Vainio (2001) speculated that 3D allowed users to better recognize their own position and the landmarks than the 2D map. Subsequently, Vainio and Kotala (2002) improved their system with a symbol showing the user’s location and viewing direction. On the basis of a small study, they concluded that the augmented 3D model illustrates motion more clearly than their 2D map alone. To conclude, it is still an open question which representation is better suitable for which tasks.

Bessa et al. (2005) studied what *features* are utilized in orientation with a 3D map. The users were first shown photographs of a certain place and asked to identify the locations and verbalize the key features they used for identification. After this they were brought to the same physical location (from a different direction than the photo) and asked to identify the same spot and specify the features that they used for identification. The key features were divided into urban furniture (lamps, seats etc.), buildings (including doors, windows), publicity (advertisements) and “other” (cars, trees, temporary elements). They concluded that only the overall geometry and a few key features were required for recognition and, further, that the selected features were not necessarily those most salient to the human visual system. Nothegger et al. (2004) evaluated the feature salience of landmarks (selected automatically by their algorithm). The task was to look at panoramic images of intersections and name the most prominent façade. They found that subjects’ criteria for selection varied notably (façade size, shape, shape deviation, color, visibility, cultural importance, and certain explicit marks). It seems that many features are attended to, but no systematic studies exist to compare 2D and 3D *interactive* maps in this respect.

1.2 Orientation: Previous Findings and Open Questions

Before embarking on our empirical study, we provide a brief review of cognitive phenomena related to orientation in general (with and without map artifacts). The goal is to discern possible differences implied to interaction strategies with 2D vs. 3D.

First, the *alignment* of the representation and the represented space is important, because some central types of human spatial knowledge are known to be *viewpoint-dependent,* even representations of dynamic scenes (Garsoffky, Schwan, & Hesse, 2002). Hence, when objects in a map do not correspond to stored representations, the user has to transform or rotate the representation, which implies mental or physical effort (Levine, 1982; Levine, Jankovic, & Palij, 1982). Mou and McNamara (2002) elaborate this view by arguing that spatial memories are defined with respect to intrinsic frames of reference, which are selected on the basis of egocentric experience and environmental cues. This basically implies that the cues available determine the availability of pre-existing knowledge for solving the task, and this is where 3D and 2D can be expected to differ.

It is also worth mentioning that orientation-dependence may be exaggerated in the case of mobile maps where display size is small. Presson, DeLange, and Hazelrigg (1989) claim that small displays do not "afford" large spatial arrays to be coded by the perceptual system when moving, as movement in the surrounding environment does. Moreover, they do not support as much perceptual exploration and scanning. Roskos-Ewoldsen et al. (1998) agree with the effect but disagree with the explanation and claim that mental representations of large-scale spaces are orientation dependent as well, and that Presson et al.'s findings were an artefact of procedure.

Second, because the use of mobile maps takes place when the users are moving (mobile), the map's support for *spatial updating*—the mechanisms involved in locating positions in space relative to oneself after a given spatial transformation—is emphasized (Wraga, 2003). Wang and Brockmole (2003) have examined the influence of the environment being divided into nested structures (e.g., city consisting of district consisting of blocks) and noted that spatial updating is not carried out in all structures simultaneously nor with same accuracy. When switching to a new environment, track of one's position relative to old environments is often lost. From this perspective, the 2D map could provide better means for users to do updating at different levels of topographical span, and this capability might in turn facilitate solving the mapping problem. However, because of the keyhole property, even the 2D map is actually quite narrow and cannot necessarily encompass more than a few blocks at the most.

Third, the small *scale* of transformations (e.g. Golledge, 1999) in mobile maps may lower the informativeness and recognizability of objects. Only recognizable objects can be utilized as landmarks that help the tasks of mapping and orientation. Therefore, 2D is expected to differ from 3D by trading off informativeness of an object to informativeness of an area. Witmer, Sadowski, and Finkelstein (2002) showed that additional aerial views enhance users' ability to *navigate* through a VE. Furthermore, in order to help in understanding the relationship between virtual and physical, landmarks in the virtual model also have to be *distinctive* (stand out), which depends on the visual as well as structural qualities of the view (Sorrows & Hirtle, 1999).

1.3 Approach

The present study is not conducted with an aim at generalizing to the general population—the small N would not allow for that in any case. Neither is the purpose to evaluate which is "better", 3D or 2D. Furthermore, instead of constructing a clear-cut hypothesis on mobile map interaction to be tested on the field, we approach the open questions from an exploratory point of view. Several variables of interest are controlled for, but the focus of analysis lies rather on providing a holistic account of mobile map interaction rather than merely reporting factors that affect performance. Thus, the purpose is qualitative: to learn and construct a working model of situated interaction in the face of the mapping problem. Triangulation—utilizing multiple methods—is important in early phases of research where the knowledge base has not yet been established and where affirmation or falsification of an established hypothesis is not possible. Observations, descriptions, and explanations of situated interactions are needed for the purposes of exploration and theory development of holistic, complex, dynamic and exceptional phenomena (Camic, Rhodes, & Yardley, 2003), such as the mapping problem. For a similar type of meta-scientific approach, see Darken and Sibert (1993, 1996).

2 Method

Our method loosely follows the *pointing task paradigm* utilized in the field of spatial cognition. The participants are brought to a spot in an urban area, and while facing a certain direction they are shown an object (building) in one of the spaces (virtual or physical) and asked to point (with the index finger) to the corresponding object in the other space. They are free to move both in the virtual and in the physical environment as they will. This method was chosen because it provides a simple operationalization for the mapping problem.

Four qualities are manipulated in the study. First, a critical quality of mobile map interaction is that the target can lie either in the physical or in the virtual environment. Therefore, our method tests search in both conditions. This independent variable (IV) is called the *source of orientation* (i.e., does the task start with knowing the target in the physical or virtual environment). Second, half of the trials use 2D, half 3D. This comparison is interesting for the differences hypothesized in the previous section. Third, to expand the analysis from proximal identification (i.e., target residing in the directly observable environment), one third of the tasks involve navigation to a remote location in an environment. Thus, two task types are used: *proximal mapping* (recognition of a target from the immediately perceivable surroundings) and *remote navigation* (moving from the present location to an out-of-sight location). As discussed above, the two representations can be expected to differ in their support for different operational scales (object vs. area). Fourth, the *misalignment angle* of the two environments in the beginning of the task was controlled for, because of the hypothesized emphasis of orientation-dependence in mobile maps.

For data collection, the empirical strategy of “triangulation” is utilized, meaning the utilization of several types and sources of data. Five kinds of measures were employed:

- *Performance measures,* for example task completion times, number of restarts, number of different types of key presses etc.
- *Subjective workload ratings* (NASA-TLX)
- *Interaction logs,* analyzed with a custom-made visualization and replay software. The first interaction in the beginning of a task automatically started the logging.
- *Video data* on users’ interaction with the device and movement in the physical environment
- *Verbal protocols* during the task and after its completion.
- Interaction logs were manually synchronized with the video data by observing the moment from the video when the participant started interaction with the application after the experimenter gave the instructions.

2.1 Participants

Eight young adults (average age = 27.5 years), accustomed to PCs (average computer use = 11.8 years), volunteered in the study. Half of them were students, the other half were full-time employed in mixed occupations. Seven of the participants were male, one of them female. They had lived in Helsinki in average of 13.8 years. Six participants had previous experience with virtual environments from 3D games but only two had previously used 3D map applications.

2.2 Virtual Model and Interaction

m-Loma, "mobile Location-Aware Messaging Application" is a working prototype of an interactive map running on mobile terminals. The model was built on top of city blueprints in 3D modeling software by referencing to digital photography. Textures were created from digital photographs taken from the facades, applying image processing software to remove the perspective distortion. The 3D model was created in two phases, first photographing and creating the 475 facade and statue textures and then modeling the 183 textured and 101 dummy buildings. Additional resources were spent in planning how a model should be constructed to optimally suit a *potentially visible sets* (PVS) engine. During our initial field studies between the two modeling phases all appropriately modeled targets were easily recognized, but those with modeling flaws such as lack of textures caused severe disorientation. The second modeling phase corrected many of the noticed flaws.

The resulting model covers an area of approximately 1 square kilometer in the center of Helsinki. The model contains buildings with realistically textured facades and flat-colored rooftops. The model also contains all statues from the area, represented as billboards. Large areas such as parks and sea are modeled as flat-colored polygonal ground. Parks may also include billboard trees. The model lies on a flat surface, even though the engine supports topographical models. Furthermore, the environment can be populated with location-based information billboards, such as restaurant logos and bus stops. Both a static, scrollable 2D raster map and a scalable vector-rendered 2D road map are supported. The raster map currently used is provided by the city of Helsinki, and contains roads, building outlines and street names. The vector map provides streets and street names only. Both 2D map types can be populated with location-based information, represented as 2D icons.

