This is a pre-print version of the article. Please cite the original article at http://dx.doi.org/10.1111/tgis.12067

Kiefer, P., Giannopoulos, I., and Raubal, M. (2014) Where Am I? Investigating Map Matching During Self-Localization With Mobile Eye Tracking in an Urban Environment. Transactions in GIS, 18(5)

# Where Am I? Investigating Map Matching During Self-Localization With Mobile Eye Tracking in an Urban Environment

**Peter Kiefer, Ioannis Giannopoulos and Martin Raubal**

*Institute of Cartography and Geoinformation, ETH Zurich*

*Wolfgang-Pauli-Str. 15*

*CH-8093 Zurich, Switzerland*

*{pekiefer, igiannopoulos, mraubal}@ethz.ch*

**Abstract**

Self-localization is the process of identifying one's current position on a map, and it is a crucial part of any wayfinding process. During self-localization the wayfinder matches visually perceptible features of the environment, such as landmarks, with map symbols to constrain potential locations on the map. The success of this visual matching process constitutes an important factor for the success of self-localization. In this research we aim at observing the visual matching process between environment and map during self-localization with real-world mobile eye tracking. We report on one orientation and one self-localization experiment, both in an outdoor urban environment. The gaze data collected during the experiments show that successful participants put significantly more visual attention to those symbols on the map that were helpful in the given situation than unsuccessful participants. A sequence analysis revealed that they also had significantly more switches of visual attention between map symbols and their corresponding landmarks in the environment, which suggests they were following a more effective self-localization strategy.

## 1 Introduction

Wayfinding is one of the main categories of spatial-cognitive tasks that people engage in on a daily basis. It can be described as purposeful, directed, and motivated movement from an origin to a specific distant destination that cannot be directly perceived by the traveler (Golledge 1999). People must utilize various cognitive and spatial abilities in order to accomplish the specific tasks included in wayfinding. These consist of creating and choosing a route, establishing and maintaining orientation with respect to one's starting location or with reference to external features or places, and recognizing landmarks and their relation to other landmarks or features in the environment (Montello and Raubal 2012). In addition, wayfinding may require a more-or-less accurate judgment of distances and directions, remembering a sequence of turns, and remembering the locations of objects and events.

Spatial orientation and self-localization are critical tasks during wayfinding in an urban environment. *Spatial orientation* refers to deciding which direction in a spatial scene one is facing, given a spatial reference frame, mostly a cognitive or real map (Davies and Peebles 2010). It necessitates recognizing one's surroundings, being able to utilize landmarks, and maintaining one's orientation for the duration of the wayfinding process (Dudchenko 2010). Without a sense of orientation, i.e. knowing where we are in relation to the location of other objects or our own previous location, we may get lost. *Self-localization* is closely linked to spatial orientation and refers to identifying one's position in a spatial reference frame, given spatial scenery. According to Peebles et al. self-localization is a "more complex combined task [. . .] which inevitably includes simultaneous orientation to some extent" (Peebles et al. 2007, p. 391). Literature suggests that self-localization is typically performed by "using local cues which are visible from our current location. These cues could be the geometry, or landmarks displayed in a map" (Meilinger et al. 2007, p. 386)[1]. While the influence of geometry on wayfinding decisions has drawn some interest (Davies and Peebles 2010; Emo 2012; Meilinger et al. 2007; Peebles et al. 2007), studies suggest that, "in geographically realistic contexts, visible salient landmarks bias people away from using optimal geometry-matching strategies" (Davies and Peebles 2010, p. 135). Achieving realism is one of the main goals of our experiments, therefore we focus in this article on orientation and self-localization based on landmarks. Landmarks are usually associated with prominence and salient features in the environment (Raubal and Winter 2002), they support the construction of a mental representation of space and they are used in the communication of wayfinding directions, preferably at decision points (Denis et al. 2007).

Over the years, researchers have run numerous studies with the goal of providing insight into people's spatial thinking and reasoning during wayfinding, such as their employed strategies (Kato and Takeuchi 2003). Different methods, e.g. interviews, behavior observation (Meilinger et al. 2007), and cognitive map drawing (Hirtle and Jonides 1985), have been employed to assess people's wayfinding problems. However, these methods do not provide quantitative and objective means of describing cognitive processes that are mainly based on visual attention, such as looking for cues during orientation and self-localization. Vision research suggests that eye movement patterns are related to an individual's cognitive processes (Liversedge and Findlay 2000), for instance in text reading (Rayner 1998). We argue that, also for orientation and self-localization tasks, a quantitative analysis of eye movement patterns allows researchers to investigate fundamental cognitive processes such as matching a symbolic landmark representation to its corresponding object in the environment. While eye tracking has been employed in lab studies on orientation (Gunzelmann et al. 2004, 2008; Peebles et al. 2007), these can only come close to the realism of the real world. As noted by Warren et al. a methodological restriction of lab experiments is that showing photographs of spatial scenes "reduces an experience of being within a three-dimensional environment to viewing a flat, two-dimensional representation of that environment from a point outside of it" (Warren et al. 1990, p. 148). This holds especially for self-localization experiments, where a 360° immersion into the spatial scene is a requirement for orienting towards different directions. For a pure orientation experiment, in contrast, a view of 60° (Peebles et al. 2007) might be sufficient. Only few orientation and self-localization experiments were performed in real environments, indoors (Meilinger et al. 2007) or outdoors (Iachini and Logie 2003), but had different foci and did not use eye tracking.

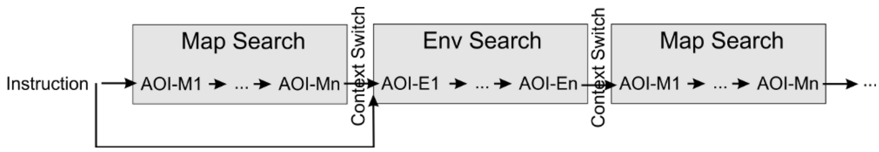



*Figure 1 Landmark-based self-localization as a sequence of map and environment search phases*

In this article we describe two experiments. In Experiment 1 (Section 4.1), participants had to judge whether an iconic map is a correct representation of the environment. In Experiment 2 (Section 4.2), participants had to mark their position on a tourist map. While Experiment 1 focuses on spatial orientation, Experiment 2 constitutes a self-localization task. In both experiments, the task could only be solved successfully by matching map symbols to their corresponding landmarks in the environment.

[1] Meilinger et al. (2007) consider the network structure (decision points, such as intersections) as a third type of cue. We subsume this under geometry, as intersections are inferred from the road network geometry.

The matching process "is probably best described as a hypothesis testing procedure, i.e., we generate a hypothesis about our current location and try to confirm or reject this hypothesis by collecting more information" (Meilinger et al. 2007, p. 386). While the cognitive processes of hypothesis creation, confirmation, and rejection remain hidden, eye tracking allows us to observe the information collection process. This process includes looking for both cues in the environment and cues on the map. Thus, it can be described as a sequence of *map search phases* and *environment search phases* with *context switches* between them (see Figure 1). It is likely that an individual with a superior orientation and self-localization strategy will, on average, have a more effective visual information collection process. In other words, he or she is likely to look for cues that are helpful for the task. Our experiments are intended to confirm this assumed relation between spatial abilities and the effectiveness of the visual information collection process, where the latter is described quantitatively by analyzing fixation patterns. This leads us to the following two research questions:

*RQ1 Do successful participants spend more visual attention on map symbols that have a visible corresponding landmark than unsuccessful participants?*

*RQ2 Do successful participants have more switches of visual attention between symbols on the map and their corresponding landmarks in the environment than unsuccessful participants?*

The results demonstrate that the gaze behavior of successful and unsuccessful participants differs, and that this difference can be described by statistics on gaze distribution and switches of visual attention. Successful participants paid significantly more visual attention to symbols on the map helpful in the given situation than unsuccessful participants. Moreover, the ratio between visual attention given to helpful and unhelpful map symbols was significantly different between successful and unsuccessful participants. In addition, successful participants had significantly more attention switches between landmarks in the environment and their correspondents on the map. A significant correlation between task success and self-estimation of spatial abilities was found, using the Santa Barbara Sense of Direction Scale, SBSODS (Hegarty et al. 2002). Besides the theoretical impact of this research, a practical application can be seen in new interaction methods based on gaze (Giannopoulos et al. 2013).

The remainder of this article is organized as follows. Section 2 presents related work on eye tracking in Geographic Information Science (GIScience). Section 3 describes the methods and technology used to collect and process mobile eye tracking data in outdoor real-world wayfinding experiments. Section 4 presents the two self-localization experiments, including their analyses and results. In Section 5 these results are discussed with regard to our research questions and related work. Section 6 presents conclusions and proposes directions for future research.

## 2 Related Work: Eye Tracking in GIScience

Eye trackers measure a person's visual attention on a stimulus. Most eye tracking studies in GIScience have been performed in the lab using remote eye trackers, which are mounted to a monitor and record a stream of pixel coordinates the user has looked at. These basic recordings are called *gazes*. It is generally assumed that perception takes place only if gaze remains almost still for a minimum amount of time. Thus, gazes are often aggregated spatio-temporally to *fixations* (see Section 3.2.2 for algorithmic details). A transition between two fixations is called a *saccade*, which is caused by a rapid movement of the eye. A comprehensive overview on eye tracking hardware and methodology can be found in Holmqvist et al. (2011) and Duchowski (2007).

### 2.1 Eye Tracking in Cartography and Map Interaction

Eye tracking studies for cartographic stimuli range back to the 1970s and 1980s. At that time, researchers analyzed whether eye movement properties, such as the average duration of fixations, differ for different tasks and stimuli (Castner and Eastman 1984; Dobson 1977; Steinke 1987). More recent work has focused on usability aspects of interactive maps, such as effectiveness and efficiency of different map designs (Çöltekin et al. 2009, 2010), the effectiveness of label placement (Ooms et al. 2012b), cartographic animations (Opach and Nossum 2011), or user group differences (Ooms et al. 2012a). The design of a map and the usability of the system in which it is shown are most likely to influence any wayfinding and self-localization decision, also when supported with the gaze-based assistants that could be built around findings presented here (see Section 6). In this article we are more interested in the basic cognitive process of matching map and environment, not in visual map or interaction design. The maps in our experiments were either purely iconic with a simple symbology, or taken from a popular cartographic product (Zurich tourist map).

