University of Zurich, Department of Geography

Geographic Information Visualisation and Analysis

A Visual Search Efficiency Study

An Evaluation of Labels, Road Junctions and Landmarks in 2D Orthogonal Maps

Master Thesis GEO 511

31.01.2014

Author: Floris Heim

Matriculation number: **07-714-272**

Advisor: Dr. Arzu Çöltekin

Faculty member: Prof. Dr. Sara I. Fabrikant

“[…] the final test of map reading is the visualization of landscape from map." (Sylvester, 1952:52)

Abstract

The goal of this master thesis is to increase the understanding of the perception of real-world environmental features frequently used as orientation guidelines in wayfinding and navigation in 2D layout map representations. An environment can be experienced in a great variety of different ways and more important from different points of view. People can be a part of the environment while moving around in it, or they can inspect the environment from above. But they can also experience the environment by studying a map, or listening to a verbal description. It is assumed that while we experience an environment or a physical map, a cognitive map is developed with similar features as the actual environment. The interaction between the physical map representation, the environment and the cognitive map is obviously open for a number of ways which can go awry. It is therefore important to understand this transaction in general to improve the presentation of cartographic map information.

In this master thesis a visual search efficiency study is conducted evaluating the perception of street labels, road junctions and landmarks in 2D layout map representations. Two main experiments were conducted, first a simple search display experiment in which subjects had to search for a predefined target item in a cartographic or satellite map representation. Secondly, a visual search experiment was done in which subjects had to determine the location of a camera according to a scene photograph in either a cartographic or a satellite map representation. In the 2D layout map representation landmark information had the fastest response time (RT) whereas road junctions were found least efficient. For the satellite 2D layout map representation road junctions were found most efficient and street label information was found least efficiently. The RT results for the "picture to 2D layout map" search task showed that for the satellite map representation road junctions are the most efficient environmental features for self-localisation, whereas for the cartographic map representation street labels are the most efficient environmental feature for self-localisation. However only between the street label condition and the road junction and landmark condition in the cartographic map representation a significant difference in RT was measured. Considering the error rate, landmark cues are the most accurate environmental features for self-localisation in a satellite map representation. In contrast to the satellite map representation, in the cartographic map representation street labels are the most accurate environmental features. This means that landmarks work well as a help in a self-localisation task but the best performance is still achieved when using a cartographic map representation looking for street labels.

Acknowledgements

I would like to thank my advisor Dr. Arzu Çöltekin who led me have all the scientific freedom I only could have dreamed of while still giving me the guidance I needed. Thanks for keeping an open mind!

I also would like to thank Kenan Bektas for providing the saliency maps used in this master thesis.

Special thanks to all my participants, you did a great job! Without you there would not be much to talk about in this master thesis!

Special, special thanks to my beautiful girlfriend, my friends and my family. Thanks for being there for me!

Table of Contents

List of Figures

List of Tables

1. Introduction

The multidisciplinary character of geovisualisation research is a great challenge and a great chance as well. It provides the possibility to connect a vast variety of different research fields to create new theoretical frameworks as well as real world applications. In geovisualisation especially the interactions between map makers, maps, and map readers, graphically represented in figure 1 by Lloyd (2005), demand for multidisciplinary theories and research methods leading in the best case to a broadened knowledge in all participating research fields.

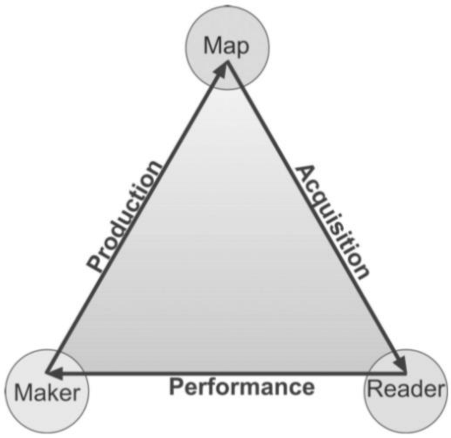


Figure 1: Trilateral relationship involving makers of maps, maps and readers of maps (Lloyd, 2005:29).

The goal of the map maker should be to create a map which allows the map reader to acquire information from the map quickly, accurately, and confidently (Lloyd & Bunch, 2003). To do so it is crucial to understand the process of information acquisition by the map reader. It is therefore only logical to connect the cartographic design process to the large research body in visual attention and visual search in cognitive psychology and neurosciences (e.g. Eckstein, 2011; Treisman & Gelade, 1980; Wolfe & Horowitz, 2004). This master thesis intends to strengthen the bond between geovisualisation research and cognitive psychology and neurosciences research by analysing the human object recognition process in 2D cartographic as well as in satellite maps in orientation and localisation tasks inspirited by research done in navigation and wayfinding studies (e.g. Allen, 1997; Caduff & Timpf, 2008; Frankenstein, Büchner, Tenbrink, & Hölscher, 2010; Pazzaglia & Meneghetti, 2010).

In this chapter the research questions and thereby the goals of this master thesis are defined. Also the main methods used to achieve these goals are described. But first the motivation which drives this master thesis is specified, and reasons why there is a need for an in depth understanding of object recognition in 2D map representations are presented.

1.1 Motivation

Possibly everybody knows the situation being on a tight schedule and looking for a meeting point in an unfamiliar environment. Even if we had memorized the route from the train station to the meeting point before, sometimes we suddenly have the feeling that something went wrong. In a situation like this most people will pull out a road map, or nowadays more likely their smart phone checking their mobile map app. What happens next in such a situation is called self-localisation. Self-localisation commonly refers to the process of identifying one's current position on a map by matching visually perceptible features of the environment, such as buildings, streets, parks or street labels to the content of the map representation (Kiefer, Giannopoulos, & Raubal, 2013). But how do we identify the real world objects we are facing during this process of self-localisation on an orthogonal map? In some extend this question corresponds to a mentioned research challenge formulized by MacEachren & Kraak (2001) stating that "a fundamental problem for geovisualization is to understand (and take advantage of) the mechanism by which the dynamic, external visual representations offered by geovisualization serve as prompts for the creation and use of mental representations (MacEachren & Kraak, 2001:8)". This research challenge formulized by MacEachren & Kraak (2001) approaches the question from the common research point of view in cartography: from the map (abstraction) to the map reader, equally as presented in figure 1 before. This point of view is in line with the early research in cartography which was mainly focused on map design decisions like what kind of symbols and structures to use in a certain map representation. In the centre of the cartographic research was clearly, and still is, the map and how to improve the transmission of intended messages and information rather than how and why they work for the map reader. This circumstance is also pointed out by critics of cartographic research because still little theoretical understanding of the interaction taking place between the map and the map reader is provided (Bunch & Lloyd, 2006). Nevertheless a growing community of geovisualisation researchers further the understanding of why maps work as they work. The majority of this research body however focuses on how to best communicate information in thematic or choropleth maps (e.g. Çöltekin, Heil, Garlandini, & Fabrikant, 2009; Garlandini & Fabrikant, 2009; R. Lloyd & Bunch, 2003; Nelson, 1994) under aspect of visual search and visual attention theories, rather than how topographic maps or road maps are helping map readers to orientate themselves in an unfamiliar environment. An overview of strategies and cognitive processes associated with navigational map reading is provided by Lobben (2004). Lobben (2004) points out that although the scope of research has broadened, a majority of spatial testing is still done by psychologists trying to understanding spatial cognition as part of the overall cognitive process. This means that the focus in these studies of map reading cognitive research continues to mainly focus on cognitive processes and not on the map itself. So while geographers primarily focus on the perception of geovisualisation, psychologists focus on the cognitive processes involved during spatial tasks without much consideration of the map representation used (Lobben, 2004). One concern of this master thesis will therefore be the integration of spatial cognitive research findings in psychological experiments in geovisualisation research and vice versa. The question of how self-localisation works will therefore be connected to the classic psychology centred research in visual attention as well as to findings in navigational map research and wayfinding research. This master thesis will hopefully also help to close the still existing research gap in the understanding of how 3D visual scenes and their 2D layout geometry in map representations help us navigate through the world.

1.2 Research Questions

A fine way to start looking for a research questions is reading a paper by Carter (2005) providing a definition of 10 generic map uses. In this master thesis I will focus on the second generic task defined by Carter (2005:3):

> "**Navigation, control and route planning** – we use maps to navigate from A to B. We also use them to plan routes and we can control whether we are still on the planned route while on the road."

The underlined part in the quotation above further describes the research direction this master thesis will take, that is to say: How can we locate ourselves on a map when we got lost in an unfamiliar environment? This leads us to the first research question:

> **RQ1: What environmental features are preferably used during self-localisation?**

The first research Question (RQ1) will be answered mainly by a literature review of findings in navigational and wayfinding research. The most important environmental features defined in RQ1 will then be tested on their efficiency in a self-localisation task. The second research question (RQ2) is therefore:

> **RQ2: How efficient are the in RQ1 defined environmental features for self-localisation on a map?**

RQ2 will test the efficiency of environmental features frequently used for example in way description tasks for a self-localisation task on a map. In this context it is important to define the map type and the map scale used, because as it is well known in geovisualisation research "[…] each map within a certain scale range requires its own level of detail depending on its purpose." (Kraak & Ormeling, 2003:75). To answer RQ2 it is therefore important to focus on environmental and artificial features present in a great variety of different maps and in a certain scale range appropriate for local navigational tasks. Because in RQ2 only environmental features present in the defined map type and map scale can be used, it is most likely that not all environmental features defined in RQ1 can be tested for their efficiency in a self-localisation task. In this context it is also of interest if frequently used environmental features (RQ1), and/or efficient environmental features (RQ2) used for self-localisation on a map, can be easily found as their corresponding map features on a map representation itself. RQ1 and RQ2 are therefore directly linked to the third research question (RQ3):

> **RQ3: How efficient can the corresponding 2D layout features of the environmental features defined in RQ1 be found on a map?**

For each defined environmental feature in RQ2 and RQ1, there will be an abstract or generalized corresponding 2D layout feature in the map representation. RQ3 focuses on how efficient these corresponding 2D layout features can be found on a map representation. The fourth research question (RQ4) combines RQ1, RQ2 and RQ3 and is therefore formulized as following:

> **RQ4: Does the reaction time (RT) needed in RQ3 to find the 2D layout of the environmental features defined in RQ1 on a (road) map correspond to their efficiency as environmental features in a self-localisation task (RQ2)?**

RQ4 refers to the question if, for example, the search time needed to search for a building in a map representation (RQ3) corresponds to the efficiency of buildings as environmental features in a self-localisation task (RQ2). Perhaps it could be that buildings can be found very quickly on an orthogonal map representation but that they are not very helpful as environmental features when standing in front of them in the real world. Research question five on the other hand contributes partly to the lively empirical debate about the influence of abstraction and generalization on map performance.

RQ5: Are certain environmental features more efficient for abstract cartographic map representations than for satellite map representations and vice versa?

The operationalization of the above defined research questions will be further addressed in the method section.

1.3 Methodology Overview and Expected Results

In this section a very short overview of the methodology and expected results will be given. The main part of the methodology will be a user study with a within subject experimental design. In the user study the research questions defined previous will be addressed by conducting different visual search tasks. During the visual search tasks the eye movements of the subjects will be recorded. The recorded eye movements will mainly be used to determine the reaction time (RT) while subjects look for different stimuli. Additionally to the visual search task, subjects will have to fill out an online survey, an spatial orientation test by Kozhevnikov & Hegarty (2001), and the Santa Barbara Sense-of-Direction Scale by Mary Hegarty, Richardson, Montello, Lovelace, & Subbiah (2002). A graphically overview of the methodology used in this master thesis is given in figure 2.

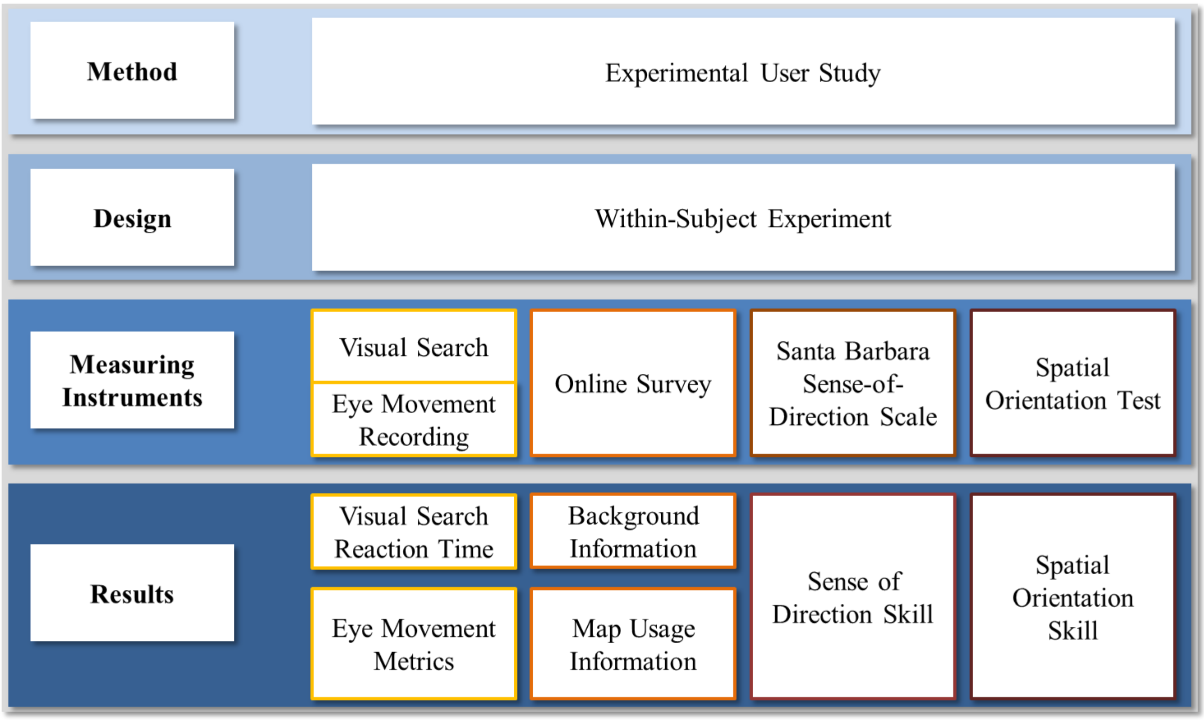


Figure 2: Research methodology overview.

The implementation of the methodology presented in figure 2 will be described in the chapter methods later in this master thesis.

1.4 Outline

This master thesis has the following structure: introduction, state of the art, methods, results, discussion, conclusion and an outlook of possible further research. In the chapter state of the art a short overview of related research fields of this master thesis is provided. The focus in the state of the art chapter lies on the main sections of navigational research and vision research. The chapter methods provides the background information about the experimental stimuli operationalization and the user study design. Afterwards the results of the user study as well as the results of the online survey, the Santa Barbara Sense-of-Direction and the Spatial Orientation test are presented in the chapter results. In the chapter discussion the results are discussed and integrated in a broader empirical research framework. The chapter conclusion points out the quintessence of the findings of this master thesis, and the master thesis is completed with an outlook for further research.

2. State of the Art

In this chapter an overview of the relevant research in map perception, navigation and wayfinding as well as in vision will be provided. Research findings in these areas of research are provided because the experiment conducted in the context of this master thesis seeks to connect theses research fields to generate new knowledge. One main element in self-localisation is the map, providing a certain amount of information about the environment around us. It is therefore important to start the state of the art chapter with an overview of map perception and map reading. In line with the research questions defined in the introduction the state of the art in map reading and perception will mainly be focused on navigation and self-localisation. The state of the art chapter is divided in two main sections. The first section covers the basic theories in navigational research, whereas the second section summarizes the main findings in vision research with emphasis on visual attention and visual search.

2.1 Navigational Research

People can experience an environment in a great variety of different ways and more important from different points of view. People can be a part of the environment while moving around in it, or they can inspect the environment from above, for example from an airplane, a mountain top or a high building. But they can also experience the environment by studying a map, or listening to a verbal description (Pazzaglia & Meneghetti, 2010). Navigation is usually defined as a coordinated and goal-directed movement though the environment, requiring the planning of a route as well as the movement to execute the planned route (Montello, 2005). Wayfinding on the other hand refers to the planning process involving typically some sort of route instruction (e.g. verbal, written, or drawn) (Caduff & Timpf, 2008). The first subsection starts with a short summary of different map uses in general, continues with a section focusing on navigation and finalises with a short summary of map perception.

2.1.1 Map Use and Perception

Board (1978) points out that, it is fundamentally important that the evaluation of maps must be based on map reading tasks that are appropriate to the map reading objectives. To reach this goal he provides a set of evaluation guide lines which should help to determine hypotheses which are based upon the way in which map readers normally use maps. In a first step Board (1978) summarizes the three basic types of purpose for using geographical information (Board, 1978:3):

- Facilitate movement from one place to another (*navigation*)
- Acquiring information on the geographical environment (*measurement* and *visualisation*)

For this master thesis especially navigation and visualisation is of interest, because both processes involve the matching of the map to the reality. In a next step Board (1978) further divides the first basic type of purpose for using geographic information, the navigational map use, into the following three relevant components for which orientation is essential (Board, 1978:3):

- Selection of the route
- Maintenance of the course
- Discovery of the objective

Out of these three relevant components of the navigational map use only the selection of the route does not necessarily involve self-localisation or any real time interactions between map representation and environment. For the maintenance of the course Board (1978) suggests that map users check or search for landmarks on the map, known to be on, or near the route selected. The third basic type of geographical information used in maps, the visualisation, is described by Board (1978) as a process in which "[…] relatively large parts, commonly the whole of the map face are scanned for clues in search for order, regularity or pattern of some significance" (Board, 1978:7). Scanning of this sort is called browsing, when the only purpose of the scanning process is the gathering of random spatial information. This information can then be used to identify nearby environmental features (Board, 1978).

According to Blades & Spencer (1987) two issues can evolve from the use of a map. The first issue can evolve because of the cognitive abilities of the map user in relating information from the map to the environment in order to travel through that environment successfully. The second issue can evolve because of the practical consequences of individual competence, or lack of competence with maps (Blades & Spencer, 1987). For both identified possible issues in navigational map use there is surprisingly little literature available concerning navigation in real-world settings. In this context also the questions remain how maps are used in combination with landmarks, and how real-world way-finding tasks differ from navigational A to B task study experiments (Brown, 2007). One possible way to understand the interaction between map user and environment during navigation or self-localisation would be to focus on how perceptual information is understood conceptually. It is believed that while learning a physical map, a cognitive map is developed with similar features as the actual environment. Nevertheless the cognitive map will differ from the real world to some extent because of the cognitive process which forms it. The cognitive process involved in building a mental or cognitive map out of conceptual information is influenced by features of the individual, including cognitive goals, stage in development, and individual differences (Taylor, 2005). The learning of an environment and therefore the building of a mental or cognitive map is most commonly achieved either from a physical map or through navigation (e.g. Montello, 2005) but people can build cognitive maps also from verbal descriptions (Taylor & Tversky, 1996). Graphic representations (maps) or verbal representations (such as written or oral directions) provide the user with a bird's eye view of the geometry of the environment, called *survey knowledge*. While traveling along a real world route the environment is seen from an on-the-ground perspective which leads to the so called *route knowledge* (Lobben, 2004). Survey knowledge and route knowledge refer to the *external perspective* and the *internal perspective* described by Lloyd, Cammack, & Holliday (1995). They describe the external perspective as the viewing of an object from a "fixed vantage point" such as a map and the internal perspective where object and observer are part of the same space. An interesting finding in this context is that it seems to be possible that people develop survey knowledge with increased exposure to an environment, enabling them to visualise a more complete picture of the area (Golledge, 1992). Nevertheless one of the main advantage of the physical map compared to navigation in a real-world environment, is the fact that maps directly present relational information between all landmark pairs, whereas in navigation the relative location information is limited to local groupings. Information about the relational properties between landmarks is important for self-localisation and navigation, because people tend to use landmarks as a reference frame to navigate or orientate themselves (Taylor, 2005). Another important benefit of salient environmental features is that they can be used for aligning the map to where a person is facing or traveling in the local surrounding. This process of map aligning to the surrounding is frequently observed in orientation and navigational tasks and considered to be an important part of successful navigation (Montello, 2005). But the most challenging ability in navigation or self-localisation is the mental transformation of a two-dimensional map into a three-dimensional form in order to visualise the area's characteristics and objects (morphology, streets, and buildings). This transformation is according to Crampton (1992) exactly what a map user has to do to successfully navigate with a map. This challenging transformation is further discussed in the next section from a wayfinding perspective.

2.1.2 Wayfinding Research

In the centre of human wayfinding research is the investigation of the processes that take place while people orientate themselves or navigate through space. Theories developed in wayfinding research are trying to explain how people find their ways in a real-world environment, how they communicate directions, and how people's verbal and visual abilities influence the processes of wayfinding (Raubal & Winter, 2002). Although wayfinding research mainly analyses navigational route planning and route descriptions it is tightly linked to map reading and map perception because environmental features used in route description might also be important for self-localisation on a maps. This connection is also pointed out by Allen (1997), who sees similar problems in wayfinding research as they might accrue when people are reading maps matching the information to the spatial environment. This process of matching the information provided by some sort of route description or map is obviously open for a number of ways which can go awry. It is therefore important to understand this transaction in general to improve verbal direction as well as cartographic map information.