For the purpose of the study, m-Loma was installed on a Dell Axim PDA with a display of 240 x 320 pixels and a simple input/output (I/O) layout (on the bottom side of the front panel) with the joystick in the middle and two buttons on the sides. The Axim also encompasses a pointing stick (a stylus), which was adopted by some participants (although this was not required) to point out the found targets. Forward and backward keys moved the view accordingly and left and right keys rotated the view horizontally. In addition, there were two keys for vertical rotation.

2.3 3D Graphics

The goal for developing the graphics engine was to allow free roaming in a photorealistic 3D space, without viewpoint position or orientation restrictions, and with sufficiently fast real-time 3D rendering in any situation. The system needed to provide a large view distance and a collision avoidance scheme to keep the viewpoint outside the buildings. Such requirements make a straightforward model visualization, for example with a VRML viewer, quite unfeasible.

To find out which buildings are visible, which is difficult at run-time in a highly occluded city environment, the 3D engine applies the PVS algorithm (Airey, 1991) that limits the number of objects to render. During a preprocess stage, the viewing space is divided in three dimensions into virtual cubes, called "*cells",* and all objects visible from each cell are recorded to a list. This large list is compressed into clusters of difference trees.

Visibility clusters may be created according to the assumed computational resources of potential devices by varying the visibility range. In addition, objects contributing only little to the scene can be discarded by defining a *hardly visible set* using a fixed pixel count. For example, only objects contributing more than 2 pixels could be accepted as "visible". Typically three PVS sets are created: one for smart phones, one for PDA's and one for PC's. The resulting cluster files require approximately 2-10MB, depending on the visibility range being used.

The m-Loma system accepts VRML models, even though in principle any boundary representation (*B-rep*) format could be used. The m-Loma buildings are created as low-polygon models suitable for mobile use. They consist of separate walls and a roof that are used as atomic objects in the visibility sense, providing a decently optimizable model for the PVS, balancing the visibility list size and model complexity.

While the model is very lightweight in geometry, all buildings are individually textured. The engine performs texture and memory management on the fly, holding only one *level of detail* of each texture in the memory, based on estimated size on screen. At large distances, when only a few pixels are visible, the dominant texture color is used to color entire meshes. Textures have been preprocessed to four levels of detail. To avoid slow file I/O, that happens especially when the user moves back and forth, configurable memory caches are used for texture data.

The calculation of potentially visible objects does not consider viewing direction. An efficient run-time *frustum culling* method takes care of rendering only those objects that lie within the current view space. Frustum culling refers to the removal of objects that do not reside within the view frustum, which is the volume of space that is projected to the screen, defined by left, right, top and lower planes, and front and back planes.

Temporal and spatial coherence are exploited to minimize the computational load. Only a part of the model is tested against the frustum at any given time. If any meshes are found visible, they are supposed to be visible for a while. Frustum culling is dominating the culling processes at sky level, especially when the user is viewing downwards. At street level, the PVS takes care of removing most of the invisible objects from the rendering pipeline. When the viewpoint is moving, geometry, visibility lists and textures are loaded from local files (or cache). To allow continuous movement, at maximum one texture per rendered frame is loaded. If the user moves rapidly, geometry is loaded first and textures later. Therefore, the user may see textures being loaded as they appear on surfaces.

The client was developed with the C language and OpenGL ES 3D API on Linux, and ported to MacOS X, Windows, WindowsCE and Symbian. OpenGL ES exposes the rendering pipeline features and allows the use of the optimization techniques described above. On average, flying around the model at street level and in the sky (including the 2D "top-down" view), real frame rates with a Dell Axim X30 624MHz PDA are 8-12 fps.

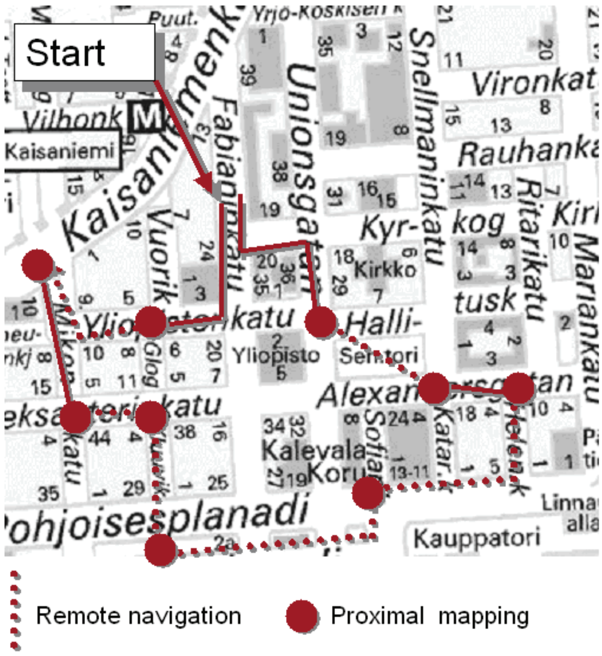

Figure 3. The route of study sites going through the centre of Helsinki

2.4 Materials and Design

Remote navigation tasks required the subject to move in both the physical environment (PE) and in the virtual environment (VE) to a remote location (not in the directly observable environment) about the distance of two blocks. *Proximal mapping tasks* required the subject to locate a point in the PE based on a view of the VE. Both tasks could be reversed so that the target could be pointed in the PE, calling for movements in the VE.

Each participant performed six navigation and ten proximal mapping tasks along a route, as shown in Figure 3. In each of the orientation sites, two proximal orientation tasks were presented: one in 2D and one in 3D; of which one was with the VE as the source environment and one with the PE. In order to balance the order of tasks we had half of the subjects taking the route in counter clockwise direction. We also altered the order of tasks along with the view mode (2D vs. 3D). These three manipulations resulted in having 2 (source VE vs. PE) x 2 (route clockwise vs. counterclockwise) x 2 (first task 2D vs. 3D) = 8 sets of materials.

2.5 Procedure

First, the subject's background and familiarity with the technology were assessed with an initial questionnaire. Next, they were familiarized with the PDA and with m-Loma.

In introducing a task, the experimenter read out the task description and then let the subject select the corresponding task from m-Loma menu, which started recording key presses of the user. Selecting the task could also be used during the task for restarting it if necessary. The tasks were sequenced so that two orientation tasks (task pair) were followed by a single navigation. The subject ended the task after she/he thought the task was completed. No questions to the experimenter were allowed during a task, but subjects were encouraged to speak during the trial and were asked after trials to describe how they solved the problem. During the trial, the experimenter stayed a few steps behind the subject so that interaction with the device and utterances could be videotaped. On *each* experimental site, a NASA-TLX questionnaire (the pen and paper scale from Vidulich & Tsang, 1986, translated to Finnish by the authors) was completed. After completing the route, the subject was interviewed.

2.6 Countering Potential Threats to Validity

The purpose of adopting a naturalistic instead of laboratory approach is to maximize contextual realism and hence “ecological validity”. In doing so, however, a considerable precision and control is given up, compromising the ability to make strong causal inferences. From the perspective of validity of the *quantitative* results, there are four problems introduced that are distinctive to this kind of field inquiry: 1) low statistical power, 2) low reliability of measurements (e.g., missing data and corrupted records due to technical problems appearing in field conditions), 3) random irrelevancies in the experimental setting (e.g., caused by abrupt uncontrolled environmental events), and 4) reliance on just one method can hinder understanding of the whole phenomenon (a.k.a. mono-method bias). These issues were addressed by the following means: 1) gathering a large number of data points per subject to ensure moderate statistical power; 2) using measurements and apparatus that were tested and found reliable in our previous field trials using video recordings, to 3) spot irrelevant events and excluding respective data from subsequent analyses; and 4) utilizing several methods comparatively.

From the perspective of the *qualitative* results, the control imposed by the experimenter of course implies somewhat unrealistic use situations. For example, the participants are forced to switch the representation modality from 2D to 3D after every task, and they cannot freely choose which tasks to do with the mobile map, or where to go. At present, it is impossible to evaluate how critical such manipulations are to “natural” use.

3 Quantitative Results

The quantitative results are analyzed here, whereas the verbal protocols and video data are reported in the next section. Due to technical problems, the second half of one participant’s data could not be analyzed quantitatively.

3.1 Task Performance

Overall, *completing the tasks* was 23% quicker with 3D (M=93.5, SD=62.9) than with 2D (M=121.1, SD=96.21), and this difference was statistically significant, $t(213)=5.50$, $p<.05$). The benefit for 3D was systematic in the sense that 2D was significantly faster to use only in two experimental sites of all 16. Moreover, there were less restarts per task for 3D than for 2D: 12% vs. 17% on average, respectively.