More methodological questions are approached by another line of GIScience research on eye tracking, which is concerned with how to analyze and interpret eye tracking data, viewing them as large quantities of spatial data. Approaches consider space-time-cube visualizations of the data (Li et al. 2010), or visual analytics methods in general (Andrienko et al. 2012). In our own work, titled *Gaze Map Matching*, we proposed to match fixations algorithmically with the vector representation of a cartographic map, just as a map matching algorithm snaps GPS points to a road (Kiefer and Giannopoulos 2012). Here, we follow the classical approach to the analysis of eye tracking data based on Areas of Interests (AOI). An AOI is a polygonal area in the stimulus the researcher considers relevant for the research question at hand. If a fixation occurs in an AOI, it is generally assumed that the participant perceived the object surrounded by the AOI. Holmqvist et al. (2011) provide details on different possibilities for AOI analyses. Our approach to an AOI-based analysis is distinct due to the mobility of the participants. which allows for head movements and locomotion.

The abovementioned usability studies take a complete eye tracking recording as input (ex-post analysis). Many eye trackers, however, also allow for real-time data access, which is the principle behind gaze-assistive systems. An example of a gaze-assistive interaction concept for small-display maps is the *GeoGazemarks* concept (Giannopoulos et al. 2012), which provides the user with an aggregation of her gaze history on different zoom levels in order to facilitate orientation. Other gaze-based concepts for the interaction with maps include the work by Stellmach and Dachselt (2012) who explored different ways of using gaze for panning and zooming on a virtual globe. Our research potentially contributes to future gaze-based interaction concepts by providing insights into the relation between gaze and errors in self-localization, which could be used by a gaze-assistive system to support the user (Giannopoulos et al. 2013).

### 2.2 Eye Tracking for Spatial Decision Making and Wayfinding in the Lab

The research presented here is one of the first to perform mobile eye tracking wayfinding studies in the real world. However, wayfinding decisions, and spatial decision making in general, have previously been studied using eye tracking in the lab. Spiers and Maguire (2008) performed an experiment with London taxi drivers who had to solve wayfinding tasks in a virtual reality simulation of London. The analysis was mainly based on retrospective verbal protocols recorded while participants were watching their task performance after the task. Eye tracking was used to cross-check those parts of the protocols where the participants had stated that they had looked at some landmark. From these self-reported gazes, 94% were found in the eye tracking data (validly calibrated eye tracking data were available for nine out of 20 participants). Besides being outdoors, our approach differs from Spiers and Maguire's in that we use the gaze tracks as main data source, not as complementary data.

Wiener et al. (2012) report on a lab experiment in which participants had to navigate through a dungeon-like virtual environment. Given an image of a spatial configuration of walls and hallways, participants had to decide whether to turn left or right. After a predefined number of scenes the goal was reached, independent of the decisions taken. The dungeon scenes were in no way spatially connected. One of the main results was that participants had a gaze bias towards the eventually chosen direction. Although this provides an interesting insight into the correlation of gaze and spatial decision-making, more complex cognitive processes involved in wayfinding and self-localization, such as the search for and matching of features in the environment and on the map, were not considered.

Emo's (2012) paper, although called "Wayfinding in Real Cities", reports on a lab experiment. In contrast to Wiener et al. her stimuli consisted of photos from a real city. She was interested in how the geometry and topology of the street scene influence gaze, a question also approached by Wiener et al. As no map was involved, this is again different to our work, although geometry can also play an important role for self-localization. Photos from real city scenes were used as stimuli in Nevelsteen and Steenberghen's (2012) study which focused on how school children perceive traffic environments; this can be used by city planners to improve traffic safety. A strength of their study was the large sample size consisting of 466 elementary school children.

Gunzelmann et al. (2008) performed an eye tracking experiment in which participants had the task of identifying the observer's position on an allocentric representation of the environment, given an egocentric view of the same space. This is similar to our self-localization task, as both require reasoning about the spatial configuration of objects. However, their experimental setting was very different (fixed perspective, artificial scenery, indoor screen study), and eye tracking data were analyzed only on the level of views, not on the level of individual objects. Their main finding was related to learning: over time, participants became more efficient and had less attention switches between the two views, which indicates that an efficient strategy has been learned.

Eye movement data were also collected in the orientation study by Peebles et al. (2007). Given a 3D model of a spatial scene, participants had the task of indicating on an according "you are here" map which direction the scene is facing. However, the eye tracking data played a minor role in their analysis as they are only reported on in one sentence: "By far the most commonly reported feature used for solving the problem was 'buildings', and the eye movement patterns in the scenes with the most salient 3D landmarks [. . .] tended to strongly focus around those landmarks." (Peebles et al. 2007, p. 398f). This is consistent with our experiments, in which visual attention of successful participants was strongly focused on salient landmarks and the corresponding map symbols. In contrast to Peebles et al. (2007), we support this finding with a quantitative analysis of eye tracking data and perform our experiment in the real world.

Wayfinding aids, such as maps and location-based services (Hirtle and Raubal 2013), need to be aligned with the surroundings during orientation. This requires mental rotation – a cognitive process studied also in a larger spatial cognition context (Hegarty and Waller 2004). An orientation experiment focusing on mental rotation was performed by Gunzelmann et al. (2004). The task consisted of (mentally) rotating a given spatial configuration of "plane" and "target", with "plane" shown at the bottom, in a way that the position of "target" could be identified on an allocentric view of the scene, given the position of "plane". Participant groups were trained for two different strategies, each of which was likely to be connected with visual attention on a characteristic set of AOIs in the stimulus. Indeed, results showed a significant correlation between gaze distribution and strategy. In our study, we did not correlate visual attention with the chosen strategy, but with task success. If we assume task success to be dependent on one self-localization strategy, our results support the findings of Gunzelmann et al. (2004). Our work is still very different since we focused on a real-world setting and also analyzed fixation sequences. In addition to empirical results, Gunzelmann et al. (2004) modeled the two strategies with an ACT-R agent model (Anderson et al. 1997). As gaze data are highly unpredictable and noisy, they had to model a certain percentage of gaze data to be “off task”, i.e. randomly distributed over all AOIs. By setting this parameter to 50% they achieved a good fit between the empirical data and the model. This relatively high value could be caused by the fact that they did not cluster gazes to fixations, which would have reduced noise. Due to the complexity of the real-world stimulus we do not model our self-localization task with an agent model, such as ACT-R. For instance, we cannot assume “off-task” fixations to be equally distributed, but rather would have to consider saliency. This is beyond the scope of this article.

### 2.3 Eye Tracking for Wayfinding in the Real World

Only few wayfinding studies have used mobile eye tracking in the real world: Schuchard et al. (2006) analyzed whether signs in a nursing home were placed at the right height for subjects with mild dementia during wayfinding.

In the scope of his Ph.D. research, Pinelo Silva (2011) performed a wayfinding experiment with mobile eye tracking in central London. Participants were guided from A to B following a specific route, then had to find the shortest route from B to A, and finally had to walk the original route from A to B. In contrast to our experiments, no map was used for help, and self-localization was not investigated. Pinelo Silvas’s questions were mostly related to the built environment and how it relates to gaze behavior, such as “identifying the physical components of the built environment that are more often used by pedestrians while performing navigation tasks” (p. 158). Similar to our experiments, he reports on severe problems with the automatic pupil detection due to changes in lighting conditions. We solved this by manually checking and adjusting the pupil detection in each single frame of the video (see Section 3.2.1.). Pinelo Silva followed a different approach and manually coded the videos.

Delikostidis (2011) considered mobile eye tracking as part of a usability methodology for pedestrian navigation systems. Among other questions, he was interested in whether a pedestrian navigation system supports the user in matching the spatial representation on the device with her mental map of the environment (which he calls “geo-identification”). His use case differed from our self-localization task, as the user’s GPS position was provided. Mobile eye tracking was considered as part of the methodology, but then abandoned due to technical restrictions of the eye tracker, so none of the results was based on mobile eye tracking.

We have previously reported on an exploratory pilot wayfinding study with mobile eye tracking in Zurich (Kiefer et al. 2012a, b). First indications were found that gaze patterns on the map and in the environment differ between successful and unsuccessful participants in a wayfinding task. Although the data of the exploratory study looked promising, there were some problems, which were accounted for in the design of the experiments described in this article:

- *Separability*: Different phases of a wayfinding decision were merged into one eye tracking recording. It was not possible to tell which part of the gaze recording occurred during self-localization, map alignment, map search, or route planning.
- *High uncontrollability*: The starting position was surrounded by tram lines. Trams arrived and left frequently, therefore making viewing conditions incomparable between participants.

While outdoor studies are always less controllable than lab studies, the uncontrollability at the tram station was unacceptably high.

- *Landmark symbology*: The symbols on the map were too small, therefore gazes could not be assigned to individual landmarks on the map, given the mobile eye tracker's inaccuracy.
- *Landmarks too distant*: The landmarks visible from the starting position were too far away so that, for gazes in the distance, it was not possible to determine from the resulting video (resolution 768x576 pixels) what exactly the participant had been looking at.

Wayfinding decisions typically result in locomotion, which requires the individual to solve a number of low-level tasks, such as steering or collision avoidance. Mobile eye tracking has been used to analyze gaze behavior during locomotion and steering (Foulsham et al. 2011; Franchak and Adolph 2010; Land and Lee 1994; Vansteenkiste et al. 2013). In this article, however, we are interested in the high-level cognitive processes involved in wayfinding.