A first step to assure the understanding of this transaction in communicative route description is a definition of the involved processes. Allen (1997) identifies two typical communicative statements when it comes to route descriptions: *Directive* and *descriptive* statements. Directive statements contain verbs like *go* and *turn* providing information about the direction and distance the questioner asked to go. Descriptive statements refer to specific objects on the way like an eye-catching building, shops or any other distinctive environmental feature on the way. Allen (1997) describes *landmarks*, *pathways* and *choice points* as the most commonly used environmental features in route descriptions. A landmark is described as "[…] an environmental feature that can function as a point of reference (Allen, 1997:366)." Landmarks as points of references are helping the "wayfinder" in the way to make a connection to the point of origin and the destination along a specified path of movement. Pathways on the other side refer to where movement in a spatial context is actual possible (streets, sidewalks, or trails). Choice points are closely connected to pathways, because they refer mainly to pathway places with certain movement or taking direction options. The most typical choice points are therefore intersections (Allen, 1997). A more in detail overview of the use and importance of landmarks for route descriptions and for “learning” the environment can be found in an article of Peters, Wu, & Winter (2010). In an experiment conducted by Peters et al. (2010) subjects were asked to write down a route description while “walking” through a virtual environment. One focus of the study was to test what kind of clues subjects use for route description. The findings of the experiment showed that, although the virtual environment was relatively sterile and only populated by textured box buildings, the majority of the participants referred to landmarks in their route descriptions. One of the earlier studies which investigated human descriptions of urban environments was conducted by Lynch (1960) who identified landmarks, along with districts, edges, nodes, and paths as one of the main elements that enhance the ability to imagine city spaces. Many other studies also conclude that landmarks are very important elements for people to learn an environment and are frequently used in route descriptions (e.g. Kiefer et al., 2013; Lovelace, Hegarty, & Montello, 1999; Michon & Denis, 2001). It is therefore save to say, that landmarks play an important role in route directions and are worthwhile to take a closer look at in the next section.

2.1.3 Landmarks

Landmarks are in general defined as “[…] a salient geographic entity that marks a locality and can be used for orientation or navigating in the environment” (Peters et al., 2010:55). A landmark can therefore be any object in the environment that is easily recognizable (e.g. buildings, rivers, parks etc.) or well-known by the navigator (e.g. working place etc.), as long as its primary property is that of a point of reference (Presson & Montello, 1988). The most general requirement of a landmark is that it must be perceptually salient in some sense (i.e. visually, auditory, olfactory, or semantically). This means a landmark has to be distinguishable, either in terms of it attributes (colour, texture, size, shape, etc.) or by its spatial location, from the other environmental objects/features in the scene perceived by the navigator (Caduff & Timpf, 2008). Landmarks are therefore often attributed to distinct objects, such as facades (Peters et al., 2010), churches, or other outstanding buildings (Lovelace et al., 1999; Raubal & Winter, 2002). Nevertheless the definition of measurable features of landmarks is a great challenge. One approach to overcame this challenge was presented by Sorrows & Hirtle (1999) classifying landmarks in terms of visual, cognitive and structural dimensions. Sorrows & Hirtle (1999: 41 ff) propose three different types of landmarks:

1) Visual prominent landmarks
2) Cognitive landmarks
3) Structural landmarks

While visual prominence is described as the visual importance of a spatial feature e.g. a building visible from many locations or standing at a location with road junctions. A cognitive landmark is a landmark standing out through its meaning, either because of its typical or atypical appearance in the context of a certain environment. A cognitive landmark might be cultural important and therefore not for every viewer salient. The importance of a structural landmark is salient to the viewer because of its role or location in the configuration of an environment. An example for structural landmarks could be spaces or intersections. Sorrows & Hirtle (1999) claim that "these three categories, visual, cognitive, and structural landmarks, encompass the reality of differences within the realm of landmarks in […] real spaces" (Sorrows & Hirtle, 1999:46). Based on the defined categories of landmark characteristics by Sorrows & Hirtle (1999) Caduff & Timpf (2008) place emphasis on what they call a trilateral relation between the feature itself, the surrounding environment, and the observer's point of view, both, cognitively and physically. Their central assumption is that "in the domain of navigation, salience emerges from the trilateral relationship between *Observer*, *Environment*, and *Geographic Feature*" (Caduff & Timpf, 2008:253). The term geographic feature is thereby defined according to the classical global landmark "as districts, edges or barriers, rivers or lakes, or unique objects, or any feature of the environment that is recognizable and may serve as a spatial reference" (Caduff & Timpf, 2008:254). Caduff & Timpf (2008) argue that the relationship between observer, environment, and geographic feature is often neglected in attention-based models of landmark extraction commonly relaying on typical visual salient features like colour, intensity, contrast, etc. (i.e. bottom-up driven, see section visual search). These attention-based models of landmark extraction might work well for robot navigation, but not for humans, because contextual and cognitive aspects have to be taken into account, too. Caduff & Timpf therefore defined the following three different types of salience (Caduff & Timpf, 2008:263):

1) Perceptual salience
2) Cognitive salience
3) Contextual salience

Perceptual salience is defined by Caduff & Timpf (2008) as exogenous or passive potential of an object or region to attract visual attention, whereas cognitive salience is considered as an endogenous or active mode of triggering attention. Cognitive salient targets can trigger attention by informative cues which provide advanced information about the location of the target, similar as cognitive landmarks defined by Sorrows & Hirtle (1999). Endogenous attention in wayfinding is also known as *Attentional Orienting* which is characterised by being initiated actively by the person in a top-down (goal-driven) manner and is therefore dependent on the observer's experience and knowledge. The knowledge about an object or an environment of an observer is manifested in a mental representation of the spatial environment. We can retrieve information about an object from this mental representation based on the *Degree of Recognition* and the *Idiosyncratic Relevance* of that object. How well an object can be identified by an observer is measured by the *Degree of Recognition,* while the *Idiosyncratic Relevance* indicates the personal importance of that object to the observer. It is believed that objects with a high degree of recognition or importance to an observer will be more likely used as reference point. In contrast contextual salience is according to Caduff & Timpf (2008) strongly linked to the modality of the task to be performed. They distinguish between two types of context (Caduff & Timpf, 2008:258):

1) Task-based Context
2) Modality-based Context

As the term task-based already implies, *Task-based Context* refers to the navigational task being performed. So for example navigation will be obviously different for a task such as sightseeing than finding a route to a certain location. The *Modality-based Context* takes the mode of travel (i.e. walking, driving or riding) into account which influences the cognitive load put on the observer, as well as the liberties of actions/choices an observer has. In summary the framework presented by Caduff & Timpf (2008) allows the integration of cognitive, observer based saliency and contextual saliency aspects besides the commonly used perceptual saliency aspects in computational models. So it is possible to achieve a more solid assessment of objects navigators may refer to as landmarks when standing at specific decision points along a route (Caduff & Timpf, 2008).

The use of landmarks for conveying route directions are interesting for map makers, because they provide an insight in how people orientate themselves in an unfamiliar or familiar environment. It is therefore interesting if the described landmarks can also be used in the cartographic map production process.

2.2 Vision Research

2.2.1 Visual Information Processing

Information about the functionality and limitations of the Human Visual System (HVS) can be gathered from the vast neuroscientific literature. The functionality and limitations of the HVS are frequently used to qualitatively predict or explain observed psychophysical results. The HVS naturally plays a major role in visual information processing, because it defines how we can perceive and process visual information. Research in neurophysiological and psychological literature on the HVS suggests that we inspect our field of view through series of brief fixations over a small region of interest. Because humans have only a sharp detailed perception in their central foveal vision (about 1-5° of the visual angle), fine scrutiny is only possible at a small part of the entire visual field. Nevertheless we spend about 90% of our viewing time in fixations while the rest of the time is spend to direct our visual attention to a new area of the visual field. These fast repositions of the fovea between the fixations are called saccades (Duchowski, 2007). In figure 3 a graphic of the human eye by Ware (2012) is presented.

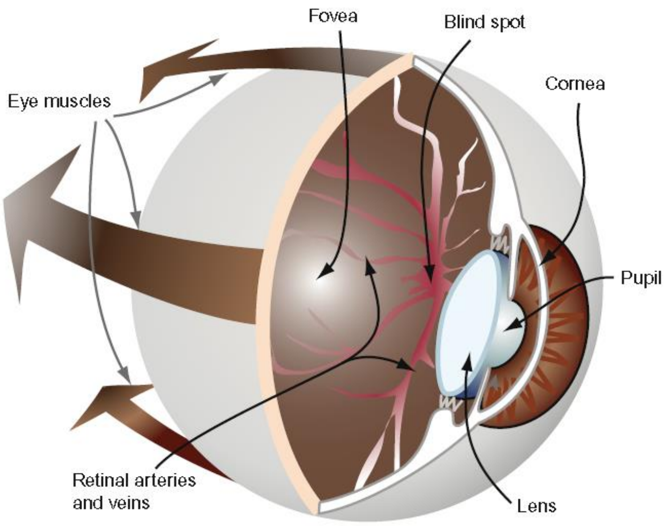


Figure 3: The human eye (Ware, 2012).

The important features of the human eye presented in figure 3 are the fovea, where vision is the sharpest; the pupil, where the light enters the eye, the lens and the cornea two principal optical elements; and on the side the large muscles who are controlling the eye movements.

Where the arteries enter the eyeball no receptors are present causing the blind spot (Ware, 2012). After the light enters the eye the visual stimuli are perceived and stored in the sensory memory. This first state does not involve any processing of the visual stimuli. The processing of the visual stimuli starts in the second state of the human visual information processing, when stored in the working memory. When stored in working memory pre-attentive processing starts, stimulating the low-level human visual system to rapidly discriminate objects and identify certain basic properties (Treisman, Vieira, & Hayes, 1992). During the pre-attentive processing a perceptual representation of the spatial scene is built in the working memory containing low-level components (e.g. size, length, colour, intensity) of the spatial objects in the scene. This perceptual representation in the working memory can then be further processed. While visual stimuli in the sensory memory are processed in parallel, objects in working memory are processed sequential, allowing us to solve goal-driven (top-down) tasks. The last step of the human information processing model involves the storage of objects in the long-term memory. When objects are already present in the long-term memory they are updated with the newer information perceived (Caduff & Timpf, 2008). The before described processes of human information processing are graphically represented in figure 4.

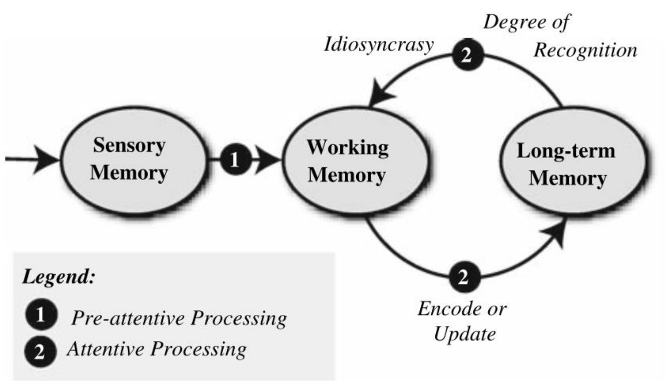


Figure 4: Human information processing model (Caduff & Timpf, 2008:260).

The graphically representation of the human information processing model by Caduff & Timpf (2008) summarises the main processes involved in human information processing. The model assumes that in each stage a representation of the spatial scene is hold; these spatial scenes are then either pre-attentive or attentive processed (Caduff & Timpf, 2008).

In the context of visual information processing and also as a concept in visual attention the terms *top-down* and *bottom-up* information processing are frequently used. Bottom-up information refers to the information on the map or in general in the environment of the observer. Bottom-up activation will therefore be the same, no matter what is expected of the target (Lloyd, 1997). One example of a bottom-up activation would be the visual "pop-out" effect where a target with a certain salient feature is detected equal fast whatever the number of distractors is (Treisman & Gormican, 1988). Top-down information on the other hand refers to the knowledge an observer already has stored in memory about a certain item. Top-down activation depends therefore on the searcher's knowledge of the target and it is therefore often referred to as goal-driven attention (Lloyd, 1997). The interaction between bottom-up and top-down information in visual search is of major importance in general visual information processing and will therefore also be further discussed in the next section about visual attention theories in visual search.

2.2.2 Visual Attention Theories in Visual Search

How do we identify objects and analyse their spatial relationship? Well in primates the identification is usually achieved by rapid, saccadic eye movements to bring the fovea onto the object, or to covert shifts of attention. The amount of information the visual system has to cope with is immense and it exceeds the capability of the brain by far. It is therefore obvious that the brain has to use a strategy which can deal with its limited resources. According to Itti & Koch (2000) "the strategy nature has devised for dealing with this bottleneck is to select certain portions of the input to be processed preferentially, shifting the processing focus from one location to another in a serial fashion" (Itti & Koch, 2000:1489). This means that although it seems that we see everything around us, we actually perceive only a very small part of our environment at the time. Visual attention can therefore be defined as "[…] the process by which one grants priority among sources of visual information" (Cameron, Eckstein, Tai, & Carrasco, 2004) . Now that we know what visual attention is, we can spend some more thoughts about the really challenging question: What deploys our attention? Visual attention research has been playing an important role, mainly in psychology, for over a century now. Over the last 50 years visual attention research has become more and more an interdisciplinary subject involving a variety of disciplines such as psychophysics, cognitive neuroscience, and computer science and others (Duchowski, 2007).

In the following sections a short and not conclusive summary of the main theories of visual attention is given. Starting with the well-known attention model; the Feature Integration Theory of Attention by Treisman & Gelade (1980).

Feature Integration Theory: The Feature Integration Theory (FIT) suggests that visual attention is basically a serial process, allowing only one item at the time to be processed. This conclusion was supported mainly by the results of feature search displays and conjunction search displays. A feature is thereby defined as "[…] a particular value on a dimension (Treisman & Gelade, 1980:99)" whereas dimension refers "[…] to the complete range of variation which is separately analysed by some functionally independent perceptual subsystem" (Treisman & Gelade, 1980:98). So for example "red" would be a feature of the dimension "colour" and "vertical" would be a feature of the dimension "orientation". A conjunction would refer in this case to the combination of two separable features for example a red "O" among red "X's" and yellow "O's". In the FIT it is assumed that a visual scene is initially coded in different dimensions such as colour, orientation, shape, spatial frequency, brightness, and direction of movement and only by visual attention these separate features are integrated into unitary objects. This assumption is, as mentioned before, based on the dichotomy of reaction time (RT) results of feature search displays and conjunction search displays as a function of set size. In general it would be assumed, that the RT would be increasing linearly with the increase of the set size. However, in the experiments conducted by Treisman & Gelade (1980) the RT did only increase with the set size for the conjunction search displays and not for the feature search displays. The FIT tries to explain this dichotomy through the assumption that "[…] features are registered early, automatically, and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention" (Treisman & Gelade 1980:98). So to characterize or distinguish objects consisting of conjunctions of more than one separable feature, attention has to be directed serially to each stimulus presented in a display which will lead to a higher RT. On the other hand, separable features are registered in a pre-attentive step and because of this the RT is independent of the set size. In essence, the FIT describes attention as a "glue" integrating separable features in a particular location into conjunctions (i.e. object) which can then be perceived as a unified whole. Before attention is deployed we can only perceive where certain feature boundaries are located, but not what those features are (Treisman & Gelade, 1980). Another well-known attention model is the Guided Search Theory by Wolfe, Cave, & Franzel (1989), currently in the actualised fourth version (Wolfe, 2007).

Guided Search Theory: The Guided Search Theory (GS) was initially proposed by Wolfe et al. in 1989 as an alternative theory to the Feature Integration Theory (FIT) by Treisman & Gelade (1980). The original GS model proposed, much like the FIT, a pre-attentive stage and an attentive stage. The first version of the GS model claims that information from the first, pre-attentive stage could be used to guide deployment of selective attention in the second, attentive stage. In the first GS version Wolfe et al. (1989) argue that the difference in task performance depends on the differences in the quality of guidance by the target item. Underlying this assumption was the conclusion that all search tasks require attention to be directed to the target item. So in general the GS model retains the FIT's covert attention as a serial processor, but unlike the FIT, attention is not randomly deployed across items in the display but is guided by parallel processing across the visual field of elements in the display (Wolfe et al., 1989). The GS model combines bottom-up and top-down properties of the elements during the processing of visual stimuli. When displayed items exceed an certain activation threshold, attention processes each item in a serially fashion starting with the item with the highest activation followed by the 2nd highest activation an so on (Wolfe, 1994a).

The GS model did undergo a series of changes since the first version in 1989. The second version of GS model (GS2) (Wolfe, 1994a) was a revision of the first GS model trying to make all aspects of the model more explicit in the light of new data. In GS2 it was also tried to account for the termination of search in target-absent trials. The third version of the GS model (GS3) (Wolfe & Gancarz, 1997) was an attempt to integrate covert deployments of visual attention with overt deployments of the eyes. The fourth and current version of the GS model's (GS4) large-scale structure is presented in figure 5 (Wolfe, 2007:103). The numbers in figure 5 refer to the descriptions provided later in this text.

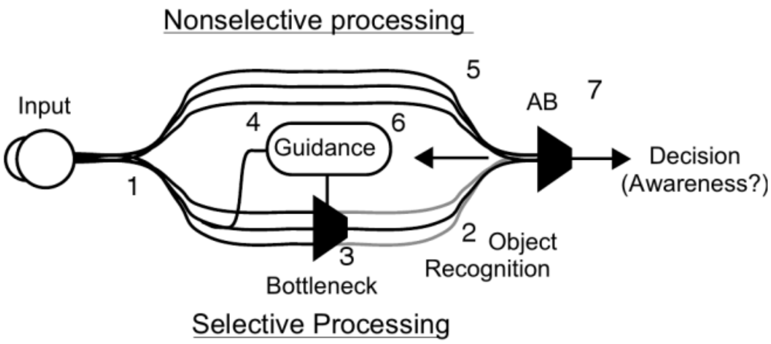


Figure 5: Large-scale structure of the GS4 model (Wolfe, 2007:103).

Figure 5 represents the large-scale structure of the GS4 model by Wolfe (2007). According to GS4 visual input is processed parallel in early vision (1) providing the input information for the object recognition processes (2) via a mandatory selective bottleneck (3). The selection process in (3) is controlled by a "guiding representation" (4), which is an abstraction of the early vision output. The GS4 assumes that a limited number of attributes (12 to 24) can guide the deployment of attention. But the extent of the attention deployment differs between different attributes. So for example salient colours are known to work very well for guiding attention, while the attention guiding properties of other attributes are still part of a lively debate. Besides the visual information processing path limited by the selective bottleneck (3), there seems to be also a nonselective processing path (5). This assumption is made because some visual tasks seem to be not limited by this selection process including for example the analysis of image statistics and some aspects of scene analysis. So to include these circumstances in the GS4 model, the second pathway, bypassing the selective bottleneck, was implemented. Because it seems likely that selection can be also guided by scene properties extracted in the second, nonselective pathway, for example the intuition of where people most likely are in an image (e.g. not in the sky), the GS4 has also a connection back (6) to the "guiding representation" (4). The visual outputs of both selective (2) and nonselective (5) pathways are subject to a second bottleneck (7) which is believed to limit the performance in so called attentional blink (AB) tasks. In AB experiments the decision and response mechanisms are analysed in rapidly presented visual sequences. So in summary the GS4 model differs from typical parallel models in the assumption, that the accumulation of information begins not at the same time for each item, but only when an item is selected. This means that if each object needs to wait till the object before is finished the GS4 becomes a strict serial process, but if objects can be processed at the same time it is a parallel model. So in general the GS4 model proposes a hybrid form with both serial and parallel properties (Wolfe, 2007).

Besides the two most known visual attention models described here in this master thesis, there are of course a great variety of other visual attention models for example the Attention Engagement Theory by Duncan & Humphreys (1992). The Attention Engagement Theory argues that the slope for the relationships between reaction time and number of distractors can take on any value based on the difficulty of the search and proposes therefore the elimination of the distinction between parallel and serial processing. At this point the interested reader is advised to read the review by Carrasco (2011) for further information.

2.2.3 Attention Model Applications

The modelling of visual attention has been a very active research over the past 25 years leading to a rich theoretical background and a variety of successful applications in computer vision, mobile robotics, and cognitive systems. In the centre of scientific interest are particularly stimulus-driven and saliency-based attention models (Borji, Sihite, & Itti, 2013). One of the most successful computational saliency models is the computational model of Saliency-Based Visual Attention by Itti, Koch, & Niebur (1998). According to Itti & Koch (2000) most computational visual search task models are based on so called saliency maps. A saliency map is a two-dimensional map encoding the saliency or conspicuity of objects in the visual environment. In their study "A saliency-based search mechanism for overt and covert shifts of visual attention" Itti & Koch (2000) describe a computer implementation of a visual search task model based on a saliency map. Their focus is thereby mainly laid on different modalities such as orientation, intensity and colour information. The original structure of the saliency model is presented in figure 6 (Itti et al. 1998:1254).