As illustrated in Figure 4, the largest difference—a 25% benefit for 3D—was present in remote navigation tasks. A slightly smaller benefit was associated with proximal orientation tasks (22%). Only the first difference was statistically significant in *Least Significant Difference* (LSD) post hoc test ($p<.05$).

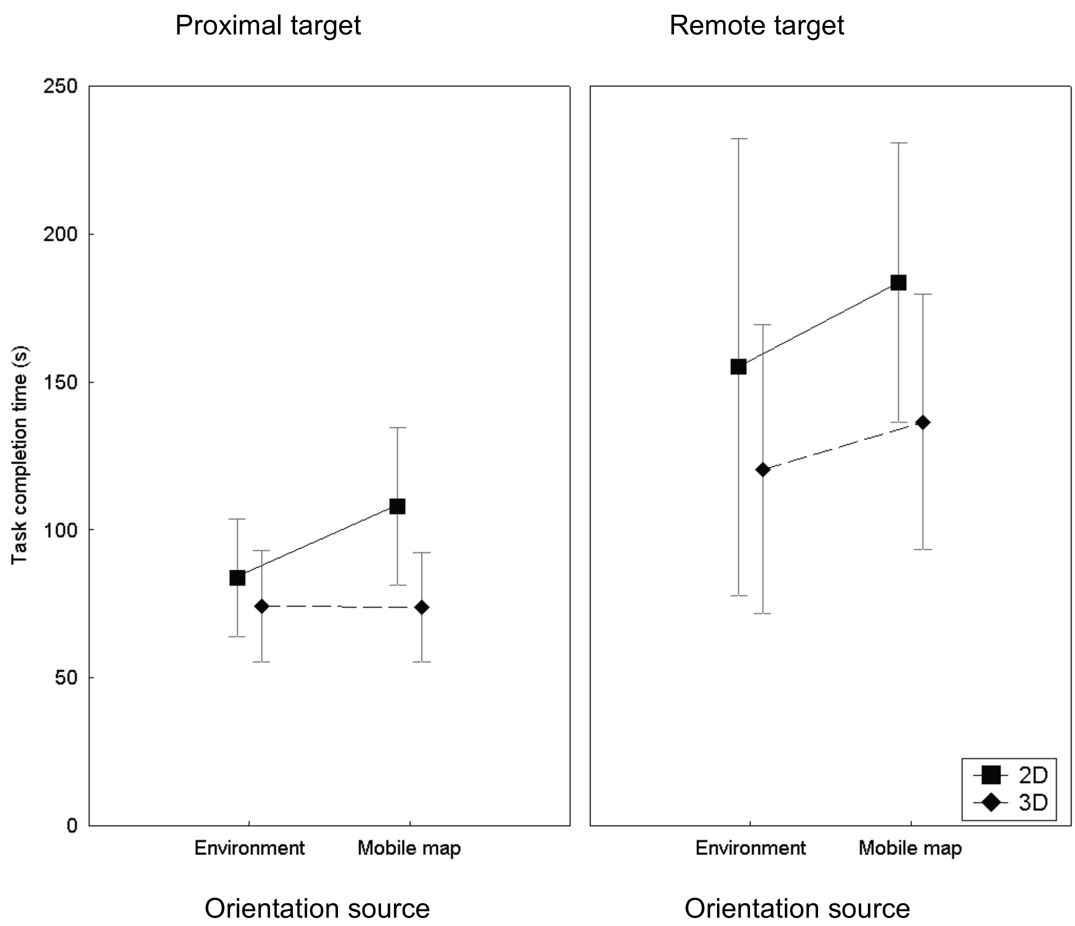

Figure 4. Task completion times. Vertical bars denote 95% CIs.

As illustrated in Figure 5, for both views, a *misalignment* of 90 degrees produced the worst results. By contrast, performance in the 180 degrees condition was close to that of no misalignment. It also appears that the only significant and borderline-significant differences between 3D and 2D are in the conditions of 0, 45, and 90 degrees.

3.2 User Input

Interaction logs were analyzed utilizing automatic scripts, of which the most surprising and reliable are reported here. (However, the omitted ones do not contrast the reported ones.)

The total *distance traveled* (in meters) in the mobile map was more than doubly longer for 2D (M=593.04 m, SD=755.8) than for 3D (M=240.5 m, SD=436.5), this difference being highly significant, $t(213)=16.89$, $p<.001$. Several other measures of interaction were examined as well. There were no differences between 3D and 2D in *horizontal rotation* behavior (degrees per task), $t(170)=.36$, $p=.54$. By contrast, the amount of *vertical rotation* (degrees per task) was different between 3D and 2D, $t(170)=17.65$, $p<.001$. There were close to zero *vertical rotation commands* in the 3D condition whereas in 2D the mean was 33.56. Similarly, there were significantly fewer *up/down* presses in the 3D condition, $t(213)=13.34$, $p<.001$. There were no significant interaction effects involving orientation-source as a factor, which hints that these observations are due to the map type not the direction of orientation in the task.

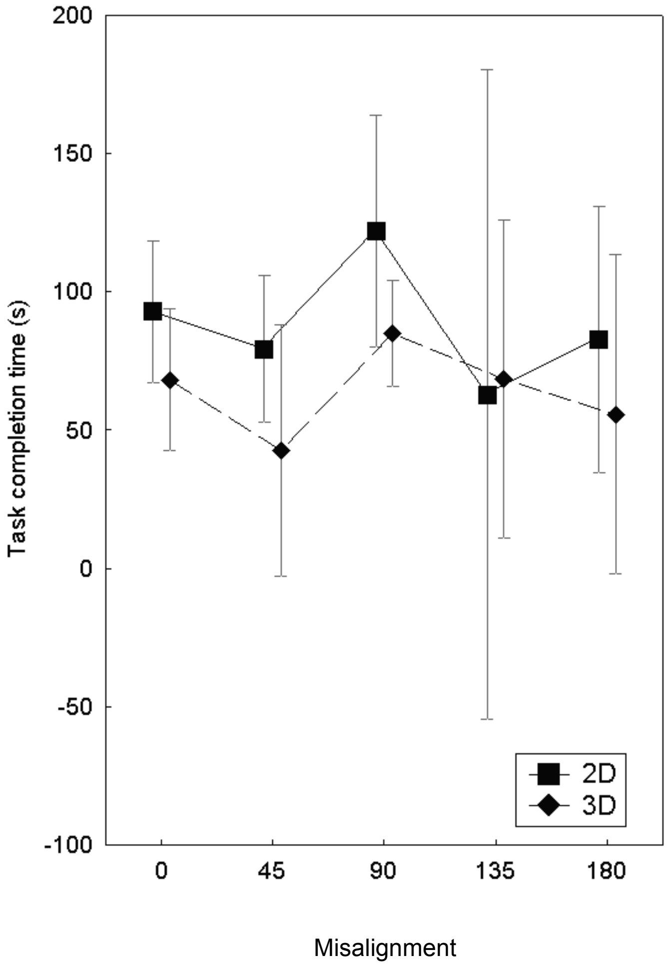

Figure 5. Task completion times in misalignment situations in proximal orientation tasks. Vertical bars denote 95% CIs.

Table 1. NASA Task Load Indices per map type.

	Measure					
Map type	Mental demand	Physical demand	Temporal demand	Effort	Performance	Frustration level
2D	2.73 (3.05)	1.34 (1.55)	2.47 (3.06)	2.57 (2.84)	3.63 (3.75)	2.47 (2.95)
3D	1.52 (1.35)	0.93 (0.90)	1.31 (1.13)	1.24 (1.16)	2.15 (2.58)	1.15 (1.59)

Note. Standard deviations are in parentheses

3.3 Subjective Workload (NASA-TLX Questionnaires)

Generally, 2D was associated with higher workload than 3D, as apparent in Table 1. A one-way ANOVA revealed that the effect of map type was significant on the five component of NASA-TLX—mental demand, temporal demand, effort, performance, and frustration—and their arithmetic mean (all *p*s<.05). The physical demand measure was an exception in this sense (*p*=.08), although the trend was similar. There were no interaction effects.