## 3 Methods and Technology

In this section we provide an introduction to the methodology used for both mobile eye tracking experiments reported on in Section 4. Table 1 provides an overview of the steps involved in the experiment and the analysis. Sections 3.1 and 3.2 describe the general approach to collect and analyze gaze data with mobile eye tracking experiments in the real world. Section 3.3 describes the specific analysis steps for quantifying the effectiveness of a visual information collection process during self-localization.

*Table 1: Overview on the steps involved in an outdoor mobile eye tracking experiment.*

| Experiment | Data Processing |
|---|---|
| Welcome, Introduction, Questionnaire | Preprocessing |
| Mount Hardware | Improve Pupil Detection |
| Calibrate | Parallax Correction |
| | Marker Detection (paper markers) |
| *For each task* | Marker Definition (environment markers) |
| Guide Participant to Start Position | Analysis |
| Calibration Checkpoints (near, far) | AOI Definition |
| Instruction | Compute AOI Gaze Sequence |
| Task | Compute Fixations |
| | Compute AOI Fixation Sequence |
| | Analyze AOI Sequences |

### 3.1 Experiment Set Up

Participants are welcomed, introduced to the experiment, and asked to fill in a questionnaire regarding their background and spatial abilities (SBSODS) (Hegarty et al. 2002). The eye tracking hardware is mounted and calibrated. Our hardware consists of the Dikablis system from Ergoneers (http://ergoneers.com/index-en.html), a monocular head-mounted mobile eye tracker (see Figure 2). It records two videos at a frame rate of 25Hz: the scene video (768 × 576 pixels, PAL, color) and the eye video (384 × 288 pixels, PAL, b&w). The eye tracker is connected via a cable to a notebook in the backpack, where both videos are stored. The eye tracker illuminates the pupil with infrared light. For studies in the sunlight, especially under varying lighting conditions, this is a major problem as the pupil detection algorithm relies on a constant and equal illumination of the eye. The problem is partially compensated through a Mexican sun hat (see Figure 2). Still, an improvement of the pupil detection is necessary during the data processing phase (see Section 3.2.1).

*Figure 2 Participant with head-mounted eye tracker. The notebook is carried in the backpack*

In contrast to remote eye tracking, which is mostly used for lab studies, mobile eye tracking enables studies in real environments. Through the freedom of head and body movements, which includes pedestrian locomotion, mobile eye tracking allows for the investigation of wayfinding processes under higher ecological validity than the conditions of a lab study. These advantages are opposed by a few problems concerning the controllability of the experiments as well as the necessary manual preprocessing. It is necessary to validate the data and also perform qualitative analyses in order to ensure validity, before starting with the analyses. Eye tracking studies in the lab are more controllable, therefore the effort for preprocessing and validation is far less extensive. However, it is an open question whether eye tracking studies in the lab have the necessary validity as compared to studies in real environments. In general there is a debate about the external validity of lab studies in many domains as summarized by Anderson et al. (1999).

Dikablis offers an automatic four-point calibration, i.e. the participant is asked to fixate four points at equal distance. One important difference between lab and outdoor studies is that the objects gazed at in the real world are typically at varying distances from the observer. This applies especially to wayfinding studies (landmarks vs. map). Monocular eye trackers, such as our system, suffer from the parallax effect, which means they can only be calibrated to one distance at one time. A far calibration will appear shifted for close distances, and vice versa. We chose a far calibration for the experiments, covering all distances larger than 10 m. Close distances are included in the analysis by post-processing the data in the parallax correction step (see Section 3.2.1). This parallax correction requires us to set at least one calibration checkpoint during the experiment, i.e. to mark an instance in the video with ultimate knowledge that the participant is fixating a given close point. We set two calibration checkpoints before each task, one at close, and one at far distance to account for minor shifts of the eye tracker.

### 3.2 Processing Eye Tracking Data

Eye tracking data come at high quantity. Experiment 2 (Section 4.2), for instance, had a total recording of 34,486 video frames over all participants (1,379 sec, or 23 min). From these data, we intend to find out for each of these frames where the participant is looking at, either on the map or in the environment (Sections 3.1 and 3.2). Second, this information must be interpreted at a higher level with respect to our research question (Section 3.3). Although the analysis can be partially automated[2], our experience showed that, due to the noisiness of the outdoor eye tracking data, much manual correction is necessary.

#### 3.2.1 Pupil detection, parallax correction, markers, and AOIs

Due to different influences in outdoor user studies, such as changing lighting conditions or mascara on eyelashes, the pupil detection algorithm integrated in the software fails frequently. In order to ensure valid data, it was necessary to go through each frame manually and adjust the pupil detection if necessary. Figure 3 displays examples of a correct pupil detection (a), two pupil detection errors (b, c), and one frame with a blink (d). In cases (b) and (c), the data analyst clicked on the correct center of the pupil with the mouse. As a next step, the parallax effect is corrected: the size of the parallax shift from the far and near calibration check points is known and removed with the Dikablis software for those parts of the task in which the participant had been looking at the map.

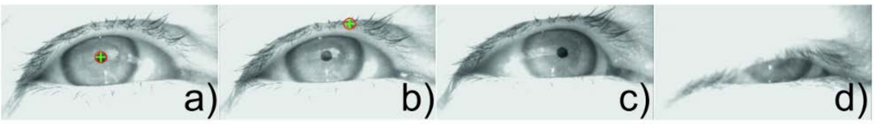



*Figure 3 Pupil detection: (a) True positive; (b) False positive (on eyelashes); (c) False negative*

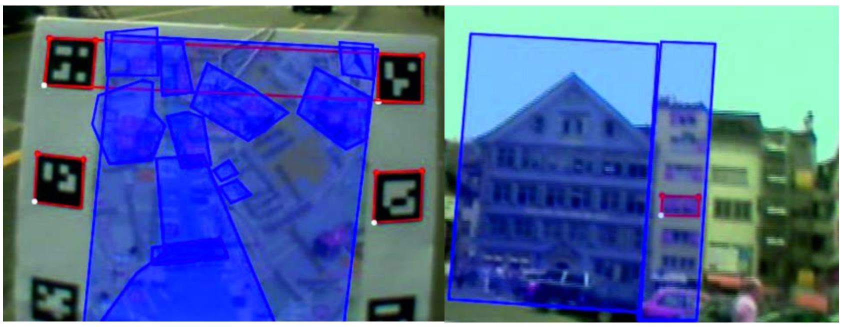


*Figure 4 Markers and AOIs. Left: Paper map with visual black&white markers and AOIs. Right: Keyframe*

Remote eye trackers return the participant's gaze in the coordinate system of the screen. From this pixel position, it is easy to automatically infer the AOI to which the participant paid visual attention. In contrast, our mobile eye tracker returns the gaze position as coordinates in the video. This cannot easily be mapped automatically to an object in the real world. In our system, this is solved with markers. Visual black-and-white markers printed on the map are recognized by the software and used

[2] In our case, with the D-LAB software belonging to the Dikablis system.

to span a plane in which AOIs can be defined relative to the markers (see Figure 4, left). AOIs around objects in the environment can be bound to so-called “keyframe” markers. The data analyst defines a keyframe marker around a distinguishable feature of the environment, such as a door or a window. The software uses a visual flow algorithm to follow these visual features in all succeeding frames until the keyframe marker leaves the visual field of the scene camera. These keyframe markers typically need to be redefined quite often because the head turns frequently during wayfinding. Figure 4 (right) displays an example of two AOIs around houses bound to a keyframe marker.

As a next step, the software computes for each frame which AOIs the gaze is hitting (a point-in-polygon operation). If several AOIs are hit, we return the deepest one in the partonomy. The output is an AOI *gaze sequence* $aseq_g = [a_1, a_2, . . . , a_n]$ which contains one AOI for each gaze. Gazes that do not hit an AOI are assigned a standard value (“NOAOI” in our case).

### 3.2.2 Fixation computation

A fixation occurs when our eyes remain relatively still for a short period of time. To compute fixations from pure gaze, it is necessary to take into account the time each gaze occurred as well as its location in terms of X, Y coordinates. As a first step we computed the fixations using a radius of 30 pixels and a time threshold of 200 milliseconds. For example, if all subsequent gazes captured in a time span of 200 milliseconds are within a radius of 30 pixels, they will form one fixation (see Figure 5).

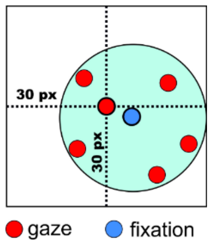



*Figure 5 Fixation computation. The gazes (red) are clustered and represented by one fixation point*

The selection of these two parameters (i.e. fixation radius and duration) for the computation of fixations has yet no established rules, which makes eye tracking studies hardly comparable. Therefore it is necessary to communicate the chosen values in order to provide more standardized and reliable results. According to Rayner, fixations occur when our eyes remain relatively still for 200–300 milliseconds (Rayner 1998). There are several publications reporting a minimum fixation duration of 100 milliseconds and others using a minimum duration of 200 milliseconds (Salvucci and Goldberg 2000; Widdel 1984). The radius threshold can be estimated from exploratory analysis of the data, where typical values are reported to be between 0.5° and 1° of visual angle (Salvucci and Goldberg 2000). We used both of these common duration thresholds, and report on both results, in order to make the results more robust. We used a minimum fixation duration threshold of 100 and 200 milliseconds, and through an exploratory analysis of the data we utilized a fixation radius of 30 pixels. For the selection of this radius we analyzed the checkpoints where participants had to fixate before each task started. Through a qualitative analysis of the videos we also ensured that no fixations were missed because of head movements, or because of smooth pursuit eye movements, which occur when a moving object is being fixated (Holmqvist et al. 2011).

In our case, each gaze point was assigned a value that specifies whether the gaze is within an AOI or not. Each fixation was assigned the value of one of the gazes that formed the fixation by selecting the most frequent AOI. For example, if the gazes that formed a fixation yielded an AOI sequence of $aseq = [a_1, a_2, a_3, a_1, a_1, a_1]$, the AOI assigned to the fixation would be “$a_1$”.