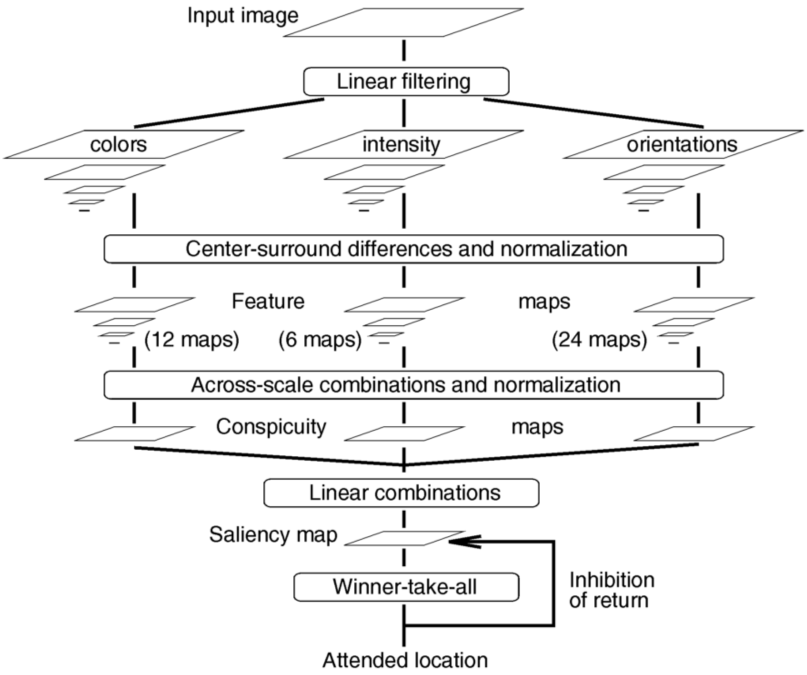


Figure 6: Architecture of the saliency model by Itti et al. (1998:1254).

The saliency model by Itti et al. (1998) is related to the Feature Integration Theory by Treisman and Gelade (1980). As presented in figure 6 the model starts by splitting the input image into a set of topographic feature maps containing features like colour, intensity and orientation. In each feature map a spatial location which locally stands out from their surrounding is determined. In the next step all feature maps are summarised to a master saliency map encoding the local conspicuity over the entire visual scene. The original saliency model by Itti et al. (1998) is a purely bottom-up model which means that it does not account for any top-down guidance of attention shifts. The original framework provides a "[…]

parallel method for the fast selection of a small number of interesting image locations to be analysed by more complex and time-consuming object-recognition processes" (Itti et al., 1998:1254). The original model has been extended in many directions since 1998, so for example by including an eye and head movement model to animate human eye/head movements (Itti, Dhavale, & Pighin, 2003) or by including also top-down task demands (Navalpakkam & Itti, 2005).

2.2.1 Clutter and Visual Attention

Clutter is mainly important because it influences your ability to find objects or persons in a crowded environment. A number of computational models exist in the recent scientific discussion which seem to work well with different simplified search task scenes.

Asher, Tolhurst, Troscianko, & Gilchrist (2013) investigated the role of scene clutter in predicting search performance in a natural image visual search task. In their study they evaluate the following clutter models: Feature Congestion (Rosenholtz, Li, & Nakano, 2007), Sub-band Entropy (Rosenholtz et al., 2007), Segmentation (Bravo & Farid, 2008) and Edge Density (Mack & Oliva, 2004). All these models are trying to predict: "[…] how hard a particular search task will be" (Asher et al. 2013:1). Interestingly they find that clutter (defined by the models mentioned above) is rather weakly correlated with the performance for natural scenes. Unlike previous studies (e.g. Rosenholtz et al., 2007) Asher et al. (2013) could not find any significant correlations between clutter measures of the picture and response time. The authors try to explain this circumstance with the more challenging task used in their study. They conclude that search performance depends mainly on the nature of the target and the relationship between the target and the background in natural pictures rather than on the clutter of the scene.

2.2.2 Shape and Visual Attention

Attneave & Arnoult (1956) define shape as a multidimensional variable not to be confused with a single dimension variable like for example brightness and hue. The definition of shape is so difficult because "[…] the number of dimensions necessary to describe a shape is not fixed or constant but increases with the complexity of the shape" (Attneave & Arnoult, 1956:452). And even if we know how many dimensions we need to describe a shape with, the particular description still remains a problem because some dimension may have more of an impact than others. An adequate framework is especially of importance in studies in which it is necessary to manipulate shape or pattern as an independent variable. This is of major importance for the generalizability of the experiment stimuli and therefore of results. So to get results with “ecological validity” experimental materials connected directly to the real situation have to be used (Attneave & Arnoult, 1956). For the goal of this master thesis this means taking natural shapes like they are used in common map representations but trying to classify them in certain shape “families” sharing specific physical features.

Shape perception is also in a variety of computer science studies a broadly discussed issue. One example is the work of DeCarlo & Santella (2002) introducing a computational system transforming natural images into line-drawing styled pictures with separated regions of homogeneous colour. Their algorithm for the creation of these abstract pictures uses bold edges and large regions of constant colour, while the content of the pictures is selected by using eye movement data from human users.

2.2.3 Text and Visual Attention

Bartz (1970) studied with an experimental search efficiency task the legibility of different typefaces used in cartographic maps. In the experiment subjects were asked to find six names on various maps and under a number of different conditions, including typographic variation on the list of names used for the search and on the map themselves. For this first part of the experiment Bartz (1970) could not determine any typographic form which was particularly superior to any other typographic form tested in the search task experiment. Though she found an increase by 300% in search time efficiency for a mixed-type test condition, in which the typographic properties of the search list names were the same as in the map. This circumstance led her to the conclusion, that the searcher’s expectation about the target appearance outperformed the difference in typographic form by far. One explanation for the significant faster search performance for a mixed-type typographic condition is, that the possible target names in the map are reduced to the ones with the same typographic properties represented in the search name list (relevant names). In general the findings of the experiment suggest that the typography of a name is of less importance than many other factors like the figure-ground relation or the location in the page and so on. In the opinion of Bartz (1970) it is therefore not valid to apply results from “text reading” experiments to “map reading” experiments.

3. Methods

In this chapter the research methodology and in particular the methods used in this master thesis are described and discussed. In a first step the stimuli which will be analysed in the user study are operationalised. According to the findings in wayfinding research landmarks, pathways and choice points are the most commonly used environmental features in route description (Allen, 1997). These three categories are the basis of the analysed stimuli in this master thesis. Landmarks will refer mainly to buildings present in a scene and the map representation, whereas for the purpose of this master thesis pathways and choices points will be mainly considered as road junctions. The third stimuli analysed in this user study will be street labels. After the first section describing the operationalization of the stimuli the measuring instruments, the preparation of the experiment displays and the procedure of the user study will be discussed in detail.

3.1 Operationalization of the Stimuli

The basic assumption for the design of the stimuli in the search and self-localisation tasks is that people can use different environmental clues such as buildings, road junctions and street labels to orientate themselves. These three features are not only present in the real world environment but also on most commonly road maps used for navigation or self-localisation. According to these three features the maps used in this experiment are in a first step divided into three different layers, containing only information about one of the before described features. This step leads to the in figure 7 presented basic main test conditions of the 2D layout map search task for the map representations “cartographic” and “satellite”.

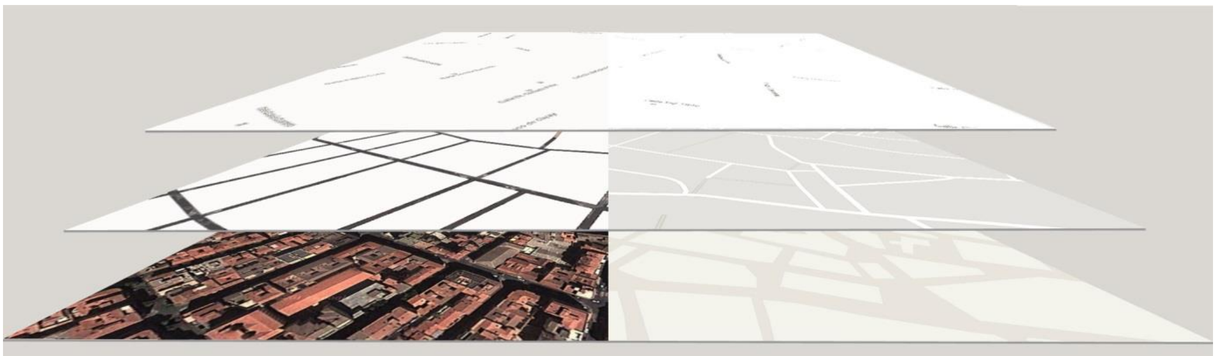


Figure 7: Test conditions “label”, “street” and “background (top to bottom).

Each condition is represented as a layer in figure 7. The top layer represents the label layer, the middle layer stands for the street layer and the bottom layer represents the background layer considered to contain the information about buildings which can be used as landmarks. All three conditions are tested for the main categories “cartographic map” and “satellite map”. Because in the experiment real-world scene pictures as well as real-world cartographic map representations are used, it is hard to control the quantitative appearance of the stimuli. The three stimuli are therefore defined in a qualitative way referring to the knowledge people have about them. This approach is also frequently used in remote sensing as a combination of Geographic Object-Based Image Analysis (GEOBIA) and ontologies. GEOBIA is devoted to develop automated methods to partition remote sensing imagery into meaningful image objects and assessing their characteristics through spatial, spectral and temporal scales, whereas ontology intends to identify concepts and their relationships within a domain. GEOBIA has become popularized in remote sensing because it allows the use of semantics, based on descriptive assessment and knowledge. This means that the approach incorporates the wisdom of the user and is not only dependent on pixel-based classification (Arvor, Durieux, Andrés, & Laporte, 2013). In a very simplified way this means that although the here defined stimuli may not be homogeneous in their spectral appearance, they can still be identified as part of a certain class (i.e. label, street, background/landmarks) because of the knowledge people have about them. Nevertheless it is important to consider attention guiding features of the stimuli chosen in the search task experiment. The search results for each stimulus will therefore be checked with a saliency map produced with an algorithm by Harel et al. (2007). This should prevent strong influence by bottom-up saliency. Examples of the saliency maps can be found in the appendix.

The layer concept presented in figure 7 is also well-known in cartography and map production where labels, information about the street network, and the cadastral map are often stored in different layers (Kraak & Ormeling, 2003). To control to some extend the influence of the environmental context each stimulus is also tested in combination with every other stimulus. So the experimental design used in the 2D layout map search tasks consists of the combinations shown in figure 8.

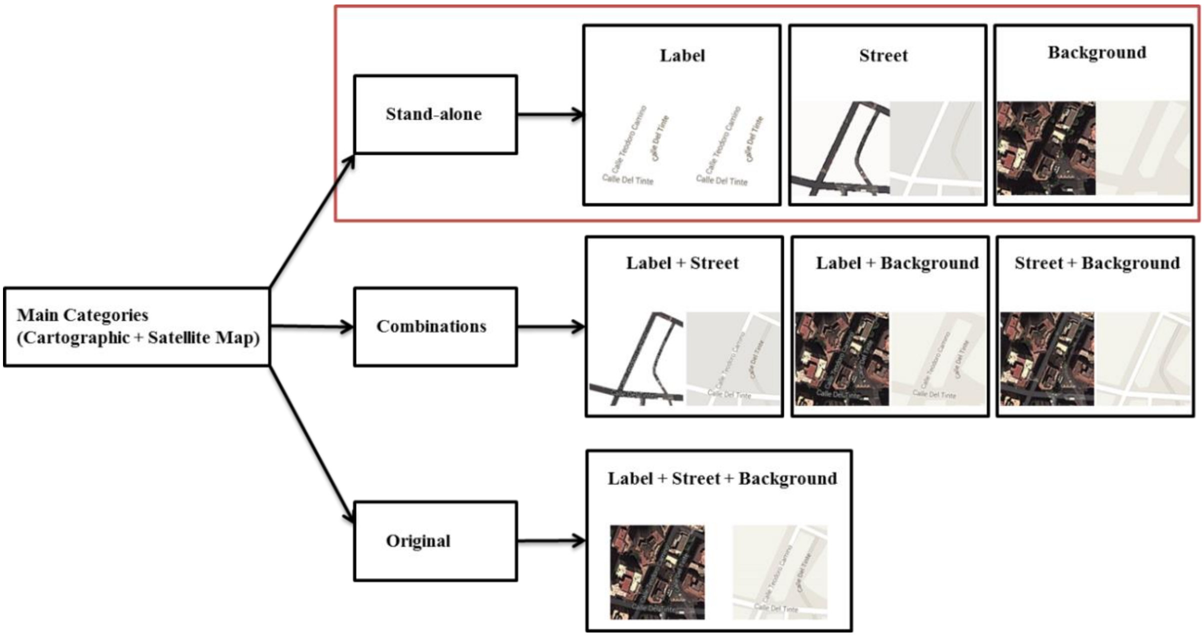


Figure 8: Subcategories used in the 2D layout search tasks.

Figure 8 shows the classification of the different subcategories used in the 2D layout map search tasks. The red framed part of figure 8 refers to figure 7 defining the three test conditions “Label”, “Street” and “Background”. They represent the basic test conditions. Additionally to the basic test conditions the combinations of each of these basic conditions are also tested as subcategories. Also the combination of all basic layers is tested in the 2D layout map search tasks. Figure 8 also provides an example picture for each subcategory for the satellite (left) and the cartographic map representations (right).

The results of the 2D layout map search task will be used to answer research question 3 (RQ3): How efficient can the corresponding 2D layout feature of the environmental features defined in research question 1 be found on a map?

In summary the three semantically defined stimuli “Label”, “Street” and “Background” are defined as the independent variables whereas the reaction time needed to find these stimuli in either a 2D layout map or in combination with a 3D environmental scene picture is defined as the dependent variable. The emphasis is placed clearly on the recognition of environmental features (i.e. in this master thesis street labels, road junctions, and landmarks/buildings) in a map, so although the amount of clues represented in the environmental pictures is controlled as much as possible, no in depth scene definition is provided.

3.2 Map Data Used

All cartographic data is derived from the online map service Google Maps[1]. The main reason for this decision is the current dominance of this map representation used for navigation and self-localisation. According to Google Trends[2] the worldwide interest in the online service Google Maps has more or less increased continuingly since 2005 as can be seen in figure 9.

Figure 9: Google Maps Website interest over time (Google Trends[2]).

Besides the online map services Google Maps, an increasing number of users use Google Maps on their smart phones as an app. For example in the USA approximately 95-100 million unique visitors per month are counted for the Google Maps website, which is about 40% of the US population, while the number of smart phone visitors did already increase to 92 million unique visitors. This means that already as much people use Google Maps on their smart phone as users use the Google web map service at least in the USA (comScore[3]).

The scene pictures for the 3D to 2D layout map search task were taken out of the Google service Google Street View[4].

3.3 The Laboratory

The eye movement laboratory where the user study took place is located in the Geography department (room Y25-L-9) on University of Zurich's Irchel campus. The technical setup of the lab consists of an eye tracking system, a Dalco workstation and an Estecom display with the following specifications represented in table 1:

---

[1] https://www.google.ch/maps
[2] http://www.google.ch/trends/explore#q=%2Fm%2F055t58&cmpt=q. Access: 26.01.2014
[3] http://www.comscore.com/Insights/Blog/Map_Searches_Shift_from_Desktops_to_Smartphones. Access: 26.01.2014
[4] https://www.google.ch/maps

Eye tracker: Tobii Model:TX300Hz (Binocular) Accuracy: 0.4 Deg. Freedom of movement: 37x17:56 Software: Tobii Studio	**Workstation: Dalco** CPU: Intel Core i5 760 (2.80 GHz, 8 MB Cache) RAM: 16 GB Disk: 2x500 GB (7,200 rpm) SATA II Video: GeForce GT 430 OS: Windows 7 Enterprise (SP 1)	**Display: Estecom** Size: 23'' diagonal Max Resolution: 1920 x 1080 pixel Colour support: 16.7 M (HiFRC) Image aspect ratio: 16:9 Response time: 5 ms Video signals: DVI/VGA, USB

Table 1: Technical setup of the eye movement lab.

More information about the eye movement lab of the geovisualisation unit is provided at the following website: http://www.geo.uzh.ch/en/units/giscience-giva/services/eye-movement-lab.

3.4 Measuring Instruments

In this section the different measuring instruments already introduced in the chapter "Methodology Overview and Expected Results" are shortly described.

3.4.1 Visual Search

The main part of this master thesis is the conduction of a visual search experiment. In a standard laboratory visual search experiment, subjects are asked to search for a certain stimulus among distractors. In the standard visual search paradigm the number of distractors is varied and the response time is measured. However, with natural images the traditional visual search task is difficult to conduct, because the traditional visual search paradigm involves the presentation of the same stimuli hundreds of times with a random placement of target and distractor stimuli. Another problem is that it is impossible to determine the set size in pictures or maps representing natural stimuli (Wolfe, 1994b). In the visual search experiment conducted in this master thesis, neither the set size is controlled, nor is the placement of the target and distractors randomised. Nevertheless the basic requirement of the visual search paradigm is met, because subjects still have to search for a certain stimuli among distractors while the response time is measured. The context in which the visual search experiment takes place is of course less controlled and the resulting results may be more questionable. But therefore this visual search experiment has one important advantage that is of course the higher ecological validity.

3.4.2 Eye Movement Recording

Eye movement recordings are often used to determine the deployment of attention in visual search tasks or scene viewing. Eye tracking provides the possibility to follow someone's path of attention while watching a picture or a map. Based on these data we can also make assumptions about what the subject found interesting or how a person perceived a scene or a map. It is therefore assumed that eye movements are bound to visual attention and that this visual attention can be measured by recording the eye movements (Duchowski, 2007). In this master thesis the recorded eye movements will mainly be used to determine the response time (RT) during the conducted visual search experiments.

3.4.3 Online Survey

The online survey was filled out by most of the participants in advance of the experiment. They needed approximately 15 minutes to answer the 30 questions asked. The responses were collected and analysed with the help of the online survey platform SurveyMonkey[5]. The survey can be subdivided in three blocks of questions with regard to the content. The first block of questions contains questions considering general background information like age, gender and education. The second block of questions contains questions which refer to the individual Google Maps usage. The third block of questions consists mainly out of the 15 questions of the Santa Barbara Sense of Direction test by Hegarty, Richardson, Montello, Lovelace, & Subbiah (2002).

[5] surveymonkey.net

3.4.4 Santa Barbara Sense-of-Direction Scale

The Santa Barbara Sense of Direction Scale (SBSOD) by (Hegarty et al., 2002) is a self-reported sense of direction (SOD) test. Self-reported means that people are asked to rate their own sense of direction. The SBSOD consists of 15 questions, each with a 7-point Likert-scale (1 = *strongly agree* to 7 = *strongly disagree*) asking the subjects how strongly they agree or disagree with the question asked. Hegarty et al. (2002) found that the results of the SBSOD are related to task results that require self-localisation. The SBSOD scale is used in this master thesis to determine individual performance differences and their possible influence on the visual search response time. The questions of the SBSOD scale can be found in the appendix.

3.4.5 Spatial Orientation Test

Kozhevnikov & Hegarty (2001) developed a spatial orientation test measuring the ability to imagine different perspectives or orientations in space. In this master thesis the revised version by Hegarty & Waller (2004) of the original spatial orientation test by Kozhevnikov & Hegarty (2001) is used. The test consists of 12 items, one on each page. The test can be found in the appendix.

3.5 Experiment Display Preparation and Specifications

3.5.1 2D Layout Map Search Task

The first eye movement recording session consisted of 52 2D Google Maps scenes. 22 scenes were taken out of a cartographic Google Maps representation and 22 scenes were from the satellite Google Maps representation. 2 scenes were only showing the label structure of a map so it does not fit in either category. The remaining 6 maps were a combination of a cartographic search reference and a satellite target.

The main focus lies on 26 pictures (12 cartographic, 12 satellite, 2 label) in which the stimuli “Label”, “Street” and “Background” were systematically varied. The remaining 26 pictures were used as control variables for texture in the satellite map representation (6 pictures) and as partial stimuli test in the original cartographic and satellite map representation. The presentation of the pictures was randomised.

Task Display Design: The design of the search task pictures of the 2D map eye movement recording session is shown in figure 10.

Figure 10:Instruction screen (Schaffhauserplatz, Zurich) and 2D map search task picture design example.

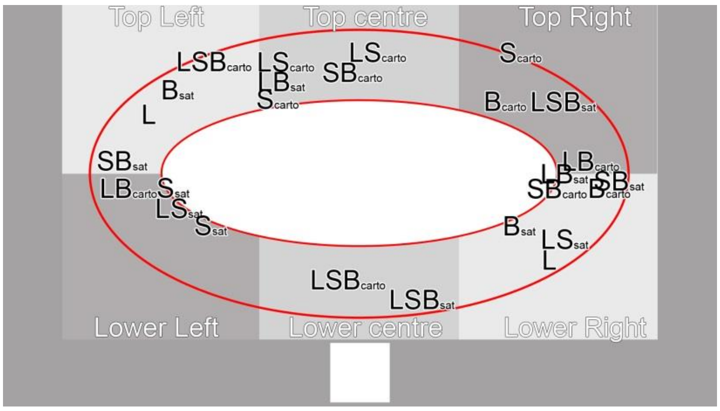
As you can see in figure 10 the search task picture design consists of a larger target search area with a smaller search reference underneath. The search reference represents a map detail out of the larger target search area. The search reference is not altered in any way. The display has a resolution of 1920 x 1080 pixels. The target map within the grey area has a size of 1600 x 900 pixels and the search reference has a size of 160 x 160 pixels.

Search Task Target Picture Location: The location used for the 2D map search tasks is Albacete, a town approximately 200 km southeast of Madrid, Spain (WGS84: 38° 59′ 44″ N, 1° 51′ 21″ W). Albacete was chosen mainly because of its symmetrical streets and square houses like a composition. Also of major importance for this choice was the colour uniformity of the satellite images and the assumption that the town would not be very familiar to the subjects.