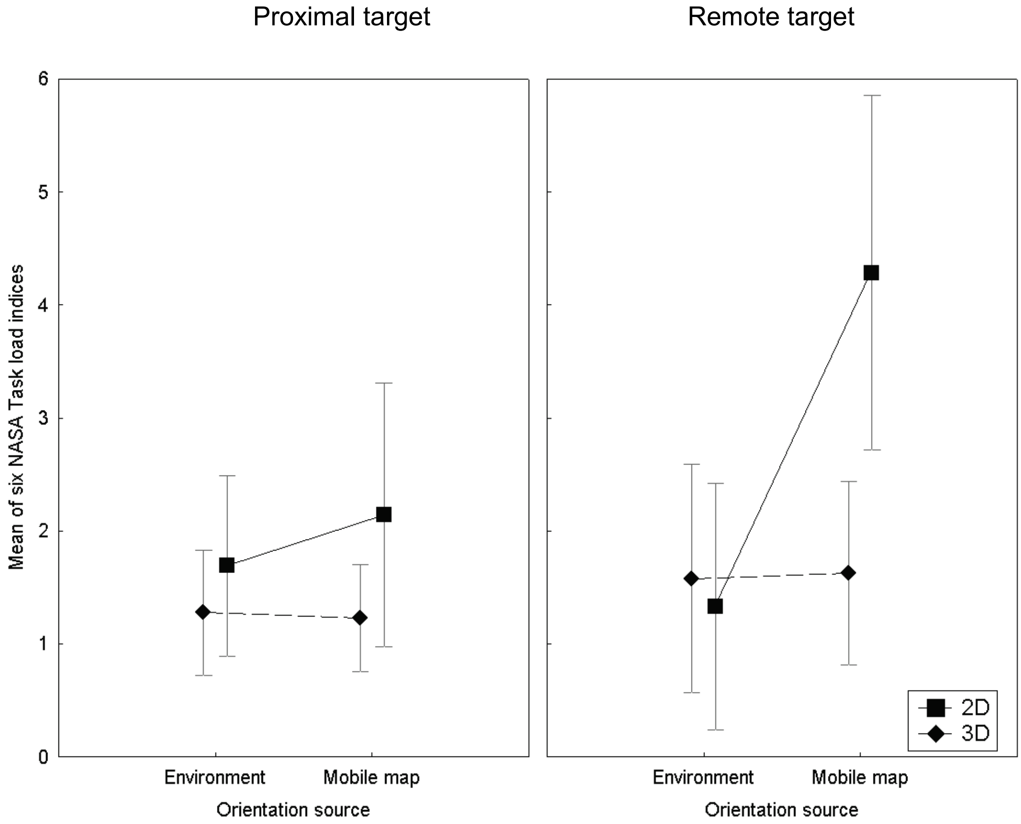

Figure 6. Subjective workload, as measured by aggregating six NASA Task Load Indices. Vertical bars denote 95% CIs.

Interestingly, the 53% total difference for the benefit of 3D resulted almost entirely from two components: (1) a slight overall benefit for 3D in the proximal orientation condition (that was however non-significant in a LSD post hoc test, p=.99); (2) a large 62% benefit for 3D in the remote navigation condition, but only when the orientation source was the mobile map (LSD post hoc test against 2D was highly significant, p<.01). The situation is illustrated in Figure 6.

Regarding the misalignment factor, the results exhibited a similar but weaker pattern to task completion times. Consequently, this report is omitted here to save space.

4 Verbal Protocols

This subsection reports participants' verbal accounts separately for the two representation types. The format of this subsection accords loosely to chronological order of course of actions in searching for a target. The benefit of this approach is that for each subtask of search, several individual differences in approach will be observed. The shortcoming, consequently, is that the description will not provide a coherent account of any single individual participant's protocol.

To illustrate search strategies, six search paths as analyzed from the logs are visualized in Figure 7. Some of the search strategies involve the user moving the viewport from top-down to street level view, or vice versa. These will be treated here according to the situation such strategies started from. (For example, if a user started with the top-down view but soon began to navigate using the street level view, all behavior after the transition to the street-level will be considered as strategy related to that viewport.)

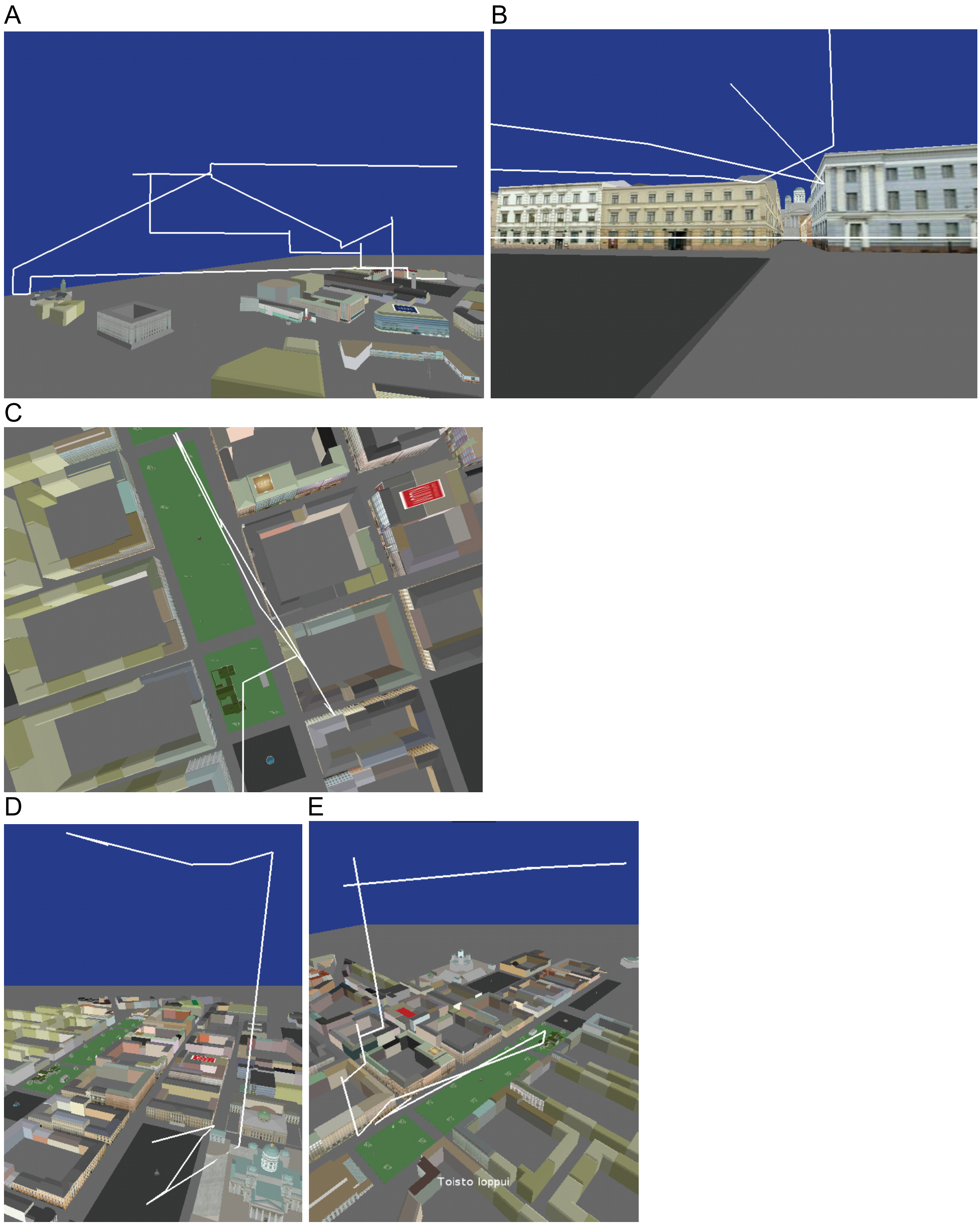

Figure 7. Illustrations of epistemic search actions from the data. A) 2D: Learning the model by exploring it from a top-down view, B) 3D: Walking around in a city square and peeking around corners, C) 2D: Flying back and forth in a square, D) 2D: Diving to street-level, E) 2D: Scanning, diving and scanning at multiple levels, and walking around at the street level. White lines represent user movement as re-constructed from the interaction logs.

4.1 2D

Understandably, a task with a 2D map was most often started with orientating oneself to the virtual space. One approach in solving alignment was to match the *axes* of the spaces:

> (1) "You had to understand how the map is rotated, and where I am myself."

A tactic to achieve alignment was to *rotate the map view horizontally a full circle*, or rotate yourself with the map in your hand, in order to find an easily matchable orientation. *Landmarks* were often searched here but they were often difficult to recognize from the 2D view:

> (2) "I looked around if there were any landmarks. […] 2D map view was difficult because you couldn't see any landmark."

Because matching of axes and finding landmarks was difficult in 2D map view, users often zoomed closer to street-level (see Figure 7) to find better recognizable cues:

> (3) "I moved lower with the device to see what is there around, what kind of environment it is… what you see from here and what should you see from here."

At times zooming and changing the viewpoint caused even more problems for orientation, as in the following case, where the user lost or forgot the location of the sought-for target:

> (4) "So I am searching for Stockmann [a department store]; It is over there. Let's go back down… I got lost."

An alternative tactic to improve recognizability was to rotate the view vertically to tilt it so that façades of buildings can be seen and therefore actually changing from 2D to 3D map view (see Figure 7):

> (5) "Looking downward like this I can see more directly onwards, I get a perspective of where I am. We are going to that direction. That is the Osuuspankki [bank] building over there."