### 3.3 Measuring the Effectiveness of Information Collection in Map-Based Self-Localization

This section describes how the results of Sections 3.1 and 3.2 (the AOI sequences) are further processed to approach our specific research questions (step “Analyze AOI sequences” in Table 1). We use the AOI sequences to describe the effectiveness of the visual information collection process, which can later be correlated with the task success. This analysis is based on the visual information collection process during map-based orientation and self-localization as shown in Figure 1. In each search phase the wayfinder looks at several cues in the environment, or on the map, respectively. The elements in the stimulus, which can be considered as cues, are surrounded by AOIs (see Section 3.2.1). As we focus on self-localization based on landmarks we do not try to cover geometric cues with AOIs, but only AOIs around landmarks and their symbolic representation on the map. If a symbol on the map represents a landmark in the environment that is visible from the wayfinder’s position, we call this a *helpful symbol*. The AOIs around the two are called *corresponding AOIs*. Symbols on the map that have no correspondent visible landmark are called *unhelpful symbols*. All fixations that fall outside an AOI are assigned a dummy AOI (“NOAOI” for the environment, and “MAP” for the map). We describe the effectiveness of the visual information collection process with both, distribution and sequence measures.

Distribution measures describe the percentage of fixations that fall into certain AOIs over the whole task duration, in our case on helpful AOIs and unhelpful AOIs, respectively. An effective visual information collection process will yield a relatively high percentage on helpful AOIs, whereas ineffectiveness is characterized by either low values for both, or a relatively high value for unhelpful AOIs (which means the wayfinder based his decision mainly on landmarks not visible from his or her position). Sequence measures try to capture the procedural view of information collection: a matching process can only be successful if a symbol or landmark is still in working memory while its correspondent is looked at. In other words, in an effective process, AOI hits on two corresponding AOIs are likely to be temporally close to each other. We assume that only AOIs seen in the previous search phase remain in working memory, which means we are interested in the co-occurrence of corresponding AOIs in succeeding search phases. We call this co-occurrence *matches*, and consider three measures based on matches:

- *#matches map -> env*: the number of all matches that occur at context switches from map to environment
- *#matches env -> map*: the number of all matches that occur at context switches from environment to map
- *#matches total*: the number of all matches that occur at all switches

Matches are computed as follows: for each search phase an according AOI set is derived from the AOI sequence (each AOI occurs at most once in this set). AOI sets of succeeding phases are compared to retrieve the number of matches that occurred at that context switch. For example, if a map phase consists of an AOI set *mapSet = {churchSymbol, operaSymbol, riverSymbol, MAP}*, and the succeeding environment phase consists of an AOI set *envSet = {cityhall, church, opera, NOAOI}*, the number of matches at that context switch would be equal to 2.

## 4 Self-Localization Experiments

### 4.1 Experiment 1: Checking Map Consistency

In this experiment, participants had to judge whether an iconic map is a correct representation of the square they are located on. This task, although not explicitly self-localization, is an important sub-task of self-localization: the matching process between map symbols and environment landmarks.

#### 4.1.1 Participants

To ensure participants were unfamiliar with Zurich, they were recruited through cooperation with a near-by hostel. Participants were provided a small monetary compensation for their efforts. Fourteen participants took part, but four data sets were lost due to errors in the recording software. The 10 remaining participants (six females) had an average age of 27.1 years (min 21, max 50, SD 10.7). Their cultural background and first language were 3x USA/English, 1x British/English, 2x Germany/German, 1x Spain/Spanish, 2x Spain/Catalan, 1xMexico/Spanish. None of the participants wore glasses. None of them was a cartographer, a geographer, or using maps in their profession in any way. Instructions were given in English or German.

#### 4.1.2 Experiment setup and procedure

The experiment took place at Münsterhof, a square in downtown Zurich, Switzerland. Participants were standing on the sidewalk at the corner of Fraumünsterstrasse and Münsterhof (WGS84: 47.369964, 8.540762). Although cars are allowed to park on Münsterhof there is little traffic.

One challenge for real-world wayfinding studies consists of getting the participants to the start location without them being able to build a cognitive map of the environment that might influence the study. To minimize the time they would spend at Münsterhof before the task, we performed the calibration at a nearby location (in front of Helmhaus, a quiet place across the river) and then guided them to the start position.

At the start position participants were given three maps in a row in random order. We refer to them as maps A, B, and C in the following (see Figure 6). Each map shows a circular square with four roads leaving the square at equal distances. Icons are grouped around the square: map A includes icons of types “blue house”, “yellow house”, and “church”. Map B shows icons for “café”, “restaurant”, “shop”, and map C uses symbols from A and B. The map area of each map was sized 28 × 28 cm and printed out on paper (size DIN A3).

The Münsterhof is not circular, the buildings are not distributed at equal distance on each side, and the symbols abstract from the actual objects. We chose these iconic abstractions for the following reasons:

- *No geometric matching*: participants were forced to use landmarks to solve the task, not geometric features of the environment, such as the exact shape of the square.
- *No single-landmark matching*: for each symbol type there was more than one possible landmark in the environment, e.g. more than one blue house. Due to this ambiguity the task could not be solved by matching one symbol/landmark only, but by using several.

On the back of each map the task instruction could be found: “There is a simplified representation of this place on the map. Do you think the objects can actually be found at the places marked on the map? Hint: Not all the objects that can be found in the real environment are represented on the map.” For each of the three maps the researchers noted the participant’s decision (“yes”, “no”). For map A, the correct answer was “yes”, for maps B and C it was “no”. Figure 7 shows the view from the participants' position. Behind the participant there was a wall not relevant to the task (the two accompanying researchers were also there).

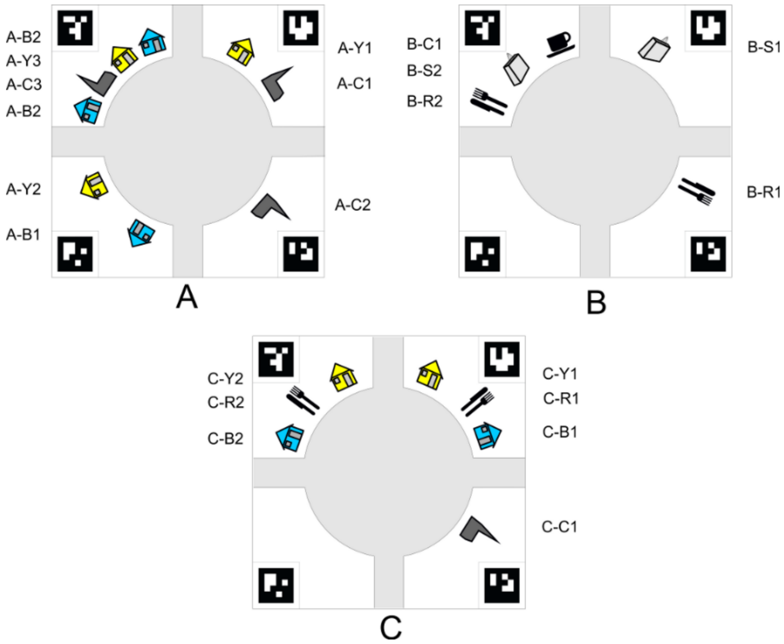



*Figure 6 Iconic maps used for Experiment 1 (symbol identifiers were not on the map)*

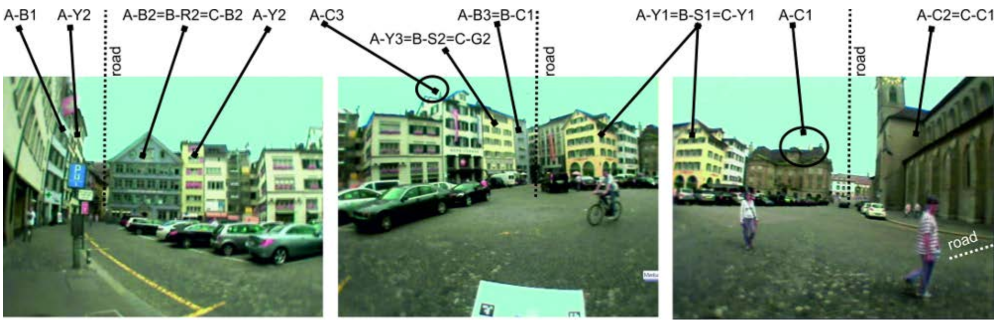



*Figure 7 Experiment 1, view from the participants' position (as seen through the scene camera)*

### 4.1.3 Results

The data analysis started with the pre-processing steps outlined in Section 3: improving pupil detection and re-calibration. In addition to the markers on the map (see Figure 6), there were six similar markers on the back of the map framing the text instructions. Map AOIs in this experiment were drawn around the map area (MAP), the instructions (INS), and each map symbol. The naming scheme for AOIs uses the map id, the type of symbol (Y = yellow, C = church, B = blue, R = restaurant, C = café, S = shop), and a clockwise numbering starting at the top (see Figure 6):

- On map A: A-Y1, A-C1, A-C2, A-B1, A-Y2, A-B2, A-C3, A-Y3, A-B3
- On map B: B-S1, B-R1, B-R2, B-S2, B-C1
- On map C: C-Y1, C-R1, C-B1, C-C1, C-B2, C-R2, C-Y2

Five key markers in the environment were used to define environment AOIs: one around a parking sign, one each around a window of A-Y1, A-Y2, B-S2, and one around a door of A-C2. Each landmark numbered in Figure 7 was surrounded by an AOI (the exact AOI polygons are omitted for reasons of clarity). Thus, all but the following map AOIs had a corresponding AOI in the environment: B-R1, C-R1, C-B1, C-R2. In the following, matches are computed based on the corresponding AOI pairs.

Every participant performed each task once (one measure). We analyzed the results of the experiment by applying the non-parametric Mann-Whitney-U test, in order to investigate if there are significant differences between successful and unsuccessful participants (between-group design).