Search Task Target Picture Preparation: The 52 2D Google Maps scenes were produced with the help of Google Maps API Styled Maps Wizard[6]. Basically the Styled Maps Wizard allows changing the visibility of different Google Maps layers. For more detailed information see also Google Maps JavaScript API Version 3, section Styles[7].

Pictures were selected mainly around the town centre, trying to equal out colour and texture differences as good as possible. For the picture preparation a Google Maps zoom level of 19 was used. The different layers and combinations of layers produced were then copy pasted as a print screen (72dpi) in Adobe Photoshop CS4 and sized and arranged for the 2D Google Maps search task. It was tried to minimize the search task picture overlay. In special cases in which this was not possible, part of the "no label" picture category were rotated and/or mirrored to prevent a learning effect during the search task experiment.

Search Reference Selection: All search references were selected within a certain distance range from the centre of the search task picture. The distance range where all search references are located in the target map is presented in figure 11. Figure 11 shows also where each of the 26 main subcategories search references where located in the target map. The locations of the search references were classified for further statistical analysis in 6 classes: Top Left, Top Centre, Top Right, Lower Left, Lower Centre and Lower Right.



Cartographic Map	
Top Left	1
Top Centre	4
Top Right	3
Lower Left	1
Lower Centre	1
Lower Right	2

Satellite Map	
Top Left	2
Top Centre	1
Top Right	1
Lower Left	3
Lower Centre	1
Lower Right	4

Figure 11: Locations of the search references in the target search pictures. The letters L = Label, S = Street and B = Background stand for the subcategories.

All search references are located between the two red ovals. The oval distance range was chosen because empirical findings suggest that the target detectability does not decline equally for the horizontal and vertical axis. Studies have reported better performance in the

[6] http://gmaps-samples-v3.googlecode.com/svn/trunk/styledmaps/wizard/index.html
[7] https://developers.google.com/maps/documentation/javascript/styling?hl=de-DE

horizontal than the vertical meridian of the visual field, also called *horizontal-vertical anisotropy* (HVA) (e.g. Carrasco, Talgar, & Cameron, 2001). Eckstein (2011) suggests that the HVA has to be taken into account to enhance visual search task performance tests.

Besides the location of the search references the selection of the stimuli is of major importance. The search references were selected according to the following rules:

Label:

- Label orientation between ± 20° degrees from horizontal direction

Street:

- Road junctions with more than two streets or more than one road junction present in the search reference
- At least more than one angle smaller/bigger than 90° degrees

Background:

- Polygon with 6 to 8 vertexes (rectangle with one distinguishable extra feature)
- Polygon orientation approximately northwest (315°) or northeast (45°)
- Polygons with similar colours
- Polygons with similar size

More selection rules would be preferable but almost impossible to implement if real world map representations are used in a search task experiment.

Control Variables: 26 pictures were used as control variables. 6 pictures were used to determine the effect of texture on the search efficiency (reaction time) in the satellite map representation. For this reason the participants had to find a cartographic map representation search reference in a satellite target map. The 6 different pictures are testing the stimuli "Background" alone and in the combinations "Street-Background" and "Label-Street-Background". 10 of the remaining 20 pictures were cartographic or satellite map representations respectively testing the following stimuli:

- Background search reference in a Label-Street-Background target map
- Background search reference in a Street-Background target map
- Street search reference in a Label-Street-Background target map
- Street search reference in a Street-Background target map

- Label search reference in a Label-Street-Background target map

Although the stimuli differ between search reference and target map, no differences between search reference and corresponding search target area were implemented. The search references of the control variables however are containing more information than the more controlled and systematic varied main 26 search references not used as control variables. For examples in the “Background” conditions there are also partly streets visible in the search references. It was however tried to reduce the information content of the not target stimuli as much as possible. Labels, icons and arrows were in all no “Label” conditions removed in the search references as well as in the search target areas (but not in the rest of the map). Also for the “Street” condition it was tried to reduce the effect of the visible background layer by selecting search references containing only little distinguishable background layer information. In the “Label” condition the participants were asked to find a certain road junction. Instead of a search reference underneath the target map a written task description was shown in the manner of the following example: “Find the intersection between Calle Jesus Nazareno and Calle Nueva”.

Task Description: The goal of the first eye movement recording session was to find the search reference map details in the larger target map and click on it with the mouse. To proceed to the next search display the space key had to be pressed.

3.5.2 3D Google Street View to 2D Layout Map Search Task

The second eye movement recording session consisted of 12 pictures representing a scene out of Google Street View. Each Google Street View picture was displayed in combination with a spatial corresponding map representation and a small picture representing different angles of view. In one half of the pictures the subjects had to pinpoint their location with the help of the Google Street View pictures on a cartographic map representation and on the other half on a satellite map representation. The stimuli used in this search task were once again the stimuli “Label”, “Street” and “Background”.

Task Display Design: In figure 12 the design of the Google Street View search task display is represented.

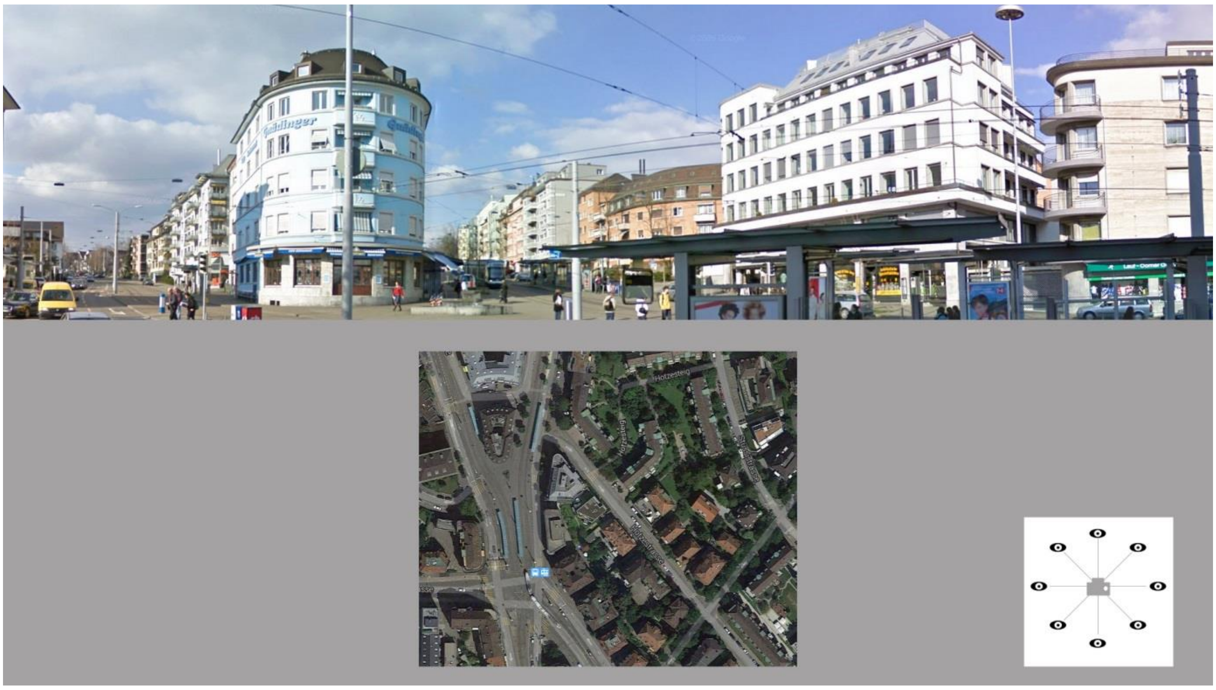


Figure 12: Instruction screen (Schaffhauserplatz, Zurich) and Google Street View search task picture design example.

As can be seen in figure 12 the display contains three relevant areas. In the upper area of the display the pictures from Google Street View are positioned. Underneath a small map is shown containing the location represented by the Google Street View picture. On the right side of the map representation a small graph is positioned representing a selection of possible angles of view. The centre of the graph represents an iconic camera and the surrounding eye-icons represent the fields of vision of the camera. The Google Street View pictures have a resolution of 1920 x 500 pixels, the corresponding maps have a resolution of 600 x 500 pixels and the orientation graphic has a resolution of 236 x 236 pixels. The display size is 1920 x 1080 pixels.

Google Street View Picture Location: Albacete, Spain (for more details see the description in the first eye movement session part).

Task Picture Preparation: The 12 panoramic Google Street View pictures were copy pasted as print screens in Adobe Photoshop where they were cropped and arranged with the map and the orientation graph. The map representations were also copy pasted as print screens in Adobe Photoshop. Because for the Google Street View eye movement recording session the stimuli were only varied in the Google Street View pictures, the map representations were not altered but left in the original Google Maps design. The orientation graph was also produced in Adobe Photoshop. The angel between the icon-eyes is 45° degrees so that the icon-eyes are representing the cardinal directions North, Northeast, East, Southeast, South, Southwest, West and Northwest. As background colour of the display a darker grey was chosen (RGB: 164, 163, 162). The map representations have always a north orientation.

Task Picture Selection: The Google Street View pictures were selected according to the predefined stimuli “Label”, “Street” and “Background”. In every picture at least two (intended) clues were presented which should help to solve the positioning task in the map representation. So for the “Label” conditions at least two labels (street names or restaurant names) were clearly visible and readable. For the test conditions “Street” and “Background” it was significantly harder to select reasonable clues. For the “Street” conditions distinguishable street intersections were chosen, comparable to the selection process of the search references in the 2D map “Street” condition. For the “Background” condition the suitable selection of Google Street View pictures and their corresponding map representation was even more challenging. The problem is that Google Maps uses a combination of real estate borders and actual building forms in the cartographic map representation. Nevertheless it was tried to select Google Street View pictures showing a building distinguishable in the Google Street View picture and in the corresponding map representation as well.

Task Description: The goal of the Google Street View search and location task was to find the locations in the map representation where the Google Street View pictures were taken. The location of the camera had to be determined by clicking with the mouse on the location in the map. In a second step participants had to determine the orientation of the camera in relation to the map representation by clicking on the corresponding eye-icon in the orientation graph. How the determination of the orientation of the camera can be achieved is shown in figure 13.

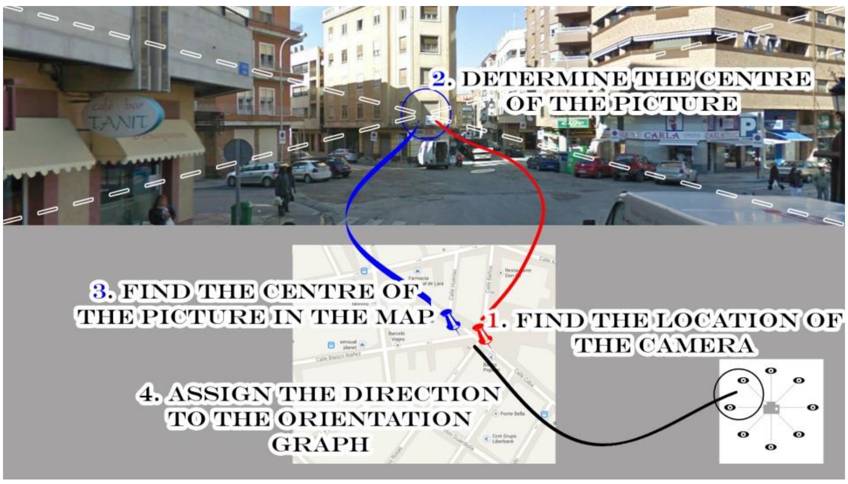


Figure 13: Explanation of the camera orientation determination task subjects had to solve.

The four steps described in figure 13 to determine the orientation of the camera in relation to the map representation were used to explain the task. The instruction display in figure 12 was used for practicing the task and for answering questions of the participants. The participants were asked to think aloud during the experiment.

3.6 User Study Procedure

The user study took place at the eye movement laboratory at the University of Zurich (Y25-L-9). The user study consisted of an online survey, two eye movement recording sessions and an object perspective taking test. The stimuli were presented on a computer screen; except the object perspective taking test which was conducted as a pen and pencil test. At the beginning of the user study the participants had to read and accept a consent form (see appendix). The task was described to the participants with the help of a printed version of the instruction screen which represents the design of the following task pictures. The first eye movement recording session took the participants approximately 20 minutes to complete. Afterwards they were offered a break if they needed one. The user study continued with the second part of the eye movement recording session. Here the participants were also shown first the printed version of the instruction screen to describe the tasks they had to solve during the following second part of the eye movement recording session. The second part of the eye movement recording session took approximately 25 minutes. Additionally to the eye movement recordings the mouse was also tracked and recorded but not further investigated in this master thesis. To complete the user study the participants had to solve an object perspective taking test designed by Kozhevnikov & Hegarty (2001) in a revised version by Hegarty & Waller (2004).

3.6.1 Participants

Thirty-seven participants participated in the experiment. They were not paid or rewarded in any way. Participants were recruited by e-mail. All participants had normal or corrected-to-normal vision. All participants attended all tests presented in this master thesis between the 30.09.2013 and the 15.10.2013.

4. Results

In this chapter the results of the user study are presented, starting with the results of the online survey. The results of the two eye movement recording sessions are presented separately after the online survey result presentation. Results of the Santa Barbara Sense-of-Direction (SBSOD) scale and the spatial orientation test are presented at the end of this chapter.

4.1 Online Survey

4.1.1 Background Information

The experiment was conducted with a total of 37 participants of which 11 (29.7%) are male and 26 (70.3%) are female. The majority of the participants are between 21 – 30 years old (23 respectively 62.2%). The second largest age group is with 5 participants (13.5%) the age group between 15 – 20 years. The majority of the participants finished High School (14 respectively 37.8%) or possess a Bachelor degree or equivalent (15 respectively 40.5%). 22 respectively 59.5% of the participants wore glasses or lenses. 15 respectively 40.5% of the participants did not have any visual impairments. 12 participants (32.4%) are related to the field of geography. For further information see the online survey section in the appendix.

4.1.2 Google Map Usage

In the second part of the online survey participants were asked to state their agreement to whether they would use a certain map type (cartographic, satellite and Google Street View) for a set of eight geographic tasks. The geographic tasks were formulated according to a master thesis by Boer (2012). Boer (2012) defined the eight most commonly performed geographic tasks according to Carter (2005). The geographic tasks subjects were asked about are: Self-localisation, Identifying locations, Route planning, Navigation and Wayfinding, Identifying Points of Interest (POI), Communication, Storage of information and Virtual tourism. For each of these geographic tasks subjects had to state their agreement or disagreement on a 7-point Likert-scale (1 = *strongly agree* to 7 = *strongly disagree*) if they would use a certain map type for the defined geographic tasks. In figure 14 the results for the usage of a cartographic map representation are presented.

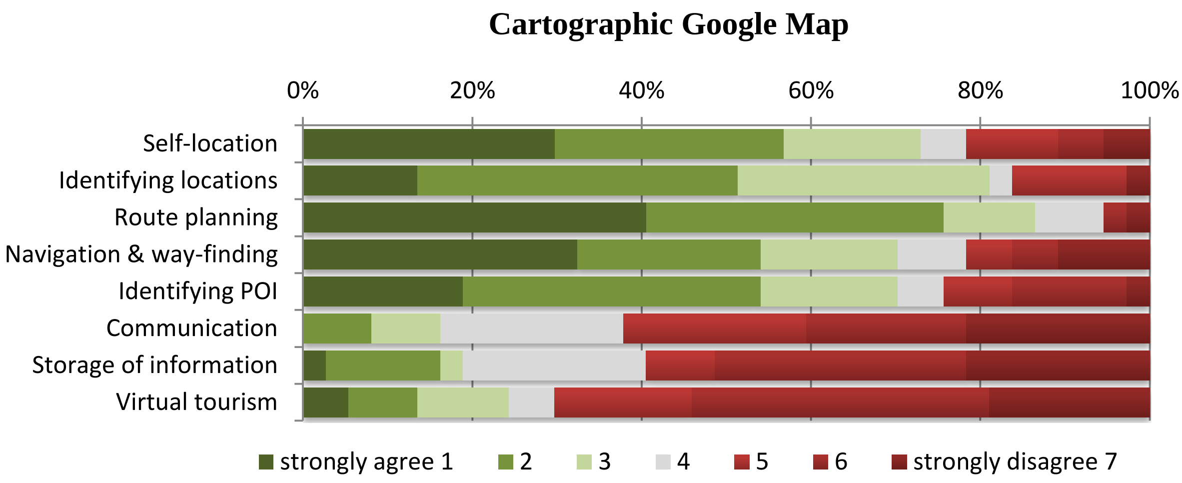


Figure 14: Levels of agreement for eight defined geographic tasks according to Carter (2005) for the cartographic Google Map representation.

As can be seen in figure 14 the majority agrees that they would use the cartographic Google Map representation for self-localisation (>73%), for the identification of certain locations (>81%), for planning a route (>86%), for real-time navigation and wayfinding (>70%) and for the identification of points of interest (>70). The cartographic Google Map representation seems to be less popular for communication (<18%), for information storage (<20%) and for virtual tourism (<25%).

Also for the satellite Google Map representation subjects had to state their agreement or disagreement on a 7-point Likert-scale (1 = *strongly agree* to 7 = *strongly disagree*) if they would use the satellite Google Map representation for the defined geographic tasks. The results of this question are presented in figure 15.

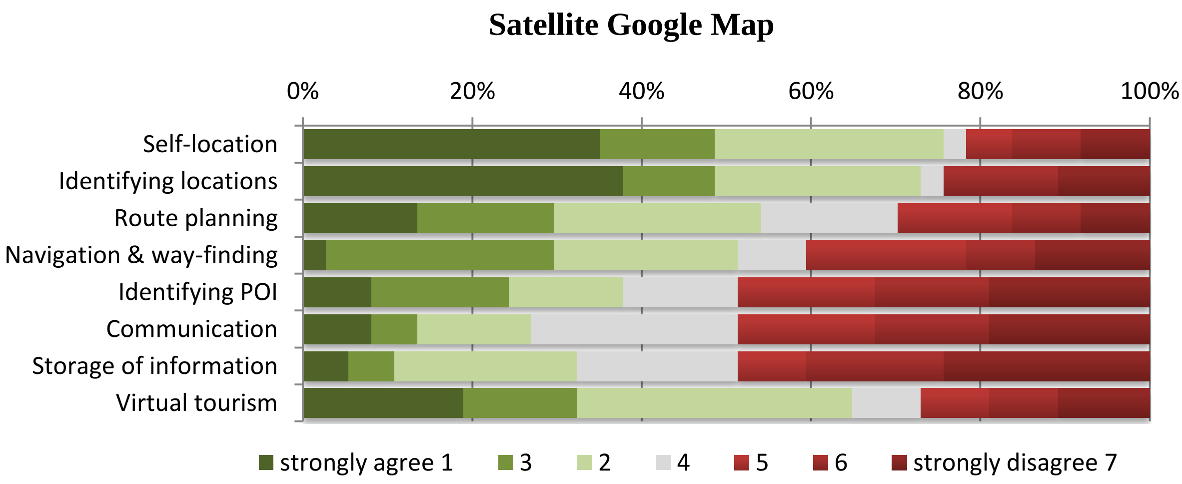


Figure 15: Levels of agreement for eight defined geographic tasks according to Carter (2005) for the satellite Google Map representation.

As can be seen in figure 15 the majority agrees that they would use the satellite Google Map representation for Self-localisation (>75%), for the identification of certain locations (>71%), for planning a route (>53%), for real-time navigation and wayfinding (>50%) and for virtual tourism (>64). The satellite Google Map representation seems to be less popular for the identification of points of interest (<38%), for communication (<27%) and for information storage (<32%).

The last relevant representation for this master thesis is Google Street View. Subjects had to answer the same questions already asked before for the Google Street View representation, too. The results of this question are presented in figure 16.

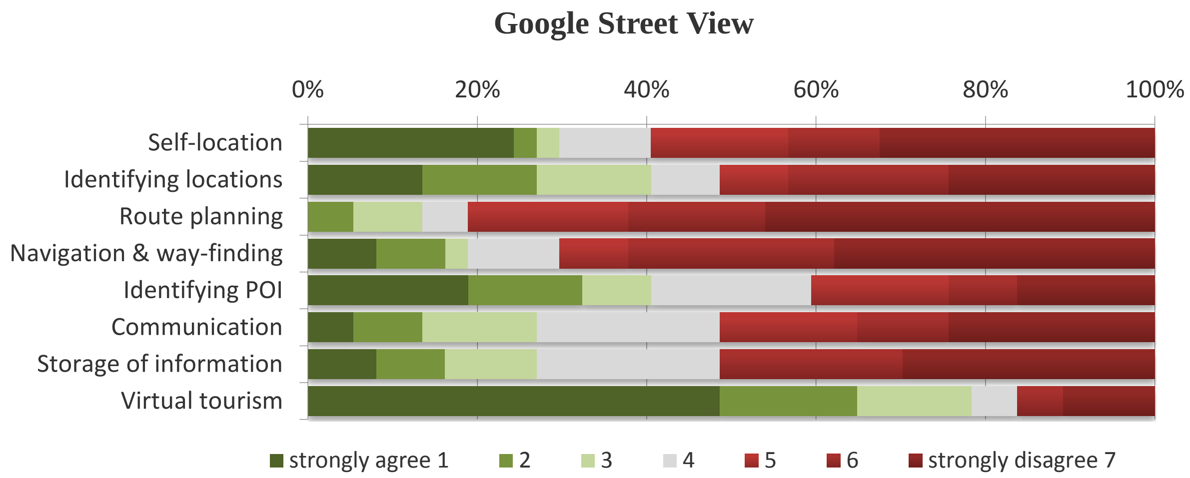


Figure 16: Levels of agreement for eight defined geographic tasks according to Carter (2005) for Google Street View.