From the vertically tilted or closer-to-ground views, a set of cues, the *colors of close-by buildings*, was made available:

> (6) "I looked where this red house next to us resides. […] That was quite a good landmark."

Often, the initially selected class of cues (e.g., colors, shapes) guided the movement:

> (7) "Ok, we're there ourselves. It [the target] is like where you go downwards and to that direction, that yellow house over there, and the yellow building, and after that a grey building; we should go there."

Sometimes the color was so distinctive that recognizing the building was straightforward without movement:

> (8) "It [the target] is the corner of that yellow building."

Users adopted their strategies according to problems they ran into. The following quote, for example, illustrates how the initial strategy (color) was changed to distance and block shape when it turned out that the color of the building on the map did not correlate with the color of the building in PE:

(9) "Oh my god, that is apparently a more yellow and a two-storey building… It could be that… No, we definitely go too far there. That one is the red building, which means that the yellow one is actually that one, which does not look like it at all, and there's no street after that like there."

The *least* liked class of cues in a plan view was the shape of blocks and street crossings. One user, for example, told that he could not recognize the target building from 2D and dove to 3D instead of using the shape of the crossing streets as the orientation cue. However, the aerial shape of the environment was useful when there were salient and indicative shapes such as open squares and the like, as in the following excerpt:

(10) "I first tried to turn around a bit to find something recognizable here. I finally recognized Esplanadi and that building over there because there's this kind of spire there."

Even when those were found, users sometimes had to zoom in to street-level view to make sure that the hypothesis is correct:

(11) "I guessed from 2D the target using Havis Amanda [a statue] as the landmark. However I went closer to make sure [of the guess]."

Salient cues were often *persevered* as strategies even when the users changed from one view to another. It sometimes happened that particular cues, such as the shapes of buildings, were not visible in the other view:

(12) "I lost my self here on the map. […] The 'spire' of the last task was not visible in this view so I thought I was in a different location."

4.2 3D

In general, comments on 3D were much more positive than with 2D and less troubleshooting talk was observed, especially in proximal mapping tasks:

(13) "I did not need to do anything there but to turn a little."

As with 2D, salient features of the built environment (i.e., easily recognizable and different from surroundings) were used extensively:

(14) "I recognized it because it was the smallest building."

As with 2D, colors and geometrical shapes were the most commonly used cues:

(15) "I inferred this [the target] from the color of this building and from the shape of the roof of that building. […] I recognized this street because of this low building over there."

Also, statues and shapes of areas were used:

(16) "I started with Havis Amanda [a statue] and Kappeli [a café] and could use them to infer the direction [of the target]. I also used the elongated shape of the Esplanadi park."

In contrast to 2D, surprisingly detailed architectural features were used as salient cues:

(17) "I used the arc of the windows in the first floor to recognize the building."

(18) "I was disappointed that the [jigsaw] shape of the Aleksi building was not modeled, because I had learned to trust to the textures shown on the display."

As with 2D, search for salient features was a poor strategy when those features were *not* modeled in the mobile map:

(19) "I'm trying to locate the round structure of the building. Let's turn around. Yes… No-ooh."

The most severe problem occurred when features were modeled inaccurately or omitted completely. However, often the features were also misinterpreted, for example, one user thought he was looking at the same crossing in the map that he was seeing in front of him. Actually, the one on the mobile map was directly behind him. There were also troubles with distinctive buildings when the users could not identify their relative position to the building. In these cases, some users circled in the area on the map or around the actual building to recognize its orientation:

(20) "I used the Railway Station as a landmark. However, I had to turn map around to see that we are on this side and not on the other side of the square."

As with 2D, users often engaged in search for the landmarks they knew of. However, this was helpful only when the memories of their locations were accurate enough:

(21) "I remembered that Fazer [a café] is on the other side of the street and thought I found the wrong location even though I had come to the correct place."

The corresponding action in 3D to circling in 2D was walking back and forth to take a peek behind corners or to rise to a level where wider perspective can be achieved:

(22) "I looked at [a church] from the map and thought what streets were parallel to its stairs. I wondered in my mind which street it could be. When I panned [the map] I saw the statue on Senaatintori [a square] and noticed that the street is [in fact] located beyond the statue."

Navigation to *remote* places with 3D was often difficult due to the additional burden of having to remember one's own location. Some users divided the problem by attempting to first identify a milestone that they knew of and walked directly there, and *then* reoriented within the map. A threat for this strategy was losing the target while moving:

(23) "I lost myself when I started to zoom out the map [arose from street-level]. Before, I was looking for the target building's façade but could not recognize it up from the plan view. I could not follow [the transition between the views]."

5 Toward A Model of Mobile Map Interaction

On the quantitative side, the results show that 3D was quicker (23% in task completion times, less restarts within the tasks) and was associated with lower workload (53% lower workload index). 3D surpassed 2D particularly in remote navigation tasks (62% for 3D in the workload measure). Several reasons stemming from the particular design of the two viewports can be pointed out to explain the apparent superiority of 3D. For example, our 2D did not present street names typically used in 2D maps. With this feature, performance with 2D could have

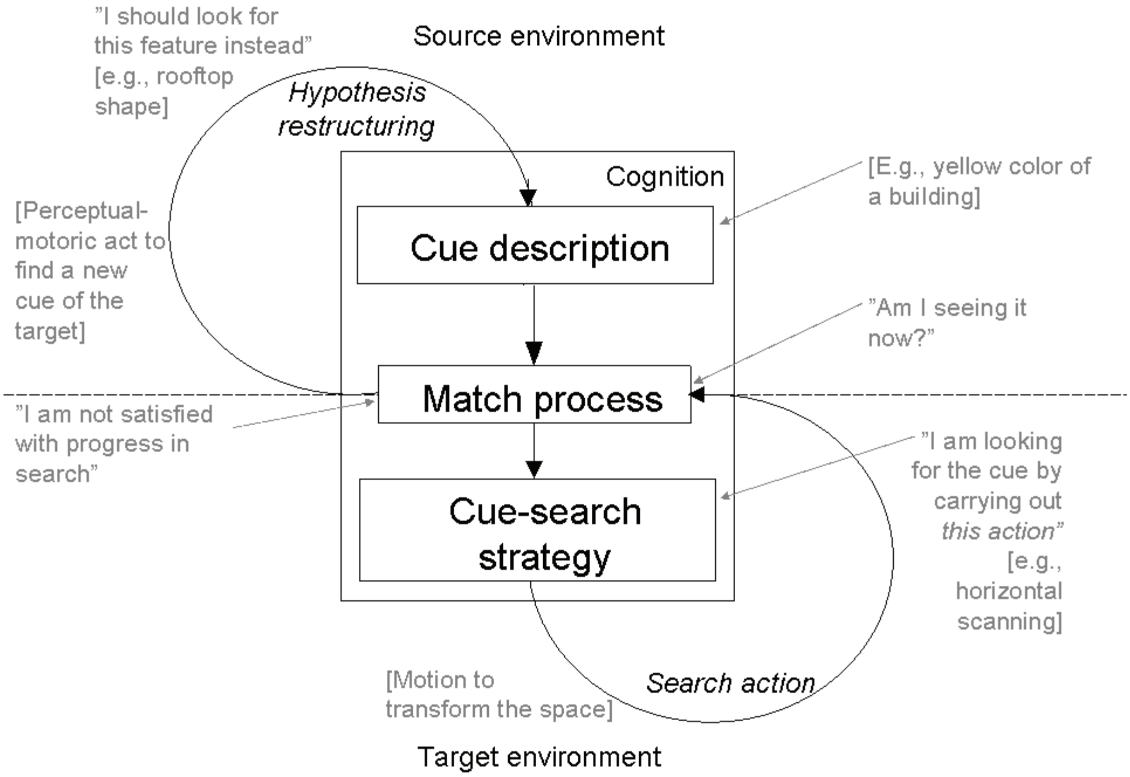

Figure 8. A model of pragmatic action in searching for targets with mobile maps.

been closer to 3D. However, the purpose of this paper was not to examine “which map type is better”, but to understand interaction with mobile maps in general. Therefore, in what follows, we put forward an explanatory model of how users interact with mobile maps.

5.1 Pragmatic Action

In the model presented in this study (see Figure 8), search for targets with mobile maps is described as pragmatic action.

In the core of our model is a *match process* that takes care of the comparison between the target description (kept in mind) and the perceived space. Any on-going comparison can be conceived as a testing of a working hypothesis. The match process itself is insufficient without action that transforms the space. Therefore, the model involves two action-perception loops called a) *hypothesis-re-structuring* and b) *search-action*. The former involves acquiring a target description by a perceptual/motor act aiming for cue extraction in the source environment (e.g., noticing that the target building is the lowest one). The latter involves acting based on a selected strategy for searching for the cue (e.g., rooftop-related cues are searched by horizontal rotation while keeping view toward rooftops). In the empirical results sections, we saw how different these strategies can be between 2D on the one hand, often focusing on plan views, and 3D on the other hand, often focusing on colors of the building horizon.