*Table 2: Results for experiment 1, listing means for successful (s) and unsuccessful (u) trials, based on fixation thresholds of 100 ms (left) and 200 ms (right).Total: 30 trials (20s, 10u).*

| **Fixation Statistics (100 ms - 30 pixel)** | | | | | **Fixation Statistics (200 ms - 30 pixel)** | | | |
|---|---|---|---|---|---|---|---|---|
| | | **avg** | **SD** | | | | **Avg** | **SD** |
| *#fixations* | *s* | *517.85* | *291.2* | | *#fixations* | *s* | *185.5* | *115.4* |
| | *u* | *563.7* | *326.3* | | | *u* | *194.2* | *108.5* |
| *#switches between map and env* | *s* | *52.5* | *29* | | *#switches between map and env* | *s* | *36.7* | *20.8* |
| | *u* | *55.6* | *37.4* | | | *u* | *37* | *18* |
| *# fixations per phase* | *s* | *10.1* | *3.3* | | *# fixations per phase* | *s* | *10.2* | *3.8* |
| | *u* | *10.9* | *3.7* | | | *u* | *8.6* | *2* |
| *% of fixations spent on map* | *s* | *21.9* | *8.9* | | *% of fixations spent on map* | *s* | *21.9* | *11* |
| | *u* | *20.7* | *8.9* | | | *u* | *21.4* | *9.9* |
| *#matches map -> env* | *s* | *6.5* | *5.5* | | *#matches map -> env* | *s* | *3.6* | *3.7* |
| | *u* | *4.6* | *2.5* | | | *u* | *2.4* | *2* |
| *#matches env -> map* | *s* | *4.95* | *4.5* | | *#matches env -> map* | *s* | *2.35* | *2.5* |
| | *u* | *3.3* | *2.3* | | | *u* | *1.5* | *1.5* |
| *#matches total* | *s* | *11.45* | *9.7* | | *#matches total* | *s* | *5.9* | *6* |
| | *u* | *7.9* | *4.2* | | | *u* | *3.9* | *3.4* |
| *#matches map->environment per minute* | *s* | *5.7* | *3.8* | | *#matches map->environment per minute* | *s* | *3* | *2.4* |
| | *u* | *3.9* | *2* | | | *u* | *1.9* | *1.4* |
| *#matches environment->map per minute* | *s* | *4.2* | *3.2* | | *#matches environment->map per minute* | *s* | *2.1* | *2* |
| | *u* | *3.2* | *2.6* | | | *u* | *1.1* | *1* |
| *#matches total per minute* | *s* | *6.8* | *6.3* | | *#matches total per minute* | *s* | *5.1* | *4.1* |
| | *u* | *4.9* | *5.2* | | | *u* | *3.1* | *2.3* |

#### 4.1.3.1 Task duration and performance

In the following we analyze the results with respect to the 30 trials of the experiment. Descriptive statistics are listed in Table 2, based on fixations. Twenty trials were successful, 10 trials were unsuccessful. Trials had an average length of 72 sec (1,800 gazes at 25Hz). Unsuccessful trials took longer (77 sec, SD 43) than successful trials (70 sec, SD 39) which may be a sign for confusion and/or a longer inference process. On average, 16 sec of visual attention were spent on the map (22% of the time). There was no significant difference on the time spent reading the map between successful and unsuccessful trials, either with a fixation duration of 100 ms *($p = 0.895$, $Z = -0.132$: [100 ms threshold])* nor with a fixation duration of 200 ms *($p = 0.912$, $Z = -0.110$: [200 ms threshold]).*

#### 4.1.3.2 Visual attention switches between AOIs

We computed the three attention switch measures introduced: number of matches when switching from map to environment, from environment to map, and number of all matches. All values indicate more matches for successful than for unsuccessful trials. However, the differences are insignificant for none of the two fixation thresholds, either for the matches from the map to the environment *($p$ = 0.218, $Z$ = -1.232: [100 ms threshold])* and *($p$ = 0.333, $Z$ = −0.969: [200 ms threshold])*, the matches from the environment to the map *($p$ = 0.355, $Z$ = −0.924: [100 ms threshold])* and *($p$ = 0.144, $Z$ = −1.461: [200 ms threshold])*, nor for the total number of matches *($p$ = 0.455, $Z$ = −0.748: [100 ms threshold])* and *($p$ = 0.153, $Z$ = −1.430: [200 ms threshold])* (all values are based on fixations and normalized to duration).

### 4.2 Experiment 2: Self-Localization

In this experiment, participants had to mark their position on a tourist map. While map material and task in Experiment 1 were deliberately chosen to be artificial, Experiment 2 is designed to be more realistic.

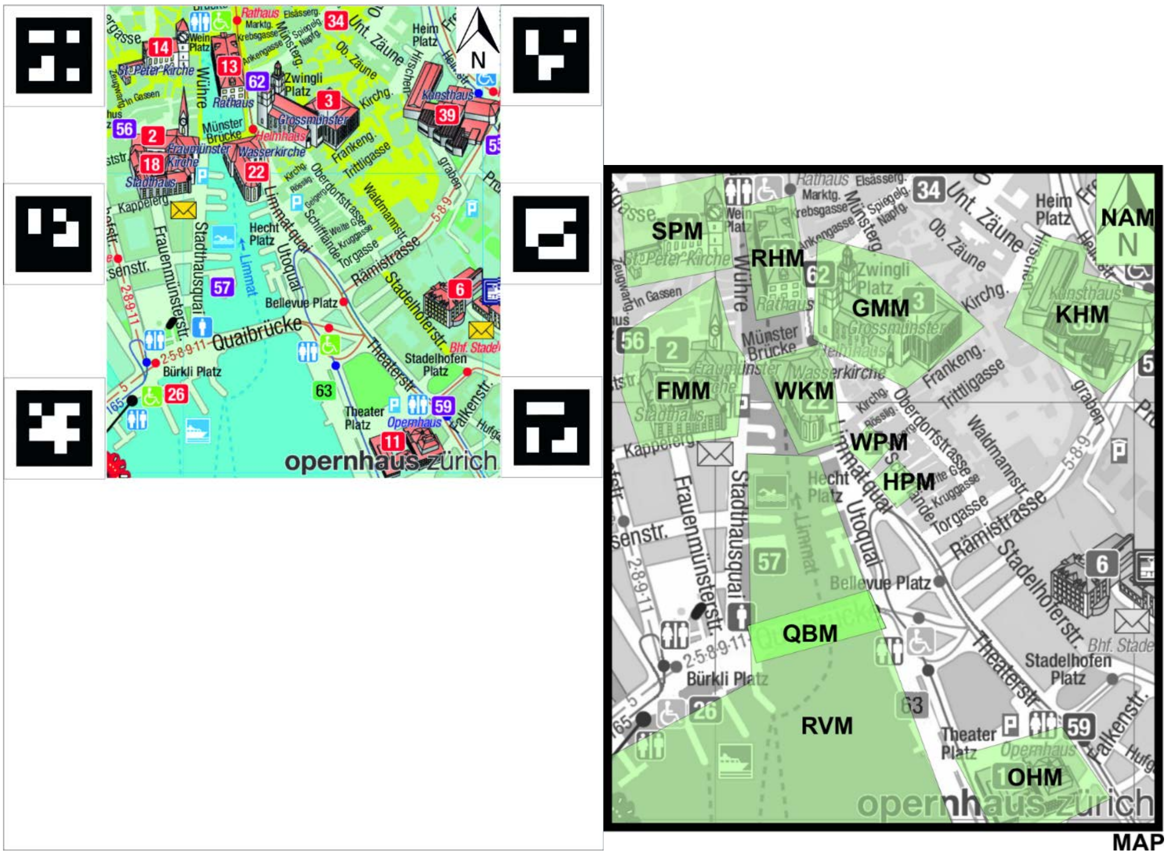



*Figure 8 Map for Experiment 2. Left: Map as seen by participants. Right: AOIs on the map for the analysis (see also Table 3)*


*Figure 9 Experiment 2, view from the participants' position (as seen through the scene camera)*

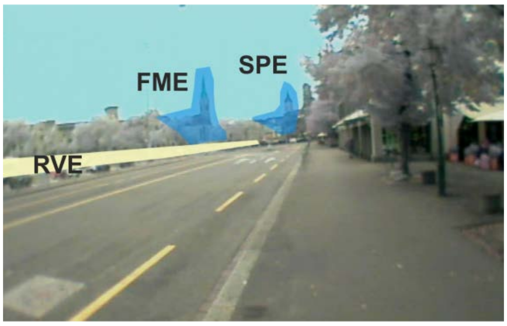
*Table 3: AOIs for the analysis of the Hechtplatz study.*

## AOIs on the map

| | |
|---|---|
| INS | Instructions (back of the map) |
| MAP | Map area |
| NAM | North arrow |
| RVM | River |
| SPM | "St. Peter" church |
| FMM | "Fraumünster" church |
| WKM | "Wasserkirche" church |
| GMM | "Grossmünster" church |
| QBM | "Quaibrücke" bridge |
| KHM | "Kunsthaus" museum |
| OHM | Opera house |
| HPM | "Hechtplatz" (the correct position) |
| WPM | The square North of Hechtplatz |
| RHM | "Rathaus" (city hall) |

## AOIs in the environment

| | |
|---|---|
| RVE | River |
| SPE | "St. Peter" church |
| FME | "Fraumünster" church |




## Others

| | |
|---|---|
| NOAOI | All other |

#### 4.2.1 Participants

As in Experiment 1, we recruited participants through a hostel. Fifteen participants took part, with seven female and eight male persons. The average age was 25 years (min 18, max 39, SD 6.4). Their cultural background and first language were heterogeneous: 1x Australia/English, 1x Germany/German, 1x Serbia/Serbian, 3x Spain/Spanish, 1x The Netherlands/Dutch, 1x South Korea/Korean, 1x Brazil/Portuguese, 1x Caucasian/English, 5x US/English. Again, none of the participants wore glasses and none was using maps in their profession. Instructions were provided in English or German.