As can be seen in figure 16 the only geographical task the majority of participants would perform with Google Street View is virtual tourism (>78%). For all other geographical tasks the participants seem to consider the Google Street View representation as less suitable or useful. Participants were also asked to state on a 7-point Likert-scale how often they use Google Map representations. The results of this question are presented in figure 17:

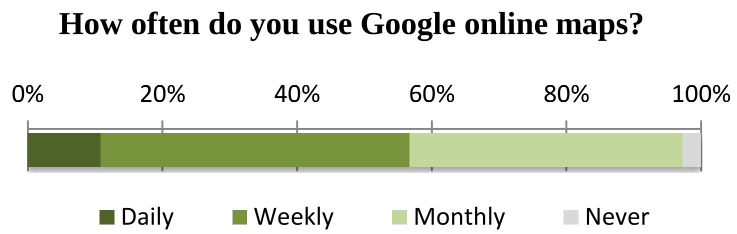


Figure 17: Google Map usage.

As can be seen in figure 17 almost all participants are familiar with the online Google Map service. Only one person never used a Google Map representation. The majority of the participants uses the online Google Map service at least once a week (46%) or once in a month (40%). Participants were also asked on what devices they use the Google Map service or if they use a printed version of the Google Map representations. The answers to this question are presented in figure 18.

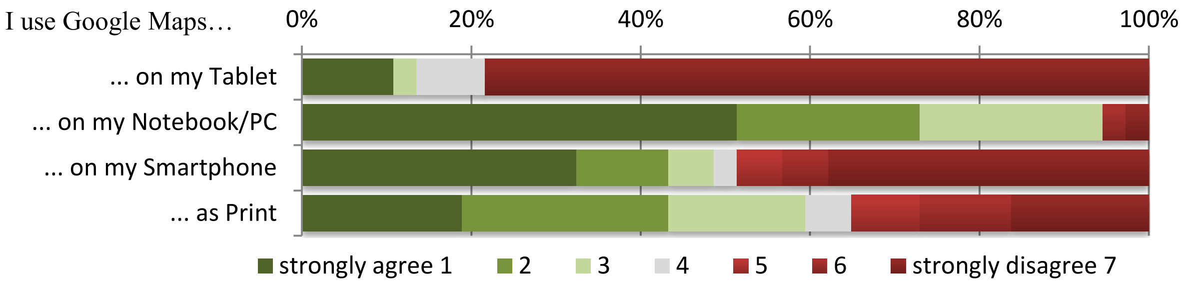


Figure 18: Utilised Devices for Google Map representation usage.

The answers in figure 18 suggest that a majority uses the Google Map services on a personal computer or a notebook (95%). Surprisingly, still about half of all participants use Google Maps also in a printed version. Participants were also asked what map representation they prefer overall. The results of this question are presented in figure 19.

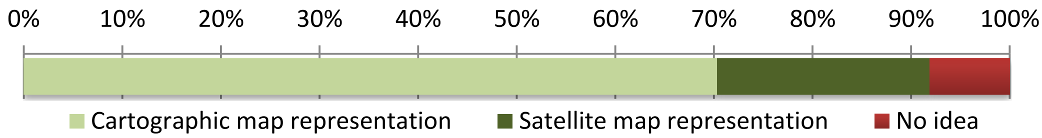


Figure 19: Map representation preference.

Figure 19 shows the representation preferences of the participants. The majority of the participants prefer the cartographic Google Map representation (70%), whereas only 22% of the participants prefer the satellite Google Map representation.

4.1.3 Orientation and Environmental Orientation Clues

Participants were also asked to rate their orientation ability in general. The results of this question are presented in figure 20.

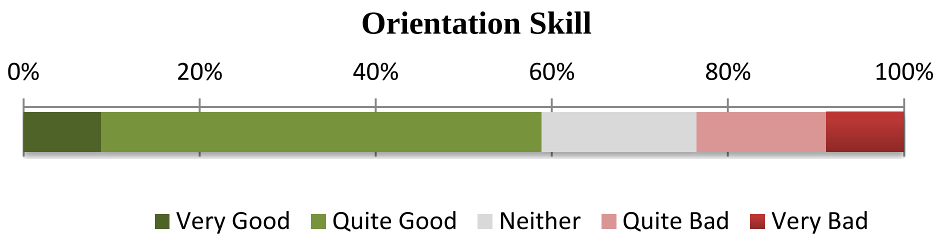


Figure 20: Self-reported orientation abilities.

As we can see in figure 20 most of the participants who filled out the online survey were confident about their orientation abilities in an unfamiliar environment. 50% of the participants stated that they think they have a quit good sense of orientation and about 9% of the participants are confident that they have a very good sense of orientation.

Participants were also asked what kind of environmental clues they usually use for the orientation in an unfamiliar environment. The results of this question are presented in figure 21.

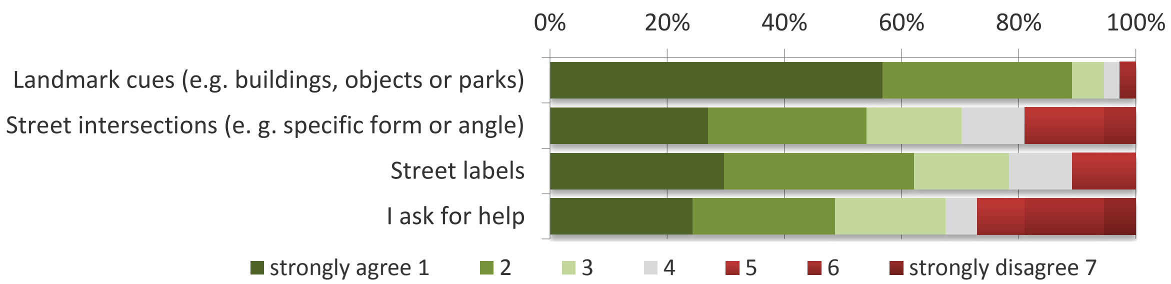


Figure 21: Level of agreement with four classic orientation clues.

As can be clearly seen in figure 21 most of the participants relay on landmark cues such as buildings, objects or parks (>94%) and street labels (>78%). The use of road junctions (>70) as environmental clues or asking other people for help (>67%) is still preferred by a majority of the participants.

4.1.4 Summary

The subject pool is unfortunately not uniform in terms of gender and age. This has to be taken into account while interpreting the data. Nevertheless the influence should not be too severe because no gender or age differences are tested in this master thesis.

The survey questions about the Google Maps usage reveal more or less what was expected. The cartographic, abstract version is mainly used for active usage such as self-localisation or navigation. In this context it has to be taken into account that Google Maps is an online service which means that most people will use the services outdoors on a mobile device with a small display. So the size of the display could have an influence on how subjects answer the question. In contrast to the cartographic map representation the satellite map representation is besides, for identifying locations and for self-localisation, mainly used for virtual tourism. This outcome is somewhat obvious, because the environmental information is much higher in the satellite map representation than in the cartographic map representation. The last visualisation used in this master thesis is the service Google Street View. Google Street View is mainly used for virtual tourism and to a lesser extent for identifying locations. Almost all subjects were highly or quite familiar with Google's online maps. This result is therefore in line with the assumption made in the chapter "Methods" that Google dominates the online map section. The familiarity with the design of Google Maps is important because it should prevent response time differences on the basis of unfamiliarity with the map design. Because the overwhelming majority uses Google Maps on their PC's or notebooks this should also prevent major differences in response time on the basis of unfamiliarity with the desktop representation design.

However, an influence on the outcome of the experiment could have the fact that the majority prefers the cartographic map representation because preferences could lead to higher familiarity with a certain kind of representation. Nevertheless this higher familiarity with a certain kind of representations does not necessarily lead to a higher performance.

Interesting is that the overwhelming majority uses landmark cues while orienting in an unfamiliar environment. This highlights the major importance of landmarks in navigation and confirms the empirical goal of this master thesis to evaluate the perception of landmark cues on a 2D layout map representation.

4.2 2D Layout Map Search Task

4.2.1 Data Preparation

With respect to the experiment design the search time (ST) was calculated by subtracting the time to first fixation (TTFF) of the search reference from the time to first mouse click (TTFMC) on the target.

$$\mathrm{ST} := \mathrm{TTFMC}_{target} - \mathrm{TTFF}_{search\ Reference}$$

To insure the data quality all missing values were controlled by checking the Tobii Studio in the program video recording or the gaze plot of each participant. The video recordings were also checked when the TTFF of the search reference exceeded 1 second. When the TTFF was missing or exceeded 1 second the time was manually determined by looking at the gaze plot to check for calibration shifts or if the TTFF could not be determined by checking the gaze plot it was determined by checking the eye-movement in slow motion in the video recording. Overall 115 TTFF values out of a total of 1036 TTFF values were checked and 74 values respectively 7.14% were corrected according to the described methods above (gaze plot and video recording). 11 TTFMC missing values were checked and none had to be corrected.

4.2.2 Dealing with Outliers

Participant: Before we can start with the data analysis we will have to deal with the outliers in the data set. In a first step we identify outliers in the individual performance of each participant by calculating a z-score for each participant. The z-score is very helpful to determine outliers because we expect that only a certain percentage of our data is greater than some absolute thresholds (positive or negative). According to Field (2005) we expect about 5% of our data to have an absolute value greater then 1.96, and about 1% to have absolute values greater then 2.58, and none values greater than about 3.29 (Field, 2005:76). The formula for the z-score calculation is:

$$z = \frac{X - \bar{X}}{s}$$

Where X represents each score, $\bar{X}$ stands for the mean of all scores and s represents the standard deviation of all scores. The z-score is calculated separately within the main categories cartographic map representation and satellite map representation. The z-score is calculated for the categories "Cartographic Map" and "Satellite Map" separately because possible preferences for one or the other map representation can be taken into account like this. Each of these two data sets includes the following subsets: "Label", "Street" and "Background" as well as the combinations of these subsets (see also chapter 3 Methods).

The results of this calculation are presented in table 2 where the table OUTLIER1 represents the cartographic map representation and the table OUTLIER 2 represents the satellite map representation.

OUTLIER1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Absolute z-score less than 2	482	93.1	94.0	94.0
	Absolute z-score greater than 1.96	19	3.7	3.7	97.7
	Absolute z-score greater than 2.58	11	2.1	2.1	99.8
	Absolute z-score greater than 3.29	1	.2	.2	100.0
	Total	513	99.0	100.0	
Missing	System	5	1.0		
Total		518	100.0		

OUTLIER2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Absolute z-score less than 2	479	92.5	93.7	93.7
	Absolute z-score greater than 1.96	22	4.2	4.3	98.0
	Absolute z-score greater than 2.58	10	1.9	2.0	100.0
	Total	511	98.6	100.0	
Missing	System	7	1.4		
Total		518	100.0		

Table 2: Outliers for cartographic (OUTLIER1) and satellite map representation (OUTLIER2).

As we can see in table 2 only one participant has a significant outlier in the cartographic map representation (OUTLIER1). The value above 3.29 means that in comparison to his/her general performance in this task he/she needed significantly longer to find the target area. The value was therefore deleted and considered as "not found". No significant outliers were found in the satellite map data (OUTLIER2).

Task: In a second step the different tasks (subcategories) are tested for significant outliers. The process is the same as it was for the participants. The data set is also first divided in the two main categories cartographic and satellite map representation. The z-score is then calculated for each subcategory (Label, Street, Background, combinations and Original) within the main categories.

For the cartographic map representation over all subcategories a total of 10 significant outliers were identified. This represents 1.95% of all data points (N=512) in the category cartographic map representation. In detail, two outliers were identified in the subcategories “Background”, “Label-Street” and “Original” and in the remaining subcategories one outlier each was identified. According to Field (2005) the identified significant outliers were assigned the value 3.29 which represents three times the standard deviation added to the mean. By rearranging the z-score formula the z-scores can be converted back to the search time values.

The same procedure was realised with the satellite map representation data set. For the satellite map category 7 significant outliers were found. This represents 1.37% of all data points (N=511) in the category satellite map representation. In terms of subcategories two outliers were found in the subcategory “Label-Street” and one outlier was found in all other subcategories expect the subcategory “Label-Background” in which no significant outlier was found. The z-score values for the satellite map representation were also converted back to the original search time values after the significant outliers were replaced with the value 3.29.

4.2.3 Descriptive Statistics

After all significant outliers had been removed taking a first look at the descriptive statistics is in order. In table 3 the main key descriptive statistic values are summarized for the cartographic map representation.

Statistics

		Label	Street	Background	Label-Street	Label-Background	Street-Background	Original
N	Valid	73	71	74	74	72	74	74
	Missing	1	3	0	0	2	0	0
Mean		14.0891	19.8976	8.4621	11.6663	12.3654	9.4827	14.6049
Median		11.0071	13.5330	6.0115	9.9861	10.6248	7.0941	12.7701
Std. Deviation		9.71976	17.35613	6.47884	7.17784	7.25251	7.72565	9.66237
Skewness		1.417	2.153	1.985	1.961	1.364	2.034	1.888
Std. Error of Skewness		.281	.285	.279	.279	.283	.279	.279
Kurtosis		3.000	4.476	4.084	5.085	2.012	3.990	4.789
Std. Error of Kurtosis		.555	.563	.552	.552	.559	.552	.552

Table 3: Summary of the main descriptive statistics for the search time results in seconds of the cartographic map representation.

As we can see in table 3 the skewness value is $g_1 > 0$ for all subcategories. This means that the distribution of the search time data is likely to be skewed to the right. The kurtosis value is an indicator for normal distribution: $g_2 - 3 = 0$ indicates a normal distribution, $g_2 - 3 > 0$ indicates that the frequency of the maximum values exceeds the normal distribution and $g_2 - 3 < 0$ indicates that the frequency of the maximum values is below a normal distribution with the same variance (Toutenburg & Heumann, 2008). As we can see in table 3 only the subcategory "Label" is probably normal distributed. All other subcategories are either > 0 (Street, Background, Label-Street, Street-Background and Original) or < 0 (Label-Background). In table 4 the key descriptive statistics for the satellite Google Map representation are presented, too.

Statistics

		Label	Street	Background	Label-Street	Label-Background	Street-Background	Original
N	Valid	74	72	74	74	72	72	73
	Missing	0	2	0	0	2	2	1
Mean		14.6333	9.2506	12.1582	18.1394	25.8124	16.8315	13.9922
Median		11.0282	7.4418	8.7349	15.6792	19.2173	14.6626	9.5444
Std. Deviation		10.72819	5.88910	7.75693	12.90538	17.02633	11.15243	12.82387
Skewness		1.627	1.748	1.471	1.774	.994	1.282	2.296
Std. Error of Skewness		.279	.283	.279	.279	.283	.283	.281
Kurtosis		3.528	2.908	2.141	3.121	.253	1.974	5.618
Std. Error of Kurtosis		.552	.559	.552	.552	.559	.559	.555

Table 4: Summary of the main descriptive statistics for the search time results in seconds of the satellite map representation.

As we can see in table 4 the descriptive statistics for the satellite map representation show a comparable picture to those of the cartographic map representation. The skewness values $g_1 > 0$ for the satellite map representation indicate that the search time data is skewed to the right, too. For the kurtosis value the picture is somewhat different, because there are also quite a couple of subcategories with kurtosis values < 0 (Street, Background, Label-Background and Street-Background). All other subcategories have values > 0. So the kurtosis values for the raw search time data are not very promising in terms of normal distribution.

Another interesting question is how well the participants performed in each subcategory and main category. With the help of the medians in table 3 and 4 we can rank the subcategories according to participants' performance for each main category. The median was used because there is for some subcategories quite a substantial difference between the median and mean values as we can see in table 3 and 4. This leads to the suggestion that the influence of the remaining (not significant) outliers is still quite substantial. But even if the means would have been used for the subcategory ranking the result would still be the same. The ranking of the subcategories is shown in table 5.

Rank	Median [sec]	Cartographic Map	Median [sec]	Satellite Map
1	6.01	Background	7.44	Street
2	7.09	Street & Background	8.73	Background
3	9.99	Label & Street	9.54	Original Map
4	10.62	Label & Background	11.03	Label
5	11.01	Label	14.66	Street & Background
6	12.77	Original Map	15.68	Label & Street
7	13.53	Street	19.21	Label & Background

Table 5: 2D layout map search efficiency ranking in [sec].

Table 5 represents the ranking of all subcategories according to the median search time results of all participants in seconds. In figure 22 the response time for the different stimuli and their combinations are presented for the cartographic and the satellite map representation.

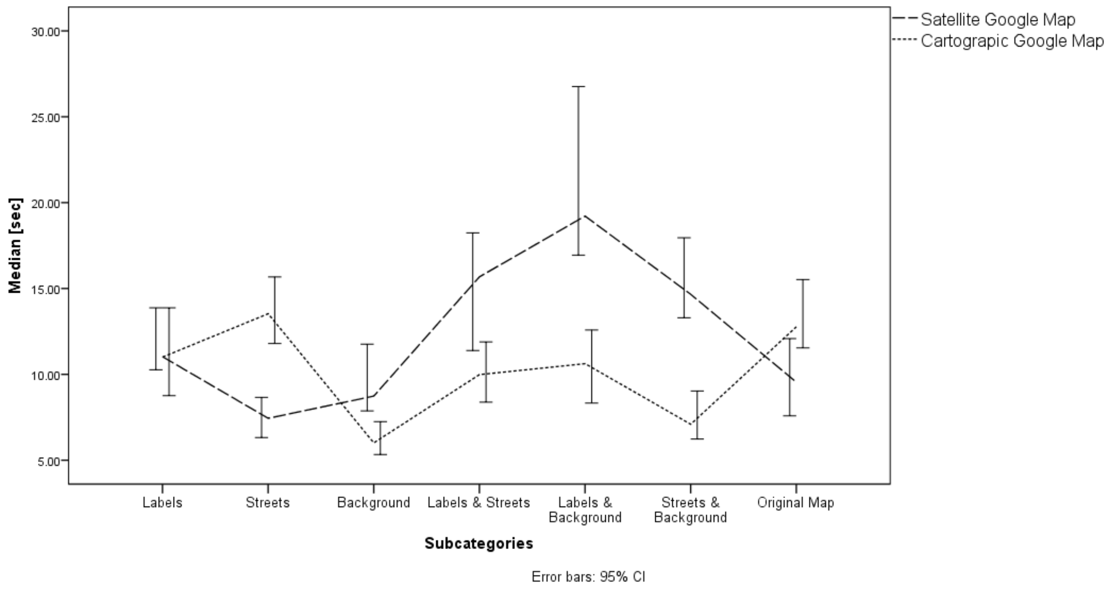


Figure 22: Task response time in [sec].

Figure 22 shows the differences in response time (RT) in seconds between the two map representations used in the experiment for the stimuli street labels (Labels), road junctions (Streets) and landmarks (Background) as well as their combinations. Figure 22 presents the median response time and the 95% confidence interval indicating the variance present in the data. In combination with the data presented in table 5, figure 22 strengthens the impression that the cartographic map representation is except for to exceptions (road junctions and original representation) faster than the satellite map representation.

In figure 23 the three defined stimuli are represented separately. Each stimulus is compared with the combinations it is also part of. This means that if we assume that one stimulus dominates the other two stimuli, we would expect a more or less horizontal line for the response time median results of the combinations. Figure 23 is subdivided in four different graphs representing a) the street label median response times; b) the road junction median response times; c) the landmark median response times and d) all stimuli median response times measured in the stand-alone condition. The stimuli presented in figure 23 stand for the median response times for the satellite Google Map representations.