According to the model, the user then acts according to the selected strategy in order to *re-position* oneself in the environment, perceiving its new state, matching, re-position, and so on. When the current strategy does not produce satisfactory matches (for an example, see Quote 9), it has to be changed, and this triggers again hypothesis-restructuring. This does not always happen internally, but is often manifested in the subject returning into a position where another description of the target can be extracted/elaborated.

The cognitive "atom" of solving the mapping problem is here called the *cue*, a perceptual entity used in the construction of a referential relationship between the two spaces. As reported, there are several cues that can be used: landmarks (statues, salient buildings, their shapes, visual details etc.), roads (shapes, types), intersections, landscapes (types, aerial shape), districts and areas (names, aerial shapes, types), distances (between and to places), and topologies (relative positions of objects and locations), and their combinations (e.g., the yellow building with pillars). What is essential is that the selected cue (e.g., crossing shape) guides the user to act in the space in a certain way, not arbitrarily. Every cue selected has some strategies that are more naturally associated with it, and some less naturally (see Quote 12). As reported, we observed that users are sometimes reluctant to change a suboptimal, or even wrong strategy based on a supposedly-salient cue, and this leads to repetition of the same search pattern. This observation implies that if a mobile map involves some features that are effective in distinguishing objects, it *also* has to support interaction means to allow search of these objects. This links interaction design to the features of the graphical model and will be further elaborated later.

The model of pragmatic action bears resemblance to that of Jul and Furnas (1997), especially using motion as part of search, assessing stimuli from the perspective of the goal and strategy, and re-structuring the strategy when it fails. However, our model describes how and why users act in the source environment during the process of re-defining the cue description, while their model does not cover this space (only the target space) because it is constructed for mono-spatial VEs. Moreover, terminology and assumptions differ. For example, Jul and Furnas assume that *cognitive maps* are utilized in connecting the strategy to the motion, whereas our model does not fix the format of spatial representation; we believe that episodic and even semantic memories are also drawn from. More critically, previous models cannot easily account for the epistemic actions listed in the following subsection.

5.2 Epistemic Action

Even though each trial started with a specific viewport (street-level or top-down), the subjects could move freely in the space. *Not* restricting user movement turned out to allow a rich variety of strategies to be observed. We report the behaviors, ten in total, interpreted as candidates for epistemic rather than pragmatic action. Some of them are illustrated in Figure 7.

Three kinds of epistemic action were recognized that were *independent* of the type of the map used (3D or 2D):

1. *Scanning the immediate context of the target*, in order to elaborate the description of the target held in working memory (e.g., by looking at colors of neighboring buildings).
2. *Scanning for landmarks* such as statues, in order to elaborate the description of the target, or to utilize of prior semantic and spatial knowledge. Typically involves larger area searched than in epistemic action number 1.

3. *Egocentric positioning* (locating oneself in the map), in order to utilize the (stronger) representations of the PE in searching for and interpreting stimuli from the VE.

Three types of epistemic action could be discerned that were *specific to the 2D view*:

4. *Rotating the view vertically,* thus gaining perspective of the buildings' façades and rooftops. This enabled the use of those features in cue-selection.
5. *Zooming* from the top-down view to the street-level to acquire a position where the target can be matched to the subject's perceived surroundings, thus transforming the cognitive task from mental rotation to direct recognition.
6. *Rotating* the mobile device physically in order to transform the top-down view to a position where the shape of the area or street crossings can be directly recognized.

There were four epistemic actions *specific to 3D*, some of which are analogous to those of 2D:

7. *Circling* (walking around a full circle in VE) and viewing the surrounding buildings, thus creating initial mental representations of their façades to support search later on.
8. *Exploring* the proximal area (e.g., moving around in an area of few blocks from the starting position, looking around) in the first experimental trial to familiarize oneself with the model, thus reducing search costs in the subsequent trials.
9. *Peeking* around a corner to enumerate action alternatives and familiarize oneself with the surroundings, thus elaborating a representation of the vicinity and reducing uncertainty in the upcoming trial.
10. *Rotating*: the view is quickly rotated horizontally in the beginning of the trial to see the surrounding buildings. May serve egocentric positioning, landmark search etc. Similar to points 7, 8, and 9.

By definition, epistemic actions do *not* directly contribute to the search process by transforming the problem space. Rather, their objective is in changing the performing agent itself. But what is actually achieved from the perspective of the task of finding a given target? We propose three main *functions* of those epistemic actions on pragmatic action. This discussion returns us to the challenges of alignment, scale, and spatial updating reviewed in the beginning of this paper.

Improving cue descriptions. In epistemic actions 1 and 4 (and perhaps to some extent in 7 and 10) the user scans the target and its immediate neighborhood. This can result in elaboration of the representation of the target held in mind, which facilitates the subsequent match process by providing a richer account of the target to match for. Thus, the immediate benefit of those epistemic actions lies in the reduction of uncertainty.

Improving the match process. One key bottleneck for the match process is posed by the limited capacity of human working memory system (Baddeley, 1986). The sought-for cue has to be kept in mind together with other relevant information (e.g., its position), and this representation needs to be filled in, updated, and augmented during a piecemeal and fragmented course of interaction. Needless to say, the keyhole property of mobile devices accentuates this problem. One way to overcome this is to link the representation of a cue to representations of the environment kept in long-term memory (see Hollingworth, 2005), thus creating more possibilities to map environmental stimuli back to internal representations. Along these lines of thought, the epistemic actions such as 2, 3, and 7 through 10 can be partly explained as attempts to shift the processing burden from the limited short-term memory to long-term memory. Because VEs entail weaker efference copy and proprioceptive feedback from that provided by head or whole-body movement in PEs, and because vestibular information provided from head and body rotations is unavailable, it is plausible that representations of PE are stronger, richer, quicker to access and are thus preferred. Further, because our subjects were familiar with the city center of Helsinki (but not necessarily with the particular targets of the study), they had less uncertainty about the PE than the VE. Therefore, it is understandable that some of the epistemic actions were carried out first and foremost in the purpose of reducing complexity in the forthcoming match process. (The situation would change were the subjects not familiar with the city but well pre-trained with the model.) The better support of 3D for aligning non-target elements of the VE to those of the PE, we believe, may also explain the observed large benefit of 3D to 2D in mental workload measures. In sum, the match process can be improved by creating richer representations of the VE, by employing prior knowledge or by constructing it on-line.

Ideally, the match process would not rely on "imperfect knowledge in the head" but only "perfect knowledge in the world" (Gray and Fu, 2004). That is, only small or no burden on working memory would be posed and the match was based solely on recognition. In general, when people perceive an object, they can very quickly decide whether they have encountered the object before (e.g., Lamberts, Brockdorff, & Heit, 2002), and recognition memory for objects is known to be both vast and surprisingly accurate. Similar to the rotation of a zoid in Kirsh and Maglio's research (1994) the purpose of epistemic action 6 was to transform matching by mental rotation to matching by recognition. Similarly, familiarity of the street-level view / façades can be utilized as in the diving action (5). Thus, switching from mental rotation to recognition reduces uncertainty and decreases time complexity in the actual task.

Improving search strategy and action. In addition to improving the match process, richer representations of the environment can of course, and do, participate in strategy selection and implementation of search action. Having a hypothesis of the target's position in relation to a landmark helps decisions on where to look for and make exhaustive scanning unnecessary. Such epistemic actions as exploring and peeking in the 3D condition, in effect, enhanced the user's ability to identify where to search for the target by narrowing the number of alternatives. Furthermore, the lower support of our 2D viewport for utilizing prior knowledge (e.g., street names missing) might explain the observed decline in the time needed for finding targets.

6 Design Implications

When designing mobile maps, designers have to implement both cartographic (amount of information presented, symbolization, generalization, simplification) as well as interaction solutions. This paper contributes to these efforts in two ways. First, the model of pragmatic action proposed in this paper suggests that one way to minimize the possibility of choosing a failing strategy is to guide and direct the user to utilize such cues that are known to be efficient. Similarly, designers can deliberately make it *more* difficult to use those cues that are inefficient. This design strategy is here called *guidance*. Second, the model of epistemic action suggests supporting 1) elaboration of target description, 2) the construction of a representation of the area in the PE and 3) utilization of prior knowledge, and 4) transforming the task from mental to perceptual. However, if guidance was optimally effective (unlike in the map utilized in this study), one could argue that users would not need to relapse to epistemic action and other "corrective" behaviors. This, we believe, is not the case. Because of notable individual differences in representing the environment, the use of cues and landmarks (e.g., Thorndyke & Hayes-Roth, 1982; Waller, 1999), and because of the situationally varying information needs, the best solutions are those that support flexible switches between efficient strategies.