#### 4.2.2 Experiment setup and procedure

The experiment took place at Hechtplatz, a small square in downtown Zurich, Switzerland (WGS84: 47.368400, 8.544395). This square was chosen because several prominent landmarks that can help for self-localization are visible from there. The trams and cars passing by on a close road (the "Limmatquai") do not obfuscate these landmarks. The eye tracker was mounted and calibrated in front of the hostel. Participants were then led to Hechtplatz. We chose a route that followed narrow roads, so that participants could not see the river or any other landmark they could potentially need for the task.

At Hechtplatz, participants were given a tourist map of the environment (see Figure 8, left) with the following instruction on the back: "Please try to find and mark your position on the map." The map is a clipping of an official city map provided by Zurich tourism (http://www.zuerich.com/en/travel-trade/Sales-Material/trade-brochures-maps.html). The printout was on DIN A3, the map area itself was sized 19 × 22 cm. The white space below the map area ensured that participants' fingers did not obfuscate the map or the markers, and that the complete map area was visible for the field camera. The landmarks on the map were sufficiently large to be analyzed in the eye tracking video.

Figure 9 shows four views from the participants' position, letters indicating cardinal directions: (SW) The river, the road, and the tram lines could help with orientation. (NW); The two churches visible here are "Fraumünster" and "St. Peter" – both have correspondents on the map; (SE) The direction towards Bellevue had no landmarks with correspondents on the map; and (NE) From this view participants could infer they were standing at the edge of a square (the Hechtplatz). No street signs could be seen without locomotion. Although participants were allowed to change their position, none of them did it extensively. They were mostly moving within a radius of approximately 2 m. While a participant was performing the task, the two accompanying researchers stayed in the background on Hechtplatz. The task was finished when the participant marked the position on the map (with a pen).

#### 4.2.3 Results

Again, the pre-processing steps from Section 3 were performed (pupil detection, re-calibration). As in Experiment 1, map and instructions were framed by markers (see Figure 8). One key marker in the environment was defined between the two churches. The AOIs listed in Table 3 were bound to this marker and used for the analysis. The AOIs on the map are displayed in Figure 8 (right), the AOIs in the environment are visualized in Table 3 (right).

Fixation sequences and fixation AOI sequences were computed as described in Section 3. In the following we analyze the results with respect to the 15 trials of the experiment. Detailed results are listed in Table 4, based on fixations. Seven trials were successful, eight were unsuccessful.

As in Experiment 1, every participant performed each task once (one measure). We analyzed the results of the experiment by applying the non-parametric Mann-Whitney-U test, in order to investigate if there are significant differences between successful and unsuccessful participants (between-group design) during self-localization.

*Table 4: Results for experiment 2, listing means for successful (s) and unsuccessful (u) participants, based on fixation thresholds of 100 ms (left) and 200 ms (right). Total: 15 participants (7s, 8u).*

| **Fixation Statistics (100 ms - 30 pixel)** | | | | | **Fixation Statistics (200 ms - 30 pixel)** | | | |
|---|---|---|---|---|---|---|---|---|
| | | **avg** | **SD** | | | | **avg** | **SD** |
| *#fixations* | *s* | *730* | *295.2* | | *#fixations* | *s* | *315.5* | *133.4* |
| | *u* | *695.7* | *179* | | | *u* | *294* | *84.2* |
| *#switches between map and env* | *s* | *41.1* | *18.5* | | *#switches between map and env* | *s* | *32.2* | *11.8* |
| | *u* | *44.7* | *13.7* | | | *u* | *36.25* | *13.11* |
| *#fixations per phase* | *s* | *19.7* | *8.8* | | *#fixations per phase* | *s* | *10.2* | *3.8* |
| | *u* | *16.1* | *3.3* | | | *u* | *8.6* | *2.2* |
| *% of fixations spent on map* | *s* | *36* | *6* | | *% of fixations spent on map* | *s* | *39.5* | *7.5* |
| | *u* | *36.5* | *11.4* | | | *u* | *40.7* | *13.4* |
| *% of fixations on helpful AOIs* | *s* | *13.1* | *7.2* | | *% of fixations on helpful AOIs* | *s* | *15.2* | *8.5* |
| | *u* | *6.9* | *5.3* | | | *u* | *5.1* | *2.7* |
| *% of fixations on unhelpful AOIs* | *s* | *14* | *7.6* | | *% of fixations on unhelpful AOIs* | *s* | *14.8* | *8.2* |
| | *u* | *14.7* | *9.6* | | | *u* | *18.6* | *8.7* |
| *#matches map -> env* | *s* | *5* | *3.1* | | *#matches map -> env* | *s* | *3.8* | *3.2* |
| | *u* | *1.1* | *1.1* | | | *u* | *0.7* | *1.3* |
| *#matches env -> map* | *s* | *4.5* | *4.1* | | *#matches env -> map* | *s* | *2.8* | *3.3* |
| | *u* | *1.2* | *1* | | | *u* | *0.3* | *0.5* |
| *#matches total* | *s* | *9.5* | *7.2* | | *#matches total* | *s* | *6.7* | *6.5* |
| | *u* | *2.3* | *2.1* | | | *u* | *1.1* | *1.4* |
| *#matches map -> env per minute* | *s* | *3.2* | *2.2* | | *#matches map -> env per minute* | *s* | *2.5* | *2.2* |
| | *u* | *0.7* | *0.6* | | | *u* | *0.4* | *0.6* |
| *#matches env -> map per minute* | *s* | *3* | *2.7* | | *#matches env -> map per minute* | *s* | *1.8* | *2.2* |
| | *u* | *0.8* | *0.6* | | | *u* | *0.29* | *0.4* |
| *#matches total per minute* | *s* | *6.3* | *4.9* | | *#matches total per minute* | *s* | *4.4* | *4.3* |
| | *u* | *1.5* | *1.2* | | | *u* | *0.7* | *1* |


*Figure 10 Participants' solutions to self-localization for the Hechtplatz study (t = true position)*

### 4.2.3.1 Task duration and performance

The positions participants marked on the map are shown in Figure 10. The distribution reveals that all participants were able to locate themselves on the correct side of the river. Participants' solutions were judged as correct or incorrect by five independent researchers. Seven solutions – all located on the correct edge of Hechtplatz – were classified as correct (marked blue). Participants needed between 0:47 and 2:39 min for the task, with an average of 1:31 min (SD 0:27). There was no significant difference concerning the task duration in correlation to task performance *($p = 1$, $Z = 0.0$: [100 ms and 200 ms threshold])*.

### 4.2.3.2 Distribution of visual attention

We approached the research question whether successful participants would spend more visual attention on helpful symbols on the map (RQ1), in our case: on RVM, SPM, FMM. All other map AOIs, for which correspondent landmarks were invisible from the experiment location, constituted the unhelpful AOIs category. The results revealed that the successful participants spent significantly more time fixating AOIs from the helpful AOIs category than the unsuccessful participants *($p < 0.05$, $Z = -2.199$: [100 ms threshold])* and *($p < 0.05$, $Z = -3.125$: [200 ms threshold])*. The difference was also significant after normalizing the AOIs based on their size *($p < 0.05$, $Z = -2.083$: [100 ms threshold])* and *($p < 0.05$, $Z = -2.315$: [200 ms threshold])* (all values are based on and normalized to fixations).

### 4.2.3.3 Visual attention switches between AOIs

Successful participants had significantly more switches of visual attention between symbols on the map and their corresponding landmarks in the environment than did unsuccessful participants.

Successful participants had significantly more matches "map -> env" than the unsuccessful participants *($p < 0.05$, $Z = -2.277$, [100 ms threshold])* and *($p < 0.05$, $Z = -2.003$, [200 ms threshold])*. The successful participants had also significantly more matches "*env -> map*" while using the fixation duration threshold of 200 ms *($p < 0.05$, $Z = -2.092$: [200 ms threshold])*, but not when using 100 ms as duration threshold *($p = 0.104$, $Z = -1.626$: [100 ms threshold])*. The total number of matches was also significantly more for the successful participants *($p < 0.05$, $Z = -2.121$: [100 ms threshold])* and *($p < 0.05$, $Z = -1.974$: [200 ms threshold])* (all values are based on fixations and normalized to duration).

#### 4.2.3.4 Self-estimation of spatial abilities

We correlated the success and fixation statistics with the self-estimation participants had provided by filling in the SBSODS before the experiment (Hegarty et al. 2002). The SBSODS consists of 15 statements about spatial and navigational abilities which are rated on a seven point Likert scale (strongly agree to strongly disagree). The self-estimate provided on an SBSODS was found to correlate highly with the demonstrated real spatial abilities of a person.

We correlated the SBSODS with the analyses of the fixation data. A Spearman Rho test revealed a significant positive correlation with the success of the participants *($p < 0.05$, $r = 0.450$)* and a significant negative correlation with fixations on unhelpful AOIs *($p < 0.05$, $r = -0.455$)*.

## 5 Discussion

The two research questions that guided our experiments were related to participants' visual attention during self-localization. We were interested in whether the effectiveness of the visual information collection process during orientation and self-localization differs for successful and unsuccessful wayfinders, and whether this can be described with quantitative measures from mobile eye tracking data.

*RQ1 Do successful participants spend more visual attention on map symbols that have a visible corresponding landmark than unsuccessful participants?*

This question was approached in Experiment 2: symbols on the map were classified as "helpful" and "unhelpful", with the corresponding landmarks being in sight for the first, and out of sight for the latter. The results demonstrated that the difference between successful and unsuccessful participants in their distribution of visual attention on helpful and unhelpful map symbols was significant. In other words, successful participants focused their attention significantly more on the helpful map symbols for solving the task. These results strongly indicate that successful participants followed a superior self-localization strategy, and that this strategy manifested itself through more effective visual exploration patterns. This is in line with Gunzelmann et al. (2004) who found gaze distributions in an orientation task to depend on the chosen strategy.