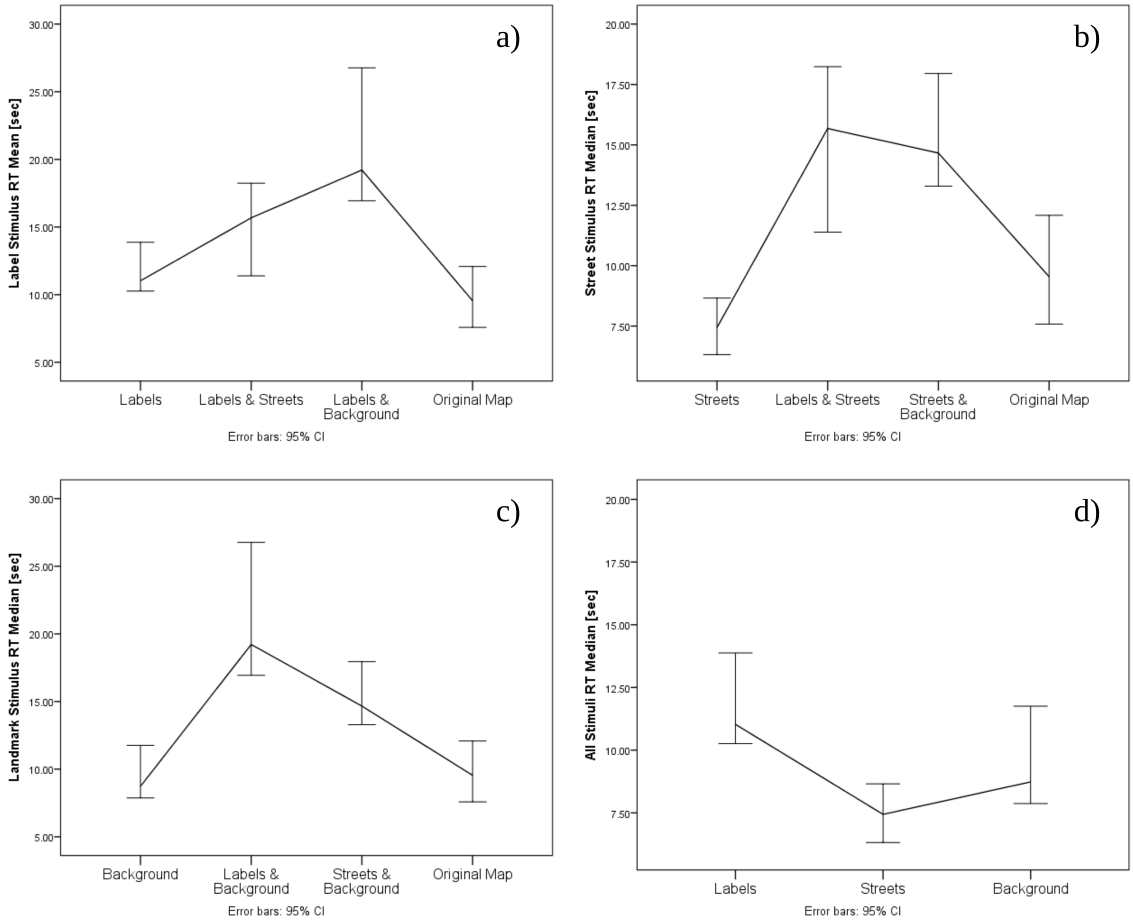


Figure 23: Median RT [sec] for the stimuli in the satellite map representation.

Figure 23 shows the different response times for all stimuli tested. We see that in graph a) representing the street label condition the median of the response time increases for the combinations. Only in comparison to the original satellite Google Map design the stand-alone “Label” condition is slower. In graph b) we see a comparable median response time development only, that this time the stand-alone “Streets” condition is slightly faster. Also in graph c) representing the landmark condition, the development of the response time is comparable to the ones presented in graph a) and b). In general the combinations of only two stimuli seem to be slower than the stand-alone conditions and the original satellite Google Map representation. Noticeable is also the high variance of the “Labels-Background” condition in the graphs a) and c). In graph d) we see that the “Label” condition seems to be slower than the “Streets” or “Background” condition.

In figure 24 the median response times for the cartographic Google Map representations are presented. The data is graphically in the same manner arranged as for figure 23.

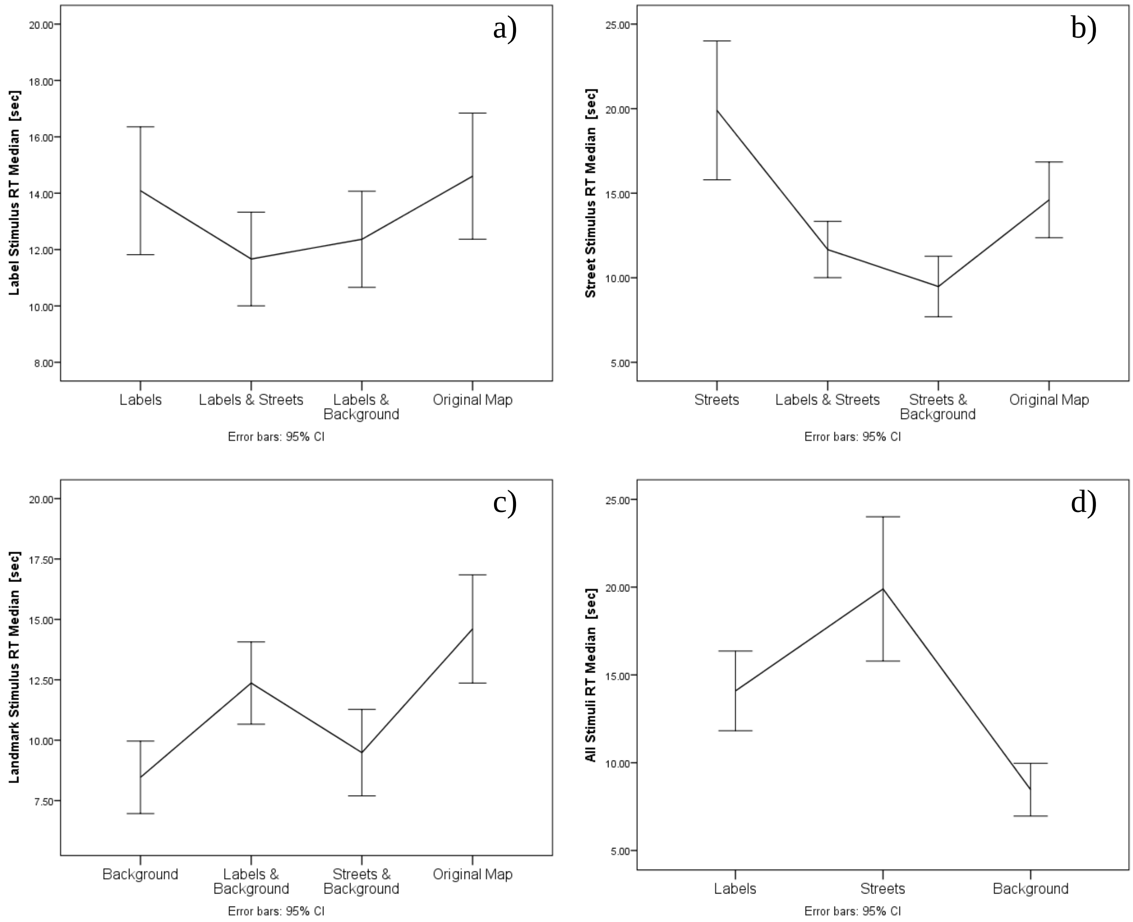


Figure 24: Median RT [sec] for the stimuli in the cartographic map representation.

Figure 24 shows the response times for the cartographic representations of Google Maps. Compared to figure 23, representing the satellite representations of Google Maps, we see for the graphs a) and b) a reversed development of response time. In the graphs a) and b) response time decreases for the combination of only two stimuli and increases for the stand-alone condition and the original Google Map representation. Only in graph c) the stand-alone condition (Background) is faster than the two stimuli condition and the original Google Map representation. Graph d) represents the stand-alone condition of the three stimuli street labels (Labels), road junctions (Streets) and landmarks (Background). Graph d) shows that landmarks are the fastest found compared to the other stimuli stand-alone conditions, whereas it seems the hardest to find road junctions on the cartographic map representation.

4.2.4 Statistical Analysis

In a first step it was tested if the data is normal distributed. Because for all subcategories the assumption of normal distribution was not given the data was Log10 transformed. According to Field (2005) taking the logarithm of a set of numbers helps to get the data values on the right side (large values) of a data set nearer to the centre. It is therefore a good way to reduce positive skew as it is present in our data set (see table 3 and 4).

As we can see in table 6 only 4 out of 14 subcategories are not normal distributed for the Kolmogorov-Smirnov test, respectively 1 for the Shapiro-Wilk test after the Log10 transformation.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Subcategories	Statistic	df	Sig.	Statistic	df	Sig.
Log10 (Cartograpic Google Map Search Time [sec])	Labels	.069	73	.200*	.978	73	.234
	Streets	.117	71	.017	.969	71	.075
	Background	.103	74	.049	.956	74	.011
	Labels & Streets	.068	74	.200*	.990	74	.851
	Labels & Background	.074	72	.200*	.987	72	.698
	Streets & Background	.084	74	.200*	.970	74	.080
	Original Map	.095	74	.097	.982	74	.354
Log10 (Satellite Google Map Search Time [sec])	Labels	.063	74	.200*	.981	74	.330
	Streets	.069	73	.200*	.976	73	.171
	Background	.118	74	.013	.980	74	.299
	Labels & Streets	.082	74	.200*	.976	74	.167
	Labels & Background	.075	72	.200*	.979	72	.262
	Streets & Background	.123	72	.008	.974	72	.149
	Original Map	.083	73	.200*	.979	73	.280

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 6: Test of normality of the Log10 transformed search times for the cartographic and satellite Google Map representation.

Even though not all subcategories are significantly normal distributed, an ANOVA was conducted to determine possible significant differences between the subcategories tested. It was assumed that the requirement for conducting an ANOVA is given, because in a visual inspection of the data the distribution was considered to be close enough to a normal distribution. Also the assumption of equal variances (Homogeneity of Variances) is with a Levene Statistic p-value = 0.082 > 0.05 for the satellite map representation and with a p-value = 3.38 > 0.05 for the cartographic map representation fulfilled. In table 7 the results of the ANOVA are presented.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Log10 (Satellite Google Map Search Time [sec])	Between Groups	7.929	6	1.321	16.017	.000
	Within Groups	41.663	505	.083		
	Total	49.591	511			
Log10 (Cartograpic Google Map Search Time [sec])	Between Groups	6.320	6	1.053	13.714	.000
	Within Groups	38.784	505	.077		
	Total	45.104	511			

Table 7: ANOVA of the RTs for the cartographic and satellite Google Map representation.

As we can see in table 7 there are significant differences between the groups for the satellite Google Map representation (F [6, 511] = 16.017, p = .000) as well as for the cartographic Google Map representation (F [6, 511] = 13.714, p = .000) at a significance level of $\propto = 0.05$. Because the null hypothesis was rejected, additionally a Bonferroni *post hoc* test was conducted to take a closer look at the differences between the groups. In table 8 we see the results of the Bonferroni test for the satellite map representation.

Satellite Map

	Label [Sig.]	Street [Sig.]	Background [Sig.]	Label-Street [Sig.]	Label-Background [Sig.]	Street-Background [Sig.]	Original Map [Sig.]
Street Labels		0.016	1.000	0.403	0.000		1.000
Road Junctions	0.016		0.418	0.000		0.000	0.263
Landmarks	1.000	0.418			0.000	0.000	1.000

Table 8: Significance levels of the stimuli in the satellite map.

In table 8 we see the significance levels of the tested stimuli “Label”, “Street” and “Background” and their combinations. The stimuli “Label” refers to street labels, the stimuli “Street” refers to road junctions, and the stimuli “Background” refers to the layer in a map representation where landmarks are considered to be present. As we can see in table 8 the response times (RT) measured for the street label condition do not differ significantly expect for the “Label-Background” condition. In comparison with figure 23 a) we see that the condition “Label-Background” is significantly slower than the other street label conditions. For the different road junction conditions we see in table 8 and figure 23 b) that the RT for the stand-alone “Street” condition is significantly faster than the “Label-Street” and “Street-Background” condition. In the landmark condition we see in table 8 and figure 23 c) a significant slower performance of the “Label-Background” condition compared to the other landmark conditions. Interesting is that neither between the stand-alone conditions “Labels” and “Background” nor between all stand-alone conditions and the original satellite Google Map a significant difference in RT can be observed. Only between the stand-alone condition

“Labels” and “Streets” a significant difference can be observed. In figure 23 c) we see that the stand-alone condition “Label” is significantly slower than the stand-alone condition “Street”.

Table 9 represents the results of the Bonferroni test for the cartographic Google Map representation.

Cartographic Map

	Label [Sig.]	Street [Sig.]	Background [Sig.]	Label-Street [Sig.]	Label-Background [Sig.]	Street-Background [Sig.]	Original Map [Sig.]
Street Labels		0.089	0.000	1.000	1.000		1.000
Road Junctions	0.089		0.000	0.002		0.000	0.591
Landmarks	0.000	0.000			0.001	1.000	0.000

Table 9: Significance levels of the stimuli in the cartographic map.

In table 9 the significance levels of the stimuli street labels (“Labels”), road junctions (“Streets”) and landmarks (“Background”) are presented. The significance levels of table 9 can be compared with the graphs presented in figure 24. In figure 24 a) the different conditions of the street label conditions are represented. In table 9 we see that the RT’s do not significantly differ between the “Label” stand-alone condition and the “Label-Street” and “Label-Background” conditions. For the road junction condition in figure 24 b) significant differences in RT between the “Street” stand-alone condition and the combined “Label-Street” and “Street-Background” conditions can be determined. According to figure 24 b) and table 9 participants needed significantly longer to identify the target in the “Street” stand-alone condition than in the combined conditions. For the landmark condition table 9 and figure 24 c) present a different picture. A significant difference in RT is found between the “Background” stand-alone condition and the “Label-Background” condition. It took participants significantly longer to find the target in the “Label-Background” condition than in the “Background” stand-alone condition. For the stand-alone conditions and the original cartographic Google Maps representation only a significant difference in RT could be found between the “Background” condition and the original map. Participants found the target significantly faster in the “Background” condition than in the original Google Maps representation. The “Background” stand-alone condition is also significantly faster than the “Label” and the “Street” stand-alone conditions.

So in summery we can conclude that for the satellite map representations especially the combination of two stimuli slow down the RT, in particular if one of the two stimuli is the “Label” stimulus. Label search in the satellite map representation seems to be particularly difficult. The search for road junctions on the contrary seem to be at least for the stand-alone condition the most efficient even there was no significant difference in RT between the stand-alone conditions “Background” and “Street”. For the cartographic map representation almost a reversed development of RT can be observed. The combinations of two stimuli seem to increase the search efficiency at least marginal. Only for the landmark (Background) condition an increase of RT can be observed when combining two stimuli. For the cartographic representation the stand-alone “Background” stimulus can be found significantly more efficiently than the other two stand-alone stimuli “Label” and “Street”.

4.2.5 Control Variables

The data of the control variables were processed in the same way as described in section 4.2.4. before. In a first step the different response times measured for the control variables in the search tasks were ranked according to their median search time results of all participants and are represented in table 10.

Rank	RT Median [sec]	Cartographic Map	RT Median [sec]	Satellite Map
1	7.7	Background (Original Map)	7.6	Background (Original Map)
2	9.77	Background (Street-Background Map)	9.23	Background (Street-Background Map)
3	10.73	Street (Street-Background Map)	12.25	Street (Original Map)
4	14.11	Street (Original Map)	14.06	Street (Street-Background Map)
5	16.8	Text (Original)	20.65	Text (Original)

Table 10: Response time (RT) ranking of the control variables in [sec].

Table 10 represents the ranking of all subcategories of the control variables according to the median search time results of all participants in seconds. The subcategories of the control variables are to some extend differently defined than the stimuli presented in the sections before. The different definition reflects the attempt to make the results of the stimuli previously analysed more robust. To reread the definition of the control variables see subsection 3.5.1.

Figure 25 presents the median response times (RT) for the control variable displays in a graphical way.

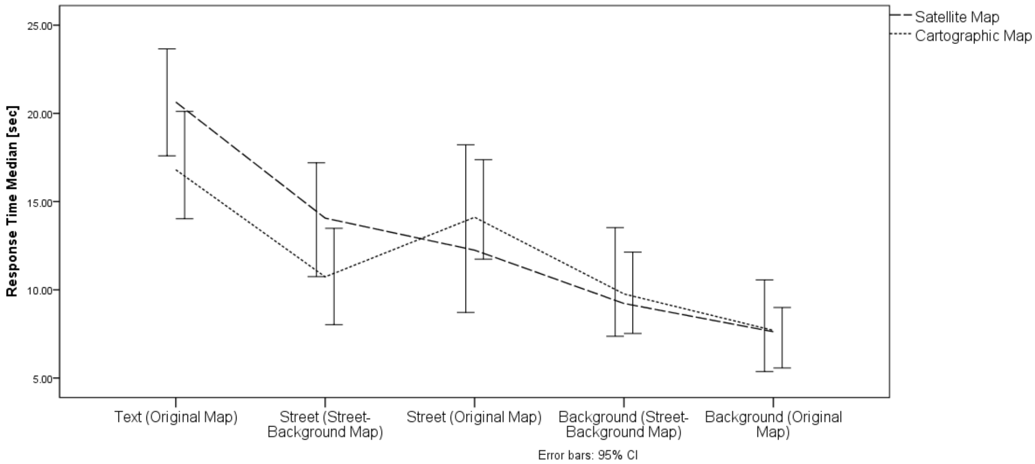


Figure 25: Response time (RT) of the control variables in [sec].

Figure 25 compares the RTs for the different subcategories of the control variables in comparison to the two used map representations. We see a decrease of RT from the "Text" to the "Background" condition for the satellite map representation. A comparable trend can also be observed for the cartographic map representation except for the "Street" (Original Map) condition which shows an increase in RT. In table 11 we see the results of the conducted ANOVA for the control variables to determine possible significant differences between the subcategories tested.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Log10 Search Time: Cartographic Map (Control)	Between Groups	5.129	4	1.282	12.177	.000
	Within Groups	36.961	351	.105		
	Total	42.090	355			
Log10 Search Time: Satellite Map (Control)	Between Groups	6.638	4	1.659	16.656	.000
	Within Groups	35.666	358	.100		
	Total	42.304	362			

Table 11: ANOVA of the control stimuli design.

As we can see in table 11 there are significant differences between the groups for the satellite Google Map representation (F [4, 355] = 12.177, p = .000) as well as for the cartographic Google Map representation (F [4, 362] = 16.656, p = .000) at a significance level of $\propto = 0.05$. Because the null hypothesis was rejected, additionally a Bonferroni *post hoc* test was conducted to take a closer look at the differences between the groups. In table 12 we see the results of the Bonferroni test for the satellite map representation.

Satellite Map

	Text [Sig.]	Street (Street-Background Map) [Sig.]	Street (Original Map) [Sig.]	Background (Street-Background Map) [Sig.]	Background (Original Map) [Sig]
Street Labels (Original Map)		0.260	0.200	0.000	0.000
Road Junctions (Street-Background Map)	0.260		1.000	0.590	0.000
Road Junctions (Original Map)	0.020	1.000		0.084	0.000
Landmark (Street-Background Map)	0.000	0.059	0.084		0.846
Landmark (Original Map)	0.000	0.000	0.000	0.846	

Table 12: Significance levels of the control stimuli in the satellite map representation.

In table 12 we see the significance levels of the defined stimuli street labels, road junctions and landmarks. No significant differences in RT at a significant level of $\propto = 0.05$ can be observed for the road junction and the landmark condition between the "Street-Background Map" and the "Original Map" representations. This strengthens the assumption that the defined stimuli are really responsible for the differences in RT and that these differences are not only based on chance. In figure 25 and table 12 we see that the label search is the most challenging search whereas the landmark search is the most efficient in the satellite map representation. These findings are in line with the previous findings although no significant differences could be found before. In table 13 the results of the Bonferroni *post hoc* test are presented for the cartographic map representation.

Cartographic Map

	Text [Sig.]	Street(Street-Background Map) [Sig.]	Street (Original Map) [Sig.]	Background (Street-Background Map) [Sig.]	Background (Original Map) [Sig]
Street Labels (Original Map)		0.005	1.000	0.000	0.000
Road Junctions (Street-Background Map)	0.005		0.087	1.000	0.139
Road Junctions (Original Map)	1.000	0.087		0.015	0.000
Landmark (Street-Background Map)	0.000	1.000	0.015		0.444
Landmark (Original Map)	0.000	0.139	0.000	0.444	

Table 13: Significance levels of the control stimuli in the cartographic map representation.

As we can see in table 13 no significant differences in RT at a significant level of $\propto = 0.05$ can be observed for the road junction and the landmark condition between the "Street-Background Map" and the "Original Map" representations. This is in line with the findings in table 12 representing the results for the satellite map representation. Also for the cartographic map representation there is a significant difference between the RT for the street label search in comparison with the road junction (only Street-Background Map) search and the landmark search.

Besides the search in a target map containing more information than the search reference, it was also tested what effect the texture of the satellite photograph has on the RT. For this reason a cartographic, abstract designed search reference was provided which had to be used to search for a target item in a satellite map representation. This means that the search reference did not contain any texture clues to help the participants with their search task. The ranking of the median RT in seconds for the search tasks without texture clues is presented in table 14.

Rank	Median RT [sec]	Target (Satellite Map)	Reference (Cartographic Map)
1	13.80	Original	Original
2	15.87	Street-Background	Street-Background
3	26.30	Background	Background

Table 14: Median RT in [sec] for the search tasks without texture clues.

The result for the original satellite map in table 14 leads to the suggestion, that the effect of texture is marginal if we compare the median RT with the median RT in the map search efficiency ranking in table 5. Whereas the median RT for a visual search without any textural clues is 13.8 seconds, the same visual search with textural clues is with 9.54 seconds only 4.26 seconds faster. A greater difference can be observed for the “Street-Background” and especially for the “Background” search. The relations between the median RT for the search tasks without texture clues in a satellite map are also represented in figure 26.