In what follows, we will explain our on-going attempts along these lines of thought. Many of the following examples have already being implemented and reported in Nurminen and Oulasvirta (in press).

6.1 Views

Because of the small screen estate of mobile devices, many informative views cannot be presented simultaneously. We have approached the problems of perspective-change and guidance by providing quick transitions between informationally complementary views, each associated with different interaction modes designed particularly for finding those cues that are informative in that view. In the improved version, three viewports are provided: street level, over the rooftops overall view (bird view), and top-down view, looking directly downward from the sky. The bird view is positioned above roof level, with viewing direction slightly tilted down, providing an overall view from the vicinity to the horizon. The street level provides a first-person view, similar to the user's perspective. Figure 11 shows the new bird's eye view and the top-down view. A virtual button is assigned to allow a fast *animated* switch between these positions and free navigation provided as an option for advanced users.

6.2 Tracks

At street level, movement is discretized and restricted to a set of invisible *tracks*, the purpose of which is to focus users' attention to informational aspects of the buildings and to prevent movement to uninformative areas between buildings. Single button-pushes advance the viewpoint in animated steps or rotate to pre-defined angles. The tracks keep the user in the middle of the street, looking forward, minimizing unnecessary maneuvering. Rotation at street level is performed in 30° animated turns, while simultaneously moving the viewpoint farther from the building in view. At 90°, the viewpoint has practically entered the building behind (Figure 9), to allow a better view of the opposing building. During this operation, the collision detection is turned off and backface culling removes the backface of the closest façade automatically. In crossings, similar angular discretization is applied, but the viewpoint is kept in the middle of the crossing. There are usually two views oriented between two alternate roads at 30° and 60°, to which the viewpoint is animated when user pushes the rotate button (Figure 10). Pushing the forward button in such a rotated view causes the user to enter the track that is better in view. Interaction related to movement is stored in an *event stack*, so the user can perform several operations in a row, and the system will respond as fast as it can.

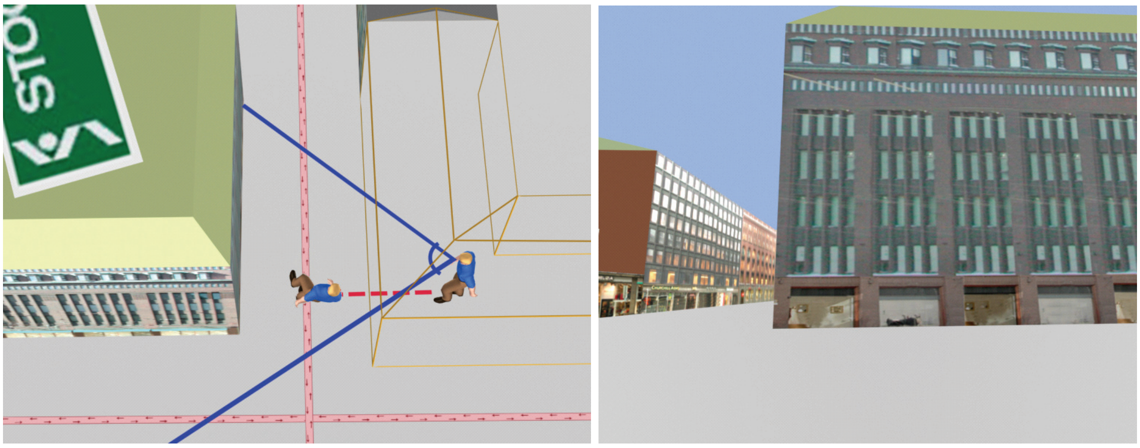

Figure 9. Virtual rails keep the user in the middle of the street (left). When rotating, distance to the opposing façade is adjusted (left) to provide a better view (right).

6.3 Positioning

Technological positioning is important from the perspective of supporting egocentric positioning, one of the epistemic actions observed with both 2D and 3D. For our current version, GPS positioning via a Bluetooth GPS device was developed, aiming at automatic support for the user to solve the mapping problem. In our solution, GPS information is used to automatically position and orient the display. When the user wants to travel away from the current position, their position in the real world is visualized on the screen by a large red dot at street level and a pointing arrow. In principle, the view can follow the GPS position, but due to the poor GPS accuracy in a city environment, it results in constant warping of the viewpoint. Therefore, this feature is selectable by the user. Currently, we are investigating means to minimize this problem, by for example forcing the view to the nearest road in the *track* mode.

6.4 Routing

Route visualizations help the user, ideally, by rendering the problem of wayfinding to that of route-following, or by externalizing the already traversed route and releasing resources from waykeeping. A routing capability was developed by creating a topological model from 2D street vector data and utilizing the *A** search algorithm. The start and end points are specified by street addresses or the current GPS position in menus, or by selecting (tapping) points in 2D or 3D map views. These points are automatically translated to the nearest streets. After specifying the points, a guiding mode is activated (Figure 11) where two arrows appear, pointing to the start (yellow) and end positions (green). The arrows are outlined when the target points are behind buildings and solid when the points are in view. The distance in meters from the current point to the start and end points is printed above the respective arrows. A polyline is also drawn along the calculated route. In the bird's eye view, it is made semitransparent and extruded to the rooftops. At street level and at bird's eye view, it is simply a polyline with route points every 40 meters, visualized as small arrows pointing at the target direction.

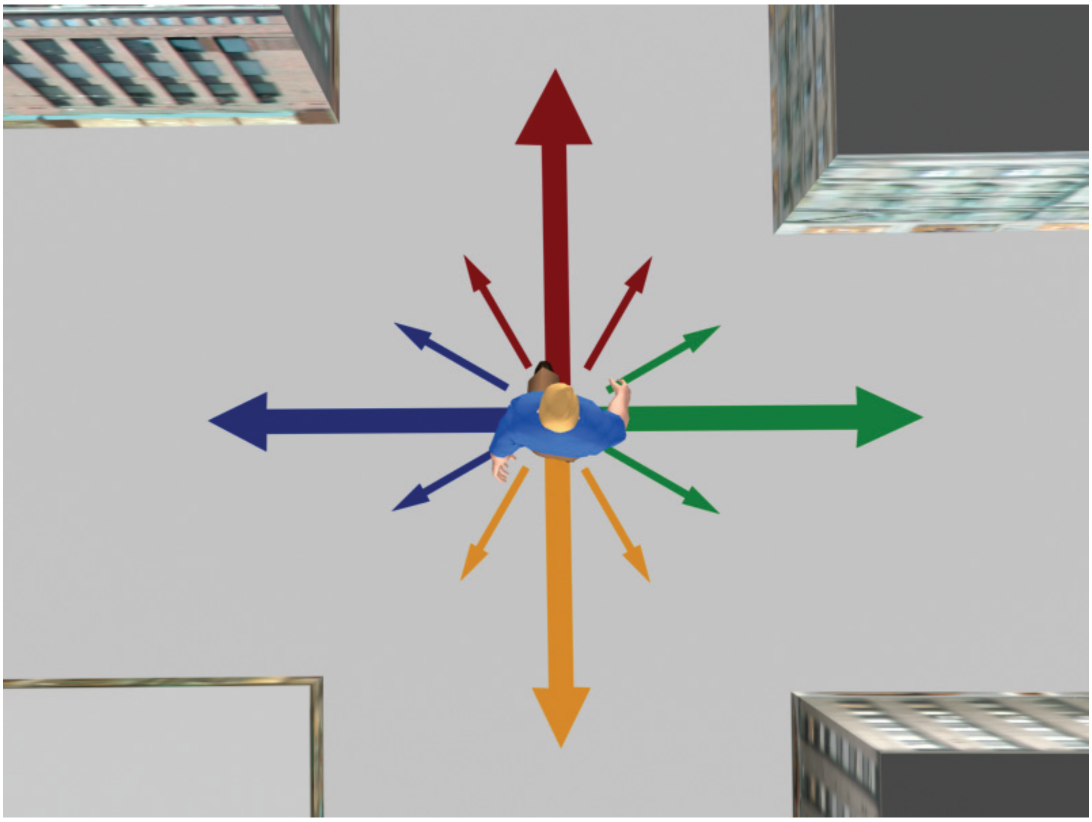

Figure 10. Possible viewing and movement directions in a crossing.

Figure 11. Guiding mode. Route visualization in bird's eye and top-down views.

When navigating, the forward and back buttons animate the viewpoint along the route, to the next or previous route point. The rotate buttons rotate the viewpoint around the current route point. The yellow begin arrow and the green end arrow virtual buttons animate the view to the beginning or end of the route. In the free navigation mode (4 DOFs), available as a menu option for experts, the route remains visualized unless deleted in the menus.