However, as self-localization is a complex process that includes several sub-processes, such as map orientation, feature matching, configuration matching, and hypothesis generation and evaluation (Thompson et al. 1990, p. 709f), we cannot tell which of these tasks they performed particularly well. Successful participants may have had superior capabilities in matching map symbols and landmarks, such as comparing the icon of a church and the real church, or superior capabilities in reasoning about configuration. Distinguishing which of these spatial capabilities contributed most to the self-localization would require further investigations, such as the retrospective verbal protocols performed by Spiers and Maguire (2008) or the Perspective Taking/Spatial Orientation Test (Hegarty and Waller 2004).

*RQ2 Do successful participants have more switches of visual attention between symbols on the map and their corresponding landmarks in the environment than unsuccessful participants?*

This question includes two new aspects, compared with RQ1: it also analyzes visual attention in the environment and takes a sequential view. The underlying assumption is that, as working memory is restricted, the fixations on corresponding landmarks and map symbols during a matching process appear in temporal proximity. Through labor-intensive data processing we were able to measure visual attention in the environment (see Section 3.2). The three sequential measures on "matches" (see Section 3.3) were computed for both experiments. For both experiments, these measures, especially when normalized to duration, underline the findings for RQ2. Successful participants were found to have more matches between corresponding landmarks than unsuccessful participants. Experiment 2 yielded insignificant results, while Experiment 1 only indicated a tendency. One possible reason for this could be that the tasks in Experiment 1 required much qualitative spatial reasoning on ordering. Thus, even if a participant had correctly matched the symbols to the landmarks, a small error in reasoning may have led to an incorrect answer.

In summary, the experiments provided strong evidence that successful and unsuccessful landmark-based self-localization can be differentiated by their typical gaze patterns with respect to fixation distribution and sequences of visual attention.

Besides the insights provided on RQ1 and RQ2, another contribution of this article consists in a novel methodological approach to wayfinding studies using mobile eye tracking in real-world experiments. In previous studies, eye tracking research on wayfinding focused on laboratory studies (Gunzelmann et al. 2004, 2008; Hayhoe and Ballard 2005; Peebles et al. 2007; Spiers and Maguire 2008; Wiener et al. 2012), mostly because of technological limitations and controllable influences. First tries of using eye tracking for outdoor wayfinding studies report on technical problems and either abandoned the method (Delikostidis 2011), or chose a manual coding approach (Pinelo Silva 2011). In our experiments we found the quality of the captured data to be very high, and through the ability to manually verify and correct the data (manual pupil detection, correction of parallax, recalibration, etc., see Section 3) we can work with high accuracy. Manual verification of the captured data is a very time consuming process, but unavoidable for research where high accuracy is of major importance. The experiments reported in this article were performed in a real urban environment for several reasons. Side effects can occur during experiments performed in virtual environments that might cause a large distortion in the collected data. According to Barrett (2004), the so called "*cybersickness*" symptoms can occur during such experiments and can be grouped into nausea, disorientation and visual symptoms. Moreover, the necessary ability to explore the environment around us through coordinated eye and head movements has not yet been transferred appropriately to virtual reality environments (Bolte et al. 2010).

In contrast to laboratory studies, the environmental and human factors that can influence an outdoor experiment are numerous. In the presented experiments we tried to minimize these factors, but still there will always be noise in data collected in real environments. Experiment times were chosen so that environmental influences, such as weather and lighting conditions, were similar for each participant. Our experiences with the pilot study (Kiefer et al. 2012a) helped in selecting an optimal location for the experiments, with clearly visible landmarks and minimal traffic impact. More problematic were unpredictable influences from pedestrians. For instance, we had to make sure that curious pedestrians would not start talking to the participant. In addition, the scheduling and organization of outdoor experiments is much more unpredictable and time-consuming than for laboratory studies. For instance, two weeks of rainy weather during the summer may cause a considerable delay in the experiment schedule.

The map design might have had an influence on our study. A different set of symbols for Experiment 1, or a different symbolization of the landmarks in Experiment 2 might have led to different results. However, visual design factors for maps are beyond the scope of this article. Map alignment is also an important factor, as it can influence the preferred strategy used for an orientation task (Gunzelmann et al. 2004), and has an influence on efficiency (Warren et al. 1990). As we were not specifically interested in mental rotation (Hegarty and Waller 2004), but in the visual information collection process, our participants were free to rotate their map, thus being able to replace mental by physical rotation. Location-aware technologies facilitate self-localization by identifying and visualizing their user's position through built-in GPS (Global Positioning System). The resulting maps can be seen as mobile correspondents of wall-mounted "you-are-here" maps (Klippel et al. 2006). One could argue that the practical impact of self-localization experiments in the age of mobile "you-are-here" maps might be limited (Meilinger et al. 2007, p. 398). Even the orientation task could be facilitated by the built-in compass. However, beside the theoretical impact described above, a practical application of our results can be seen in new interaction methods based on gaze (Giannopoulos et al. 2013), refer to Section 6.

Spatial orientation and self-localization are typically studied under the conditions of a "drop-off task" (Thompson et al. 1990), in which "a person is either viewing or is actually immersed within a scene, and has to look for the first time at a map of the area to try to match it" (Peebles et al. 2007, p. 390). This ensures controllability with respect to the subject's pre-knowledge about the spatial environment. For outdoor real-world experiments, such as ours, this is harder to achieve than for laboratory studies in which arbitrary spatial scenes can be displayed on the screen. We approached this issue by selecting as participants only newly arrived tourists who had never been to the study region. We considered blindfolding while approaching the starting position (Iachini and Logie 2003), but refused the idea due to potential eye tracker decalibration. Recruiting participants unfamiliar with the location was a difficult endeavor, as some of the people we approached at the hostel were already familiar with the city of Zurich. This does not pose a problem for laboratory studies where an arbitrary spatial surrounding can be displayed. As for most eye tracking experiments, the impossibility of wearing glasses and eye tracker at the same time restricted the choice of participants. Another potential, but probably small, bias in our experiments could be caused by the fact that the majority of the participants were tourists staying at a hostel. Although time consuming and less controllable, we believe that mobile eye tracking experiments in the real world have higher ecological validity than laboratory studies, and thus have the potential to contribute important new insights for studying human wayfinding behavior.

## 6 Conclusions and Outlook

In this article we reported on new insights into the process of self-localization based on landmarks by utilizing a novel methodology using mobile eye tracking in the real world. We described two experiments, one focusing on the matching between landmarks and map symbols on an iconic map, the other investigating the self-localization process as a whole using a tourist map. As the most striking result from these studies we could observe differences in the gaze behavior between successful and unsuccessful participants. These differences were measured with statistics on the distribution of visual attention on different map symbols, and with statistics on the switches of visual attention between landmarks in the environment and their corresponding map symbols.

The focus of this research was on self-localization, because our goal is to analyze the different phases of wayfinding separately. Thus, as a next step we will investigate visual attention during the other phases that were not treated here, such as route planning and plan execution. This article treated self-localization mainly as a problem of matching landmarks with map symbols. However, other features of the environment, such as street geometry, also influence the self-localization process (Meilinger et al. 2007). This could be another issue for future research in real-world studies in order to acquire a more holistic view of self-localization.

Another direction for future research is the replication of previous wayfinding studies in the real world, which were originally performed in the laboratory. For instance, it would be interesting to investigate whether the gaze bias towards the eventually chosen direction reported by Wiener et al. (2012) would also appear in a real-world study. In general, we believe that methodological work on the comparison of laboratory and real world studies would be worthwhile to investigate the balance between ecological validity and controllability.

One practical application of the findings from this experiment could be gaze-based assistive systems, such as outlined in Giannopoulos et al. (2013). A gaze-based assistive system could detect that a user's visual information collection strategy during orientation or self-localization is ineffective, e.g. because there are only few corresponding landmarks in subsequent search phases, and then provide information on the correct matching. Another issue for future research in this context is gaze-based activity recognition (Bulling et al. 2011): a gaze-based assistant could detect from characteristic gaze patterns that the user is trying to orient him- or herself, concluding that he or she is disoriented and needs help.

**Acknowledgements**

We would like to thank Florian Straub for his contributions to an earlier version of this work. Mikko Schmitter, Ruth Kläy, and Katharina Kneer helped tremendously with the data pre-processing. Thank you also to Jördis Graf and Cristina Iosifescu for co-supervising our experiment sessions. The City Backpacker Hotel Biber, Zurich kindly supported our recruitment of participants. We also thank the participants for taking part in our experiments. The feedback of three anonymous reviewers provided important and helpful suggestions to improve the content of this article.

**References**

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the Psychological Laboratory: Truth or Triviality? Current Directions in Psychological Science, 8(1), 3-9. doi: 10.2307/20182544

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. Human-Computer Interaction, 12(4), 439-462.

Andrienko, G., Andrienko, N., Burch, M., & Weiskopf, D. (2012). Visual Analytics Methodology for Eye Movement Studies IEEE Transactions on Visualization and Computer Graphics, 18(12), 2889-2898.

Barrett, J. (2004). Side effects of virtual environments : a review of the literature / Judy Barrett. Edinburgh, S. Aust: DSTO Information Sciences Laboratory.

Bolte, B., Bruder, G., Steinicke, F., Hinrichs, K., & Lappe, M. (2010). Augmentation techniques for efficient exploration in head-mounted display environments. Paper presented at the Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology, Hong Kong. pp. 11-18

Bulling, A., Ward, J. A., Gellersen, H., & Tröster, G. (2011). Eye Movement Analysis for Activity Recognition Using Electrooculography. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(4), 741--753.

Castner, H. W., & Eastman, J. R. (1984). Eye-Movement Parameters and Perceived Map Complexity - I. Cartography and Geographic Information Science, 11(2), 107-117.

Çöltekin, A., Fabrikant, S. I., & Lacayo, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. International Journal of Geographical Information Systems, 24(10), 1559-1575.

Çöltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-Movement Analysis. Cartography and Geographic Information Science, 36, 5-17. doi: 10.1559/152304009787340197

Davies, C., & Peebles, D. (2010). Spaces or scenes: Map-based orientation in urban environments. Spatial Cognition & Computation, 10(2-3), 135-156.

Delikostidis, I. (2011). Improving the usability of pedestrian navigation systems. PhD Thesis, University of Twente, Enschede, The Netherlands, 263 pages

Denis, M., Michon, P.-E., & Tom, A. (2007). Assisting pedestrian wayfinding in urban settings: Why references to landmarks are crucial in direction-giving. In G. Allen (Ed.), Applied Spatial Cognition - From Research to Cognitive Technology (pp. 25-51). Mahwah, NJ: Lawrence Erlbaum.

Dobson, M. W. (1977). Eye Movement Parameters and Map Reading. Cartography and Geographic Information Science, 4(1), 39-58.

Duchowski, A. T. (2007). Eye Tracking Methodology: Theory and Practice (2nd ed.). London: Springer.

Dudchenko, P. (2010). Why People Get Lost - The psychology and neuroscience of spatial cognition. Oxford, New York: Oxford University Press.

Emo, B. (2012). Wayfinding in real cities: experiments at street corners. In C. Stachniss, K. Schill & D. Uttal (Eds.), Spatial Cognition VIII (pp. 461--491). Berlin Heidelberg: Springer.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. Vision Research, 51(17), 1920-1931. doi: http://dx.doi.org/10.1016/j.visres.2011.07.002

Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. Vision Research, 50(24), 2766-2774. doi: http://dx.doi.org/10.1016/j.visres.2010.09.024

Giannopoulos, I., Kiefer, P., & Raubal, M. (2012). GeoGazemarks: Providing gaze history for the orientation on small display maps Proceedings of the 14th International Conference on Multimodal Interaction (ICMI '12) (pp. 165-172). New York, NY, USA: ACM.

Giannopoulos, I., Kiefer, P., & Raubal, M. (2013). Mobile Outdoor Gaze-Based GeoHCI. Paper presented at the Geographic Human-Computer Interaction, Workshop at CHI 2013, Paris, France, pp. 12-13.

Golledge, R. (1999). Human Wayfinding and Cognitive Maps. In R. Golledge (Ed.), Wayfinding Behavior - Cognitive Mapping and Other Spatial Processes (pp. 5-45). Baltimore: Johns Hopkins University Press.

Gunzelmann, G., Anderson, J. R., & Douglass, S. (2004). Orientation tasks with multiple views of space: Strategies and performance. Spatial Cognition and Computation, 4(3), 207-253.

Gunzelmann, G., Douglass, S., & Khooshabeh, P. (2008). Learning to Orient Using a Map Display: Evidence from Eye Tracking In C. Hölscher (Ed.), Spatial Cognition 2008: Poster Presentations (SFB/TR 8 Report No. 016 - 08/2008) (pp. 85-88). Freiburg, Germany: Universität Bremen / Universität Freiburg.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. Trends in Cognitive Sciences, 9(4), 188--194. doi: 10.1016/j.tics.2005.02.009

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. Intelligence, 30(5), 425-447. doi: 10.1016/S0160-2896(02)00116-2

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. Intelligence, 32(2), 175-191.

Hirtle, S., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. Memory & Cognition, 13(3), 208-217.

Hirtle, S., & Raubal, M. (2013). Many to Many Mobile Maps. In M. Raubal, D. Mark & A. Frank (Eds.), Cognitive and Linguistic Aspects of Geographic Space - New Perspectives on Geographic Information Research (pp. 141-157). Berlin, Heidelberg: Springer.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van De Weijer, J. (2011). Eye Tracking - A Comprehensive Guide To Methods And Measures. New York: Oxford University Press.

Iachini, T., & Logie, R. H. (2003). The role of perspective in locating position in a real-world, unfamiliar environment. Applied Cognitive Psychology, 17(6), 715-732. doi: 10.1002/acp.904

Kato, Y., & Takeuchi, Y. (2003). Individual differences in wayfinding strategies. Journal of environmental psychology, 23(2), 171-188.

Kiefer, P., & Giannopoulos, I. (2012). Gaze map matching: mapping eye tracking data to geographic vector features Proceedings of the 20th International Conference on Advances in Geographic Information Systems (pp. 359-368). New York, NY, USA: ACM.

Kiefer, P., Straub, F., & Raubal, M. (2012a). Location-aware mobile eye tracking for the explanation of wayfinding behavior. Paper presented at the AGILE'2012 International Conference on Geographic Information Science, Avignon, France.

Kiefer, P., Straub, F., & Raubal, M. (2012b). Towards Location-Aware Mobile Eye Tracking Proceedings of the 2012 Symposium on Eye-Tracking Research and Applications (ETRA 2012) (pp. 313-316). New York, NY, USA: ACM.

Klippel, A., Freksa, C., & Winter, S. (2006). You-are-here maps in emergencies - the danger of getting lost. Journal of spatial science, 51(1), 117-131.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. Nature, 369(6483), 742-744.

Li, X., Çöltekin, A., & Kraak, M. J. (2010). Visual Exploration of Eye Movement Data Using the Space-Time-Cube. In S. I. Fabrikant, T. Reichenbacher, M. van Kreveld & C. Schlieder (Eds.), Geographic Information Science (Vol. 6292, pp. 295-309): Springer.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. Trends in Cognitive Sciences, 4(1), 6-14.

Meilinger, T., Hölscher, C., Büchner, S., & Brösamle, M. (2007). How Much Information Do You Need? Schematic Maps in Wayfinding and Self Localisation. In T. Barkowsky, M. Knauff, G. Ligozat & D. Montello (Eds.), Spatial Cognition V Reasoning, Action, Interaction (Vol. 4387, pp. 381--400): Springer Berlin / Heidelberg.

Montello, D., & Raubal, M. (2012). Functions and applications of spatial cognition. In D. Waller & L. Nadel (Eds.), Handbook of spatial cognition (pp. 249-264). Washington, DC: American Psychological Association.

Nevelsteen, K., & Steenberghen, T. (2012). Eye Movement Analysis of Elementary School Children in Traffic Environments. In J. Gensel, D. Josselin & D. Vandenbroucke (Eds.), Proceedings of the AGILE'2012 International Conference on Geographic Information Science. Avignon, France.

Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., & Witlox, F. (2012a). Interpreting maps through the eyes of expert and novice users. International Journal of Geographical Information Science, 26(10), 1773-1788. doi: 10.1080/13658816.2011.642801

Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., & Witlox, F. (2012b). Investigating the Effectiveness of an Efficient Label Placement Method Using Eye Movement Data The Cartographic Journal, 49(3), 234-246.

Opach, T., & Nossum, A. (2011). Evaluating the Usabilty of Cartographic Animations with Eye-Movement Analysis. Paper presented at the 25th International Cartographic Conference, Paris.

Peebles, D., Davies, C., & Mora, R. (2007). Effects of geometry, landmarks and orientation strategies in the ‘drop-off’orientation task Spatial Information Theory (pp. 390-405): Springer.

Pinelo Silva, J. D. (2011). Spatial congruence theory: visual cognition and wayfinding in the urban environment. PhD Thesis, UCL (University College London), London, 240 pages

Raubal, M., & Winter, S. (2002). Enriching Wayfinding Instructions with Local Landmarks. In M. Egenhofer & D. Mark (Eds.), Geographic Information Science - Second International Conference, GIScience 2002, Boulder, CO, USA, September 2002 (Vol. 2478, pp. 243-259). Berlin: Springer.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3), 372-422.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. Paper presented at the Proceedings of the 2000 symposium on Eye tracking research & applications, Palm Beach Gardens, Florida, USA. pp. 71-78

Schuchard, R. A., Connell, B. R., & Griffiths, P. (2006). An environmental investigation of wayfinding in a nursing home. Paper presented at the 2006 Symposium on Eye Tracking Research & Applications (ETRA 2006), San Diego, California. pp. 33-33

Spiers, H. J., & Maguire, E. A. (2008). The dynamic nature of cognition during wayfinding. Journal of environmental psychology 28(3), 232-249. doi: 10.1016/j.jenvp.2008.02.006

Steinke, T. R. (1987). Eye movement studies in cartography and related fields. Cartographica: The International Journal for Geographic Information and Geovisualization, 24(2), 40-73. doi: 10.3138/J166-635U-7R56-X2L1

Stellmach, S., & Dachselt, R. (2012). Investigating gaze-supported multimodal pan and zoom. Paper presented at the 2012 Symposium on Eye Tracking Research & Applications (ETRA 2012), Santa Barbara, CA, USA. pp. 29-36

Thompson, W., Pick, H., Bennett, B., Heinrichs, M., Savitt, S., & Smith, K. (1990). Map-based localization: The drop-off problem Proceedings of the DARPA Image Understanding Workshop (pp. 706-719). San Mateo, CA, USA: Morgan Kaufmann.

Vansteenkiste, P., Cardon, G., D'Hondt, E., Philippaerts, R., & Lenoir, M. (2013). The visual control of bicycle steering: The effects of speed and path width. Accident Analysis & Prevention, 51(0), 222-227. doi: http://dx.doi.org/10.1016/j.aap.2012.11.025

Warren, D. H., Rossano, M. J., & Wear, T. D. (1990). Perception of map-environment correspondence: The roles of features and alignment. Ecological Psychology, 2(2), 131-150.

Widdel, H. (1984). Operational Problems in Analysing Eye Movements. In G. G. Alastair & J. Frank (Eds.), Advances in Psychology (Vol. Volume 22, pp. 21-29): North-Holland.

Wiener, J. M., Hölscher, C., Büchner, S., & Konieczny, L. (2011). Gaze behaviour during space perception and spatial decision making. Psychological Research, 1-17. doi: 10.1007/s00426-011-0397-5