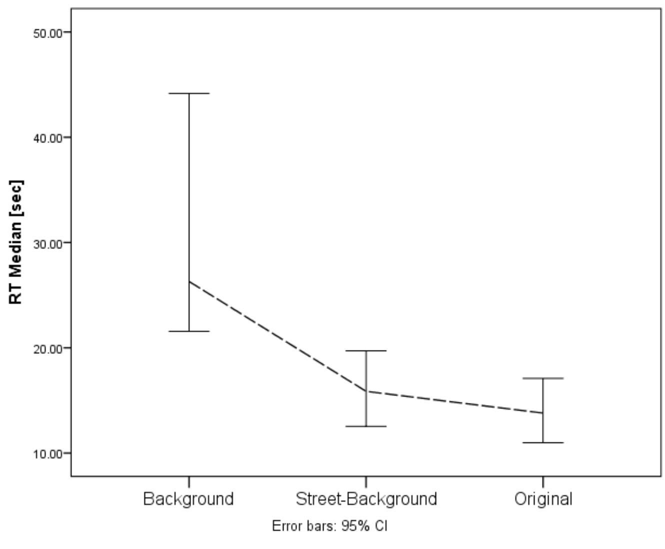


Figure 26: Response time [sec] for the search tasks without textural clues in a satellite map.

The RT for a visual search in a satellite map representation without any textural information is represented in figure 26. As we can see it seems specifically difficult to find a landmark in a satellite map representation without the textural information provided in the satellite map. It seems to be easier to find road junctions and targets if all stimuli are combined (Original). In table 15 we see the amount of participants who either gave up searching for the target or clicked on the wrong target.

	Cases							
	Valid			Missing		Total		
	N	Percent		N	Percent	N	Percent	
Background	49	66.2%		25	33.8%	74	100.0%	
Street-Background	63	85.1%		11	14.9%	74	100.0%	
Original Map	73	98.6%		1	1.4%	74	100.0%	

Table 15: Amount of faults in the visual search without textural clues.

The same picture as in figure 26 emerges by looking at table 15. With 25 (33.8%) faults in the "Background" condition it is clear that for the landmark search in a 2D layout map the textural clues are of major importance. In contrast to the RT results for the landmark search no higher error rate is measured for the search in the original map representation containing all the defined stimuli street labels, road junctions and landmarks.

4.2.6 Summary

The results presented for the 2D layout map visual search task reveal two main findings. The first is that the tested stimuli in the satellite 2D layout map representation seem to be more efficient if presented on their own, whereas for the cartographic 2D layout map we see to a certain degree a revers trend. One explanation for this trend in the satellite map representation could be that less distracting context decreases the RT. This is definitely for the "Street" condition the case in which the background information is missing. In contrary to this for the cartographic map representation additional information seems to be more helpful then distracting. The RT results in the cartographic representation also reveal a strong influence of the "Background" stimulus. The "Background" or landmark containing layer seems to decease the RT significantly especially in combination with the stimulus "Streets" referring to the road junctions present in the map representation. The second main finding is that the search for street labels in satellite maps increases the RT. The effect is significant for the combination of the conditions "Labels" and "Background". For the cartographic representation the influence of the stimulus "Label" is inconclusive. Because no significant differences in RT between the "Label" stand-alone condition, the "Label-Street" condition, the "Label-Background" condition and the original map was found it seems plausible that street labels are the most important feature in a cartographic map search. The reason for this conclusion is the assumption that if one feature dominates all other features the RT will be continuous for all conditions because the other features are not taken into account when searching for a target item.

The control stimuli confirm the findings in the satellite map representation that street labels are the most difficult feature to find in satellite map representations. The RT is even higher than in the “Label-Background” condition suggesting that the landmark information decreased the RT or the representation of the label and their orientation helped the subjects to find the target item. The “Background” conditions in the control stimuli display (7.6 and 9.23 seconds) were comparable to the RT in the regular “Background condition” (8.73 seconds). The RT in the control “Street” conditions were however twice as high than in the regular “Street” condition (12.25 and 14.06 to 7.44 seconds). These results lead to the conclusion that road junctions would be efficient if they were more distinguishable in the satellite map representation. For the control “Street” condition in the cartographic representation the RT (10.73 and 14.11 seconds) where comparable to the regular “Street” condition (13.53 seconds).

The absence of textural clues increases especially the RT for the “Background” stand-alone condition. The median RT is with 26.3 seconds three times the RT in the regular “Background” condition (8.73 seconds). The effect of the texture decreases with the increase of more information in the map. The “Street-Background” condition is with 15.87 seconds comparable to the regular “Street-Background” condition (14.66 seconds). Also the difference in RT between the original map in the “no-textural-clue” condition and the regular satellite map condition is with 13.8 to 9.54 seconds more or less in the same range.

4.1 3D Google Street View to 2D Layout Map Search Task

4.1.1 Data Preparation

In contrast to the 2D layout map search task in the “picture to 2D layout” search task no search time was calculated, because participants did not need to look at a search reference first. They could either start with the environmental picture or with the map representation. The measurement of response time (RT) begins with the start of the picture presentation and ends with the first mouse click on the map representation. This means that the time to first mouse click (TTFMC) equals the RT of the participants. After the participants located the location of the camera in the map representation they had to determine the orientation of the picture in comparison to the north of the map representation. For more information see chapter 3.5.2.

4.1.2 Descriptive Statistics

In table 16 the different stimuli tested are ranked according to their RT in the satellite map representation and the cartographic map representation.

Rank	Median RT [sec]	Satellite Map	Median RT [sec]	Cartographic Map
1	50.12	Road Junctions	39.00	Street Labels
2	61.42	Street Labels	60.28	Landmarks
3	65.66	Landmarks	62.87	Road Junctions

Table 16: Median RT [sec] for the "picture to 2D layout map" search task.

In table 16 we see that for the satellite map representation road junctions are faster found than street labels or landmarks. For the cartographic map representation we see that street labels are the most efficient whereas road junctions are the least efficient feature. The results of table 16 are graphically presented in figure 27.

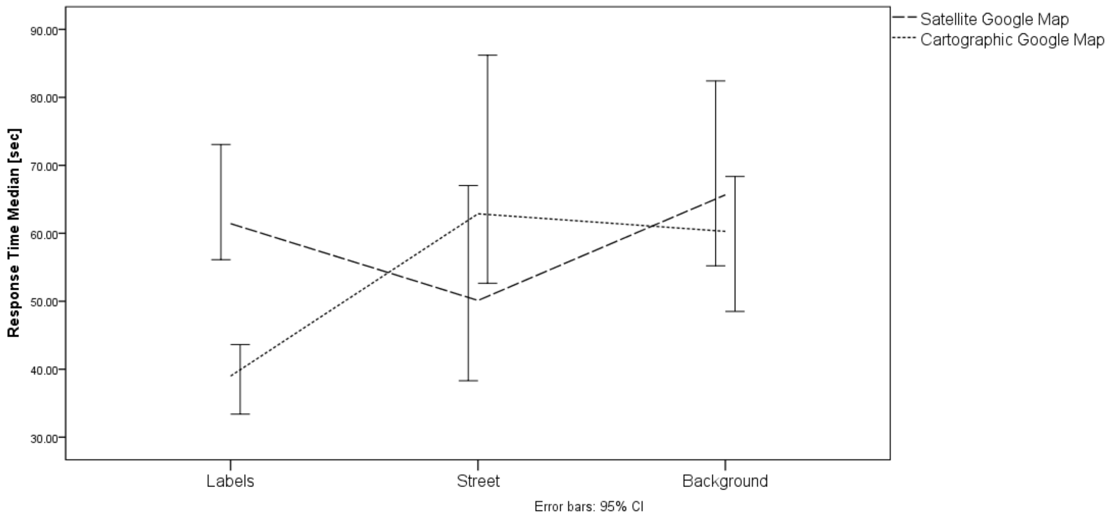


Figure 27: Median RT [sec] for ftimuli of the "picture to 2D layout map" search task.

The only clearly distinguishable difference is the difference between the satellite street label condition and the cartographic street label condition. In general the variance is very high making it difficult to identify any differences within the cartographic or the satellite map representation. Nevertheless the street labels in the cartographic representation seem to be the most efficient feature to determine the camera's location in the map representation. A very interesting trend is represented in figure 28. Figure 28 shows the error rate in the different stimuli condition for the two different map representations.

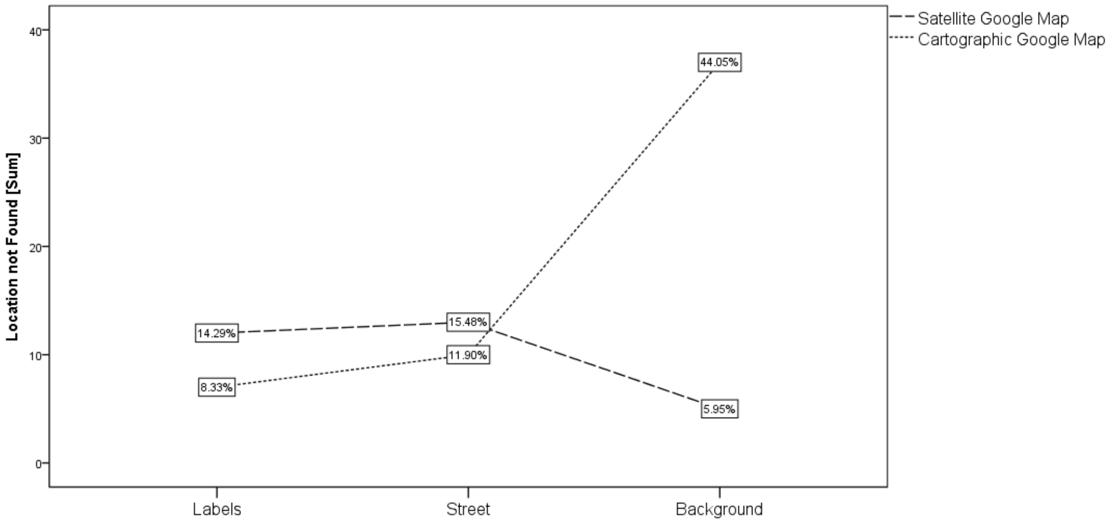


Figure 28: Error rate in the "picture to 2D layout map" search task.

As we can see in figure 28 the error rates are almost identical for the "Label" and the "Street" condition for both map representations. However for the "Background" condition a quite substantial difference in the error rate can be observed. Also we can see in figure 28 that the error rate is lowest with 5.95% for the "Background" condition for the satellite map representations. This observation leads to the suggestion that the self-localisation with the help of landmarks in satellite map representations is not particularly fast but highly accurate. Whereas for the cartographic map representation the street labels are efficient in terms of RT as well as under the aspect of the error rate.

4.1.3 Statistical Analysis

Because the raw data of the "picture to 2D layout map" experiment are not normal distributed they were Log10 transformed. As we can see in table 17 the Log10 transformed RT values are then normal distributed.

Tests of Normality

	Sub	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Log10 (Satellite Google Map Search Time [sec])	Street Labels	.089	62	.200*	.977	62	.294
	Road Junction	.126	61	.018	.981	61	.463
	Landmark	.063	69	.200*	.992	69	.935
Log10 (Cartograpic Google Map Search Time [sec])	Street Labels	.058	67	.200*	.986	67	.643
	Road Junction	.104	51	.200*	.975	51	.341
	Landmark	.097	37	.200*	.978	37	.674

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 17: Tests of Normality of the "picture to 2D layout" RT results.

Because the data is normal distributed (table 17) we can proceed with an ANOVA to test for differences between the subcategories for the cartographic and satellite Google Map representations. The results of the ANOVA are presented in table 18.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Log10 (Cartograpic Google Map Search Time [sec])	Between Groups	1.705	2	.853	11.962	.000
	Within Groups	10.834	152	.071		
	Total	12.539	154			
Log10 (Satellite Google Map Search Time [sec])	Between Groups	.436	2	.218	3.306	.039
	Within Groups	12.460	189	.066		
	Total	12.896	191			

Table 18: ANOVA of the "picture to 2D layout map" RT results.

As we can see in table 18 there are significant differences between the groups for the cartographic Google Map (F [2, 154] = 11.962, p = .000) as well as for the satellite Google Map (F [2, 191] = 3.306, p = .039) at a significance level of $\propto$ = 0.05. Because the null hypothesis was rejected a Bonferroni *post hoc* test was conducted to take a closer look at the differences. In table 19 we see the results of the Bonferroni test for the satellite and cartographic map representation.

Multiple Comparisons

Bonferroni

Dependent Variable	(I) Sub	(J) Sub	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Log10 (Cartograpic Google Map Search Time [sec])	Street Labels	Road Junction	-.22203*	.04961	.000	-.3421	-.1019
		Landmark	-.19526*	.05468	.001	-.3276	-.0629
	Road Junction	Street Labels	.22203*	.04961	.000	.1019	.3421
		Landmark	.02677	.05765	1.000	-.1128	.1663
	Landmark	Street Labels	.19526*	.05468	.001	.0629	.3276
		Road Junction	-.02677	.05765	1.000	-.1663	.1128
Log10 (Satellite Google Map Search Time [sec])	Street Labels	Road Junction	.09842	.04630	.105	-.0134	.2103
		Landmark	-.00707	.04493	1.000	-.1156	.1015
	Road Junction	Street Labels	-.09842	.04630	.105	-.2103	.0134
		Landmark	-.10549	.04512	.061	-.2145	.0035
	Landmark	Street Labels	.00707	.04493	1.000	-.1015	.1156
		Road Junction	.10549	.04512	.061	-.0035	.2145

*. The mean difference is significant at the 0.05 level.

Table 19: Bonferroni post hoc test of the "picture to 2D layout map" RT results.

In table 19 we see that the only significant differences in RT are the differences between the street label condition and the road junction as well as the landmark condition in the cartographic map representation. For the satellite map representation no significant differences in RT could be found between the subcategories.

4.1.4 Summary

The RT results for the "picture to 2D layout map" search task show that for the satellite map representation road junctions are the most efficient environmental feature for self-localisation, whereas for the cartographic map representation street labels are the most efficient environmental feature. However only between the street label condition and the road junction and landmark condition in the cartographic map representation a significant difference in RT was measured. When comparing the error rates for each stimulus in the two different map representations a different picture arises. In figure 28 we see that landmark cues are the most accurate environmental features for self-localisation in a satellite map representation. In contrast to the satellite map representation it was shown that in the cartographic map representation street labels are the most accurate environmental features. This means that landmarks work well as a help in a self-localisation task but the best performance is still achieved when using a cartographic map representation looking for street labels in an unfamiliar environment.

4.2 Orientation

During the “picture to 2D layout map” search task participants had also to determine the orientation of the picture in relation to the map representation. For more details see also section 3.5.2.

In figure 29 each symbol stands for a picture and its orientation in relation to the map representation. The stimuli presented in the pictures are shown in the legend of figure 29. On the Y axis we see the error rate which accrued for each cardinal direction and stimulus.

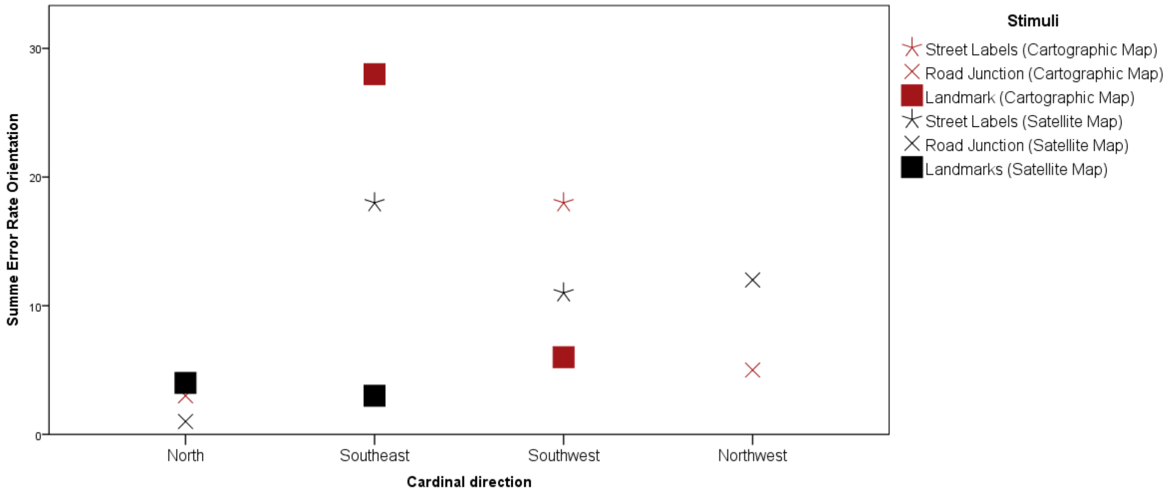


Figure 29: Error rate for different cardinal directions.

We see in figure 29 that the least errors were made when the picture was oriented to the North. This was expected because no mental rotation of the scene was necessary to solve the task. The error rate is particularly high for cartographic landmark pictures with a Southeast cardinal direction. A reason for this could be that participants selected a wrong location and therefore also answered the orientation question wrong. To check this assumption the error rate of the camera positioning task with a satellite map and the orientation task error rate is compared in figure 30.

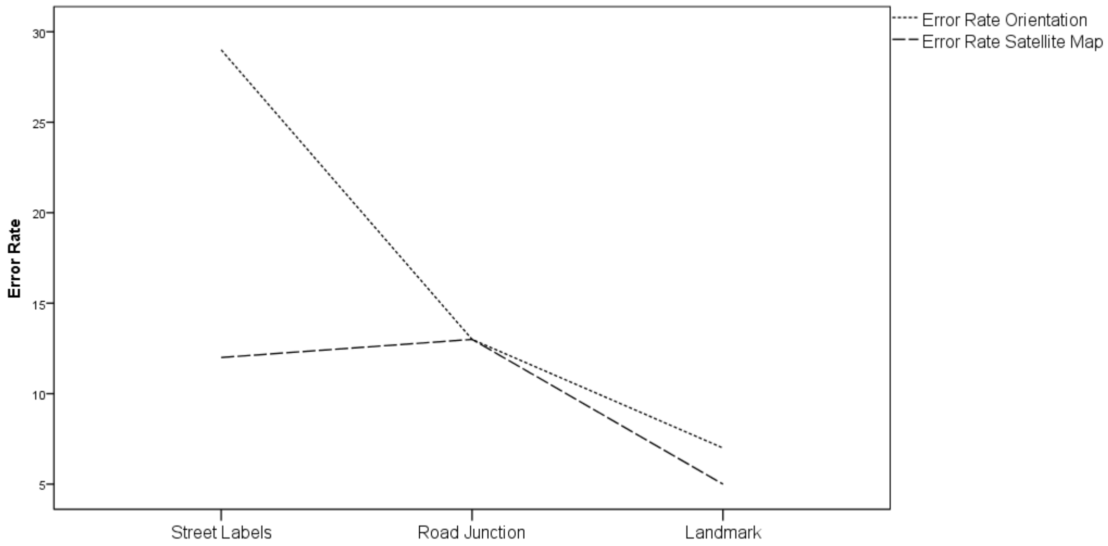


Figure 30: Comparison error rate location task and error rate orientation task.

As we can see in figure 30 the location task error rate correlates with the orientation task error rate except for the street label condition. The great difference in error rates for the street label condition is surprising because for the satellite map the location error rate is comparable between all subcategories although they are differently orientated (SE and SW). The most plausible assumption is that the need to rotate the picture about 180° degrees increases the orientation error rate. Nevertheless the influence of the orientation did not have a great impact on the landmark condition even though the landmark condition was also in one case Southeast orientated. The road junction condition was orientated North and Northwest so maybe the partly faster RT is due to the easier rotation task. In figure 31 we see the same graphic for the cartographic map representation.

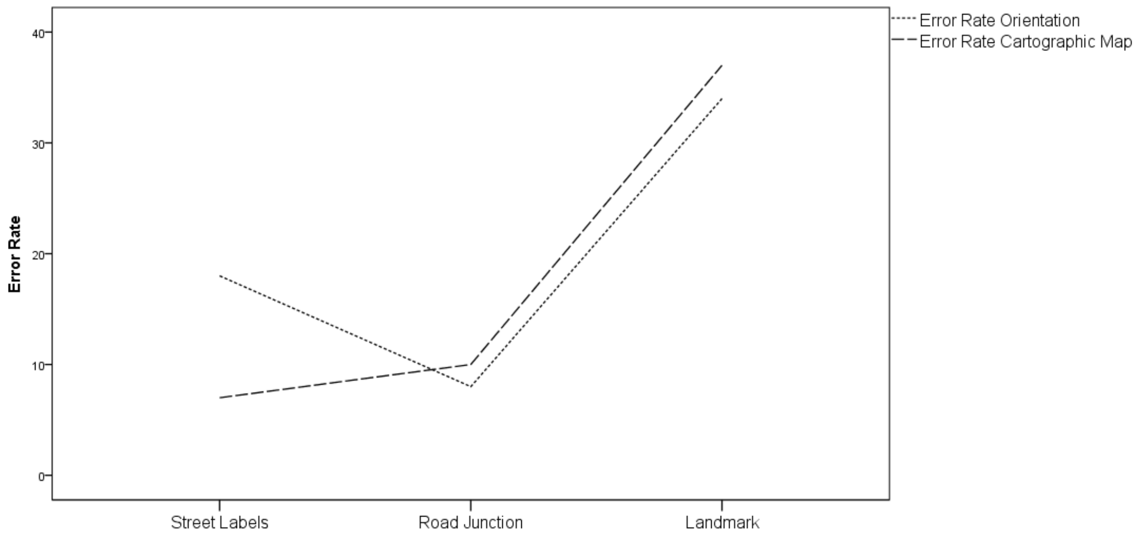


Figure 31: Comparison error rate location task and error rate orientation task.

In figure 31 only in the street label condition a difference between the two error rates can be observed. Both street label condition pictures were orientated to the Southeast. So the harder rotation task seems to have a strong impact on the orientation error rate and may also have an impact on the RT even though participants were precisely instructed to only search for the location first and then determine the orientation. Nevertheless the influence should not be too severe because, as we can see in the cartographic road junction condition, even though the rotation task was easy participants performed slowest in this condition. In contrast to the road junction condition in the cartographic map representation the road junction condition in the satellite map representation was the fastest with exactly the same orientation as the cartographic road junction condition. To determine the exact effect of the orientation of the picture it would be necessary to test all orientations systematically. In the course of this master thesis this was not possible also because of the experiment length.

4.2.1 Summary

The orientation of the picture in relation to the map representation may have an influence on the RT but at least does not dominate the RT results. This can also be seen in table 20 presenting a Spearman correlation between cardinal direction and overall RT (both map representations and all stimuli).

Correlations

			RT all Stimuli (both map representations)	Cardinal direction
Spearman's rho	RT all Stimuli (both map representations)	Correlation Coefficient	1.000	.070
		Sig. (2-tailed)	.	.209
		N	321	321
	Cardinal direction	Correlation Coefficient	.070	1.000
		Sig. (2-tailed)	.209	.
		N	321	444

Table 20: Correlation test between RT (all stimuli of both map representations) and cardinal direction.

As we can see in table 20 no significant correlation was found (r = .070, p = .209 > 0.05). The result in table 20 shows at a minimum that the differences in RT are not onlyobserved because of the orientation of the pictures.

For the satellite map representation we see in figure 30 a great difference between the error rates of the street label condition for the location task and the orientation task. One possible reason for this finding could still be the orientation of the pictures because it might be more difficult to determine the orientation without distinguishable other environmental features like landmarks or road junctions. This assumption is also supported by the finding that also for the cartographic map representation a difference between location based error rate and orientation based error rate is observed.

4.3 Eye Movement Metrics

In this master thesis the eye movement data was mainly used for the calculation of the RT. A vast variety of different eye movement information was also collected but not analysed. This section only takes a glimpse at the vast amount of eye movement date recorded. The goal of this section will be to answer which features of the experimental display were most fixated. Of interest is especially if there are differences in the fixation count for the picture representation and the map. The fixation count could provide the information if participants spent more time searching in the picture for clues or if they spent more time searching the map representation in the display.

For quality reasons 3 participants were exclude from the eye movement metric calculations mainly due to calibration issues. Figure 32 presents the fixation counts for the pictures in the satellite map condition. The fixation counts are shown separately for each stimulus analysed in this master study.

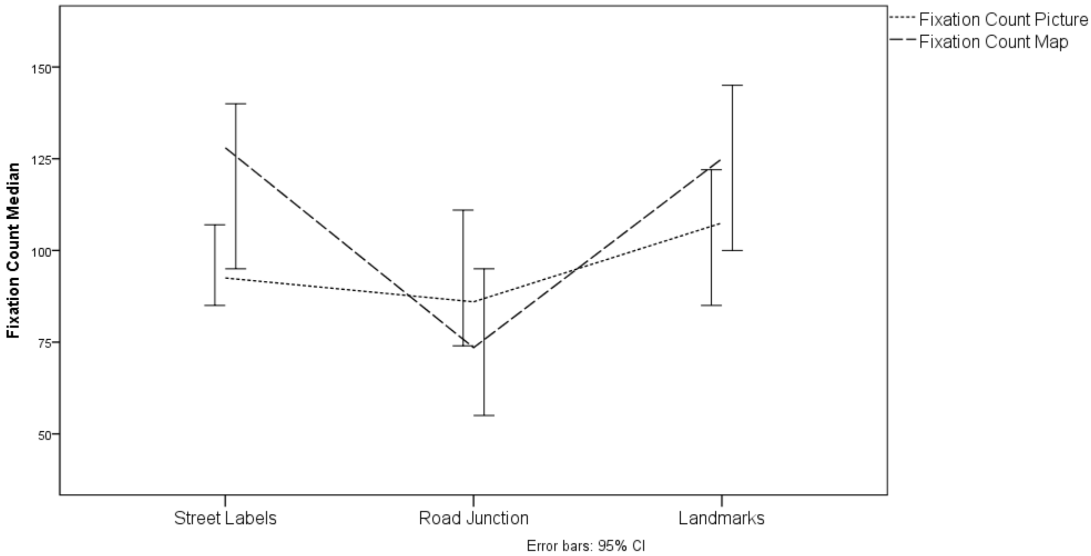


Figure 32: Satellite map fixation count versus picture fixation count.

Figure 32 compares the fixation counts for the picture and the map present in the experimental displays of the satellite map representations. As we can see in figure 32 participants had a higher fixation count value for the map representation than for the picture representing the environmental scene when searching for street labels. The differences for the road junction and the landmark condition seem to be less significant. Nevertheless it seems to be easier to detect road junctions on the map representation than to determine how they look like on a map. The landmark search in a satellite 2D layout map seems to be harder than the identification of possible helpful landmarks in the picture representation. In figure 33 the count of fixations on the map and on the picture is analysed for the cartographic map representation, too.

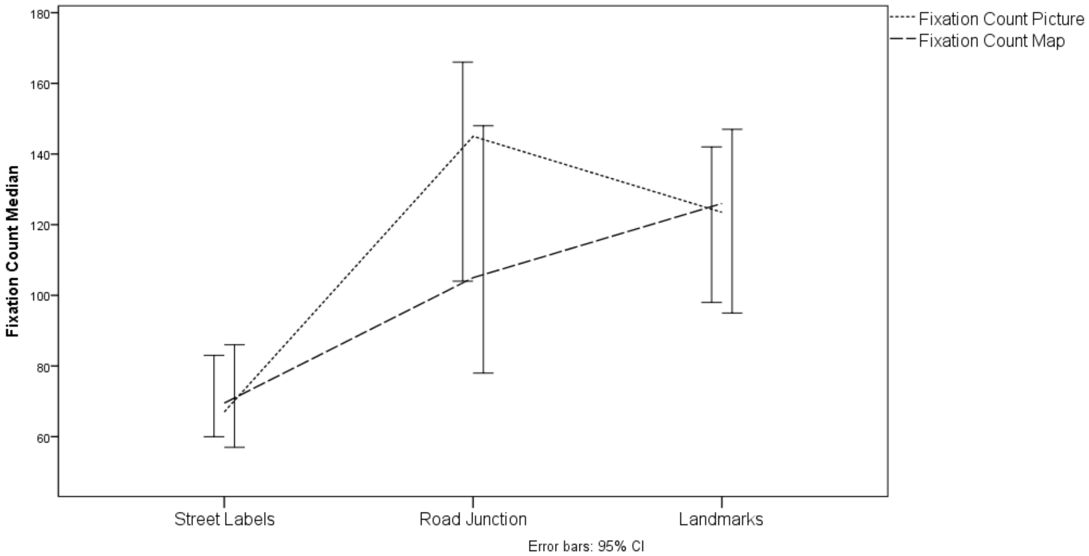


Figure 33: Cartographic map fixation count versus picture fixation count.

Figure 33 shows the differences in fixation counts on the cartographic 2D layout map representation and the picture containing a location also represented in the cartographic 2D layout map. In figure 33 we see that the differences between fixation counts on the picture or the map are not very different for the street label and the landmark condition. A higher difference in fixation counts can be observed for the road junction condition in which the median fixation count was higher on the picture than on the map. This leads to the assumption that the participants searched more in-depth in the picture of the scene because it was harder to imagine the appearance of the road junction in a 2D layout map than to find it on the map itself.

4.3.1 Summary

The analysis of the fixation counts on the picture and the map representations reveals a minor difference for the road junction condition in both map representations. Subjects had a lower fixation count in the satellite map condition for the fixation count of the map compared to the picture whereas in the cartographic map condition subjects had a higher fixation count for the picture compared to the map representation. Additionally, the analysis shows a higher fixation count for the map representation in the satellite map condition for street labels and landmarks compared to the picture. No differences were detected for street labels and landmarks in the cartographic map representations.

4.3.2 Santa Barbara Sense of Direction Scale and Spatial Orientation Test

The Santa Barbara Sense of Direction scale by Hegarty et al. (2002) and the spatial orientation test by Kozhevnikov & Hegarty (2001) were used to determine if differences in spatial orientation abilities have an impact on subjects mean RT in the Google Street View picture to 2D layout map condition. In a first step the self-reported sense of direction test was analysed. In figure 34 the results of the self-reported sense of direction (SOD) test are presented.

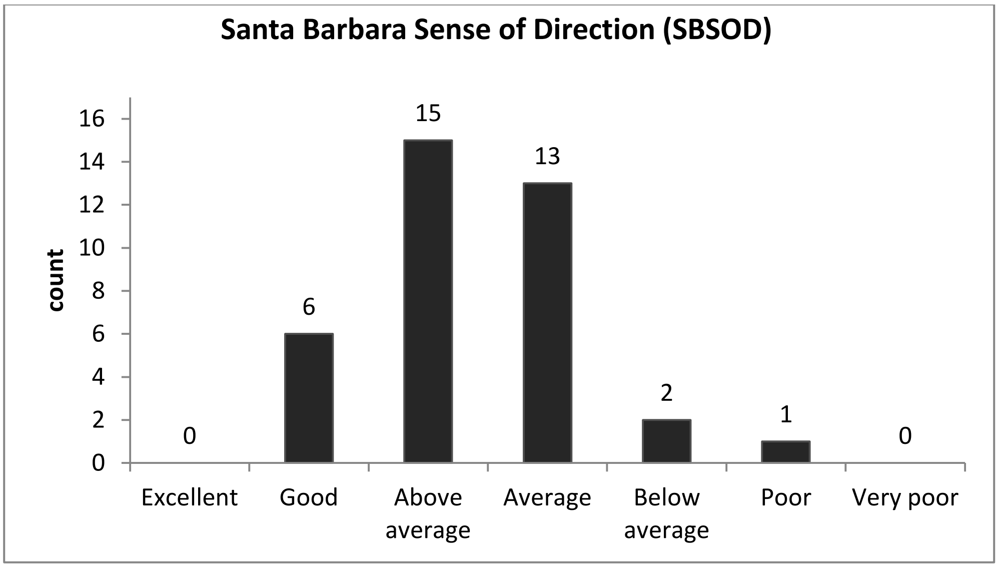


Figure 34: Santa Barbara Sense of Direction scale results.

As we can see in figure 34 the majority of the participants rated their sense of direction as either “Average” or “Above average” (75.7%). The results of the self-reported SOD suggest that the subject pool is very homogeneous in terms of their sense of direction skills. To determine if the results of the SBSOD test have a significant influence on the subjects mean RT measured in the localisation task (picture to 2D layout map representation) a Spearmans' rho correlation was performed. The result is presented in table 21.

Correlations

			Mean Response Time (Location Task)	Santa Barbara Sense of Direction
Spearman's rho	Mean Response Time (Location Task)	Correlation Coefficient	1.000	.097
		Sig. (2-tailed)	.	.570
		N	37	37
	Santa Barbara Sense of Direction	Correlation Coefficient	.097	1.000
		Sig. (2-tailed)	.570	.
		N	37	37

Table 21: Correlation test of RT and SBSOD results.

As we can see in table 21 no significant correlation was found (*r* = .097, *p* = .570 > 0.05). This result was expected because the subject pool is too homogenous. In a next step it was tested if the spatial mental rotation ability has a significant impact on the subjects mean RT results for the picture to 2D layout map search experiment. In figure 35 the histogram of the results of the spatial orientation test is presented.

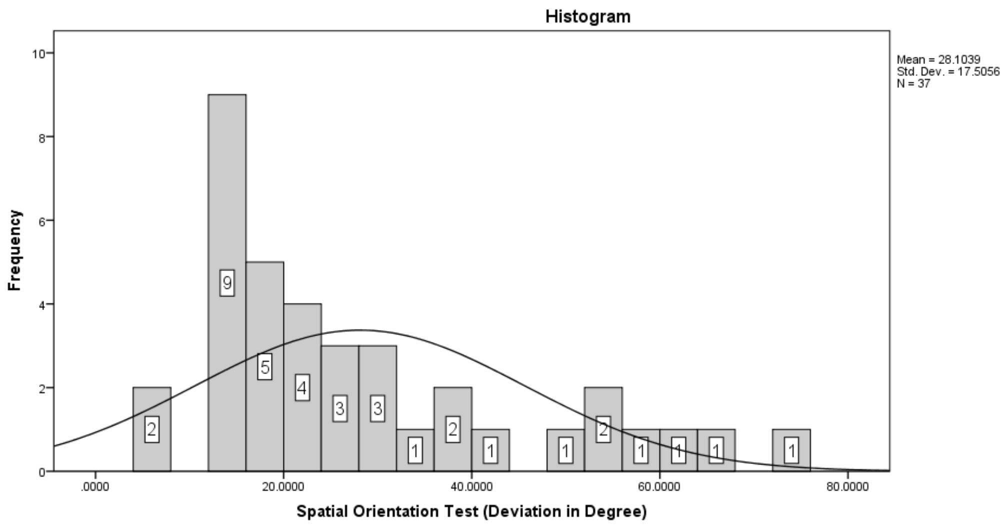


Figure 35: Spatial Orientation Test results represent mean deviation in degrees from the correct answer.

The in figure 35 presented results correspond to the mean deviation in degrees from the correct answer. So if a participant answered all 12 questions absolutely correct (which means drawing the line at exactly the right angel) he or she would have a mean deviation from the correct answer of 0° degrees. As we can see the majority has a deviation from the correct angel of about 20° degrees. To determine if the results of the spatial orientation test correlate with the subjects mean RT measured in the localisation task (picture to 2D layout map representation) a Pearson's correlation was performed. In table 36 the result of the Pearson's correlation test is presented.

Correlations

		Spatial Orientation Test	Mean Response Time (Location Task)
Spatial Orientation Test	Pearson Correlation	1	-.028
	Sig. (2-tailed)		.871
	N	37	37
Mean Response Time (Location Task)	Pearson Correlation	-.028	1
	Sig. (2-tailed)	.871	
	N	37	37

Figure 36: Correlation test of RT and spatial orientation test results.

As we can see in figure 36 no significant correlation was found (r = -.028, p= .871 > 0.05).

So at least for this subject pool the RT results in the "picture to 2D layout map" condition are not dependent on individual differences in sense of direction or spatial orientation abilities.

5. Discussion

In this chapter the results of the user study are summarized and discussed with a focus on the research questions formulated in the introduction of this master thesis.

How do we orientate, navigate and find our way in an unfamiliar environment? The first research question tries to identify the environmental clues which are helping us to orientate while navigating.

RQ1: What environmental features are preferably used during self-localisation?

In communicative route description the most commonly used environmental features are *landmarks*, *pathways* and *choice points* (Allen, 1997). Choice points refer mainly to road junctions where under normal condition the navigator has to decide which way to go. Landmarks and road junctions are of major importance for route descriptions, navigation and for self-localisation (Kiefer et al., 2013; Lovelace et al., 1999; Michon & Denis, 2001; Peters et al., 2010). This was also confirmed in the online survey conducted in this master thesis in which 94% of the subjects asked, said that they are using landmarks to orientate in an unfamiliar environment. Road junctions were considered by 70% of the subjects asked as an important environmental feature for orientation as well. For self-localisation also street labels were considered to be useful for orientation (78%). These three environmental features were therefore selected for a visual search evaluation experiment.

RQ2: How efficient are the in RQ1 defined environmental features for self-localisation on a map?

The efficiency of the in RQ1 defined environmental features street labels, road junctions and landmarks were tested for two different map representations: a cartographic, abstract map representation and a satellite, realistic map representation.

For the cartographic map representation a significant difference in RT was found between the search for street labels and the search for road junctions and landmarks. The RT for the street label search was significantly faster than for the other two environmental features. The same trend was observed for the location error rate which represents the amount of subjects who did not find the location or chose a false location in the search task. For the cartographic map representation a significant increase in the location error rate was observed for the landmark search. The RT for the road junction search was similar to the landmark search RT but the location error rate was significantly lower than for the landmark search.

For the satellite map representation no significant differences in RT between the search for street labels, road junctions or landmarks were found. However a minor decrease in RT could be observed for the road junction search. Noticeable was also the high accuracy in determining the location of the camera in the landmark search conditions.

RQ3: How efficient can the corresponding 2D layout features of the environmental features defined in RQ1 be found on a map?

For the 2D layout environmental feature search the environmental features street labels, road junctions and landmarks were tested in four different combinations each. All in RQ1 defined environmental features were tested in a "stand-alone" condition, in combinations of two's, and in an altogether condition. The combination of all environmental features in one map representation equals the original map representation. This test design was used because it allows testing if one of the three environmental features dominates the other two environmental features. The manifestation of such a dominance could be determined if the RT would be equal for all conditions. This dominance could only be observed in the search for street labels in the cartographic map representation in which no significant differences in RT could be found between the combined conditions street labels, street labels and landmarks, street labels and road junctions, and in the original map representation. In agreement with the assumption formulated before it can be assumed that the search for street labels dominates the other environmental clues present in the cartographic map representation. This observation can most likely be explained with typical map reading strategies focused on street label search. Although street labels seem to dominate the other environmental features in the cartographic map representation, the search for street labels was not the most efficient search. In the cartographic map representation the fastest RTs were measured for landmark search whereas for road junctions the slowest RTs were measured.

In contrast to the cartographic map representation in the satellite map representation road junctions had the fastest RTs whereas the search for street labels was observed to perform poorest.

RQ4: Does the reaction time (RT) needed in RQ3 to find the 2D layout of the environmental features defined in RQ1 on a (road) map correspond to their efficiency as environmental features in a self-localisation task (RQ2)?

For the cartographic map representation the RT for the street label search measured in the 2D layout map representation was significantly slower than the RT for the landmark search whereas for the self-localisation task the street label search was significantly faster than the landmark search RT. So for the cartographic map representation no correlation between the RT needed for self-localisation and the RT needed to find the corresponding environmental feature in the 2D layout map was found. There seems to be though a minor correlation between the RT needed for the search of road junctions in a 2D layout cartographic map and the RT needed for self-localisation. In both tasks the RT in visual search for road junctions was the slowest of the three defined environmental features.

For the satellite map representation the RT was the fastest for the search for road junctions in the 2D layout satellite map. The same trend could be observed for the self-localisation task although no significant differences could be determined. Self-localisation is not only about speed but also about accuracy. In terms of accuracy the search for landmarks in a satellite map representation for a self-localisation task proved to be very satisfactory.

In general an equal RT trend for the 2D layout map search and the self-localisation task was observed for the satellite map representation suggesting a correlation between the search efficiency in a 2D layout map representation and the efficiency for self-localisation of an environmental feature.

RQ5: Are certain environmental features more efficient for abstract cartographic map representations than for satellite map representations and vice versa?

In the cartographic map representation labels could be found faster than in the satellite map representation whereas road junctions could be found faster in the satellite map representation. In contrast to the cartographic map representation it seems to be particularly difficult to search for street labels in the satellite map representation. In general it seems to be an advantage if textual clues are present for the identification of road junctions and landmarks whereas it seems to be a disadvantage for the street label search. The difficulty to search or read labels in complex natural contexts is also pointed out by Bartz (1970) who emphasises the importance of figure-ground relations in text search.

6. Conclusion

The experiment presented in this master thesis provides a connection between important environmental features in a real world environment and their equivalent on a map representation. Although the conducted experiment can certainly be improved in terms of controllability and reliability it provides an interesting competitive approach for the evaluation of labels, road junctions and landmarks in a laboratory environment.

The evaluation of the different perception of environmental features from a birds-eye view or from an in-scene view is a challenging but interesting task. It connects the concepts of mental map and visual scene perception with cartographic design challenges. The in this master thesis presented results support the empirical findings that people can either build up a mental map from graphic representations (survey knowledge) or by an on-the-ground perspective (route knowledge) (Lobben, 2004). The study setting of this master thesis provides a framework which compares the general search efficiency of different environmental features in a 2D layout map with a specific on-the-ground derived perspective of these environmental features in a self-localisation task. The results presented suggest that street labels are the overall most efficient environmental features for self-localisation on a cartographic map representation. Although landmark features could be found very quickly on a 2D cartographic layout map the response time (RT) and the error rate increases in the self-localisation task. This difference points out that a simple comparison without a change of perspective is, without any textural information, not transferable to a self-localisation task. Nevertheless certain trends in RT results were observed for both conditions. Road junctions were for example overall faster found in the satellite map representation for both tasks whereas landmark clues were found to be very accurate for the self-localisation tasks.

The main advantage of the cartographic, abstracted map representation is the high readability of the street labels whereas the textural information in the satellite map representation provides additional information about the environment useful for self-localisation without the street labels present.

7. Further Research

At the beginning of this master thesis a statement from MacEachren and Kraak was quoted:

“A fundamental problem for geovisualization is to understand (and take advantage of) the mechanism by which the dynamic, external visual representations offered by geovisualisation serve as prompts for the creation and use of mental representations (MacEachren & Kraak, 2001:8)”.

The visualisation of information and the question how information is perceived and understood by the user will be a fundamental challenge for some time longer in geovisualisation. The processes involved in building the mental map are complex and still not finally understood. Further research should broaden the knowledge in this research field and hopefully provide new design guidelines to increase the future map reading experience. One important point is also the conduction of real-world experiments, for example with a mobile eye tracking device, equal as it was conducted by Kiefer et al. (2013). This could help to increase the knowledge how environmental features are perceived and used for self-localisation on a map representation. Also a further research area in this context is the evaluation of small displays for example for smart phones and the creation of appropriate