6.5 Street Names

The street vector data was also used to generate a 2D vector map, with street names rendered along the streets. The complexity of the vector map geometry increases as the viewpoint is zoomed closer. Similarly, the street name layouts follow the changing street geometry. The 2D map can be selected through menus or with a specific virtual button. The routing engine is also used to resolve the closest street address. The user can point anywhere in the map area, and the closest point at street is visualized with a dot, and the street address written in the top of the screen. In the 3D view, it is also possible to enable automatic street addressing. In this case, the current viewpoint position is constantly fed to the routing engine that returns the current closest-matching address, which is kept updated at the top of the screen.

6.6 Compass

In addition, for North-South orientation, a *compass reading* is visualized in the top of the screen to help the user to 1) utilize prior survey knowledge from map artifacts and 2) align representations of the PE with representations of the VE.

Acknowledgements

m-Loma is developed at the INIT Laboratory of the Helsinki University of Technology. We thank Ville Helin, Nina Koivisto, Paula Löppönen, Mikko Martikainen, and Mikko Rasa for their generous help. Ville Tikka and Lassi Liikkanen participated to an earlier version of this manuscript published as a workshop paper at MobileHCI 2005. The first author was funded by the Academy of Finland projects Context (2002-2005) and ContextCues (2006-2009).

References

Abowd, G.D, Atkeson, C.G., Hong, J., Long, S., Kooper, R. & Pinkerton, M. (1997). Cyberguide: A mobile context-aware tour guide. ACM Wireless Networks, 3, 421-433.

Airey, J.M. (1991). Increasing update rates in the building walkthrough system with automatic model-space subdivision and visible set calculations. An unpublished Doctoral Dissertation, University of North Carolina, Chapel Hill.

Baddeley, A., 1986. Working Memory. Cambridge University Press, Cambridge, UK.

Baudisch, P, & Rosenholtz, R. (2003). Halo: A technique for visualizing off-screen locations. CHI Letters, 5 (1), 481-491.

Bessa, M., Coelho, A., & Chalmers, A. (2005). Alternate feature location for rapid navigation using a 3D map on a mobile device. In the Proceedings of the 3rd international conference on Mobile and Ubiquitous Multimedia (MUM'05), ACM Press, New York, pp. 5-9.

Burigat, S., & Chittaro, L. (2005). Location-aware visualization of VRML models in GPS-based mobile guides. In Proceedings of the 10th International Conference on 3D Web Technology (Web3D 2005), ACM Press, New York, pp. 57-64.

Camic, P.M., Rhodes, J.E., & Yardley, L. (2003). Naming the stars: Integrating qualitative methods into psychological research. In: P.M. Camic, J.E. Rhodes, & L. Yardley, Qualitative Research in Psychology. Washington DC: American Psychological Association, pp. 3-17.

Cheverst, K., Davies, N., Mitchell, K., Friday, A., & Efstratiou, C. (2000). Developing a context-aware electronic tourist guide: Some issues and experiences. In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 2000), ACM Press, New York, pp. 17-24.

Darken, R.P., & Sibert, J.L. (1993). A toolset for navigation in virtual environment. In the Proceedings of the ACM Symposium on User Interface Software & Technology '93 (UIST '93), ACM Press, New York, pp. 157-165.

Darken, R.P., & Sibert, J.L. (1996). Navigating large virtual spaces. International Journal of Human-Computer Interaction, 8, 49-71.

Garsoffky, B., Schwan, S., & Hesse, F.W. (2002). Viewpoint dependency in the recognition of dynamic scenes. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28 (6), 1035-1050.

Golledge, R.G. (1999). Human wayfinding and cognitive maps. In R.G. Golledge (Ed.), Wayfinding behavior. Cognitive mapping and other spatial processes. John Hopkins University Press, Baltimore.

Gray, W.D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. Cognitive Science, 28 (3), 359-382.

Heidmann, F., Hermann, F, & Peissner, M. (2003). Interactive maps on mobile, location-based systems: Design solutions and usability testing. In Proceedings of the 21st International Cartographic Conference (ICC), 'Cartographic Renaissance', Durban, South Africa, CD-ROM, pp. 1299-1305.

Hollingworth, A. (2005). The relationship between online visual representations of a scene and long-term scene memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31 (3), 396-411.

Jul, S., & Furnas, G.W. (1997). Navigation in Electronic Worlds: A CHI 97 Workshop. SIGCHI Bulletin, 29 (4), 44-49.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. Cognitive Science, 18, 513-549.

Laakso, K. (2002). Evaluating the use of navigable three-dimensional maps in mobile devices. An unpublished Master's Thesis, Helsinki University of Technology, Department of Electrical and Communications Engineering.

Lamberts, K., Brockdorff, N., & Heit, E. (2002). Perceptual processes in matching and recognizing of complex pictures. Journal of Experimental Psychology: Human Perception and Performance, 28 (5), 1176-1191.

Levine, M. (1982). You-are-here maps: Psychological considerations. Environment and Behavior, 14 (2), 221-237.

Levine, M., Jankovic, I.N., & Palij, M. (1982). Principles of spatial problem solving. Journal of Experimental Psychology: General, 111 (2), 157-175.

Milgram, P., & Conquhoun Jr., H. (1999). A taxonomy of real and virtual world display integration. In Mixed reality: Merging Real and Virtual Environments. Berlin: Springer-Verlag, pp. 1-16.

Mou, W., & McNamara, T.P. (2002). Intrinsic frames of reference in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 162–170.

Nothegger, C., Winter, S., & Raubal, M. (2004). Selection of salient features for route directions. Spatial Cognition and Computation, 4 (2), 113-136.

Nurminen, A. Oulasvirta, A. (in press). Designing interactions for navigation in 3D mobile maps. To appear in L. Meng, A. Zipf (Eds.) Mobile Maps, Springer-Verlag.

Presson, C.C., DeLange, N., & Hazelrigg, M.D. (1989). Orientation specificity in spatial memory: What makes a path different from a map of the path? Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 887–897.

Rakkolainen, I., & T. Vainio (2001). A 3D city info for mobile users. Computers and Graphics, Special Issue on Multimedia Appliances, 25 (4), 619-625.

Rantanen, M., Oulasvirta, A., Blom, J., Tiitta, S., & Mäntylä, M. (2004). InfoRadar: Group and public messaging in mobile context. In Proceedings of the Nordic conference on Computer-Human Interaction (NordiCHI'04), ACM Press, pp. 131-140.

Roskos-Ewoldsen, B., McNamara, T.P., Shelton, A.L., & Carr, W. (1998). Mental representations of large and small spatial layouts are orientation dependent. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 215–226.

Ruddle, R.A., Payne, S.J., Jones, D.M. (1999). The effects of maps on navigation and search strategies in very large-scale virtual environments. Journal of Experimental Psychology: Applied, 5 (1), 54-75.

Sorrows, M.E., & Hirtle, S.C., (1999). The nature of landmarks for real and electronic spaces. In Freksa, C. and D.M. Mark (Eds.), Spatial Information Theory. Lecture Notes in Computer Science, Vol. 1661. Springer, Berlin, pp. 37-50.

Stuart, R. (1996). The Design of Virtual Environments. McGraw-Hill.

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

Trevisan, D.G., Vanderdonckt, J., & Macq, B.M. (2004). Conceptualising mixed spaces of interaction for designing continuous interaction. Virtual Reality, 8 (2), 83-95.

Vainio, T., & Kotala, O. (2002). Developing 3D information systems for mobile users: Some usability issues. In the Proceedings of the The Second Nordic Conference on Human-Computer Interaction (NordiCHI'02), ACM Press, New York.

Vidulich, M.A., & Tsang, P.S. (1986). Collecting NASA Workload Ratings: A Paper and Pencil Package. NASA Ames Research, Moffett Field, CA.

Waller, D.A. (1999). An assessment of individual differences in spatial knowledge of real and virtual environments. Unpublished doctoral dissertation, University of Washington.

Wang, R.F., & Brockmole, J.R. (2003). Human navigation in nested environments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29 (3), 398-404.

Witmer, B.G., Sadowski, W.J., & Finkelstein, N.M. (2002). VE-based training strategies for acquiring survey knowledge. Presence: Teleoperators and Virtual Environments, 11, 1–18.

Woods, D.D., & Watts, J.C. (1997). How not to have to navigate through too many displays. In Helander, M. G., Landauer, T. K. and Prabhu, P. (Eds.), Handbook of Human-Computer Interaction, 2nd edition. Elsevier Science, Amsterdam.

Wraga, M. (2003). Thinking outside the body: An advantage for spatial updating during imagine versus physical self-rotation. Journal of Experimental Psychology: Leaning, Memory, and Cognition, 29 (5), 993-1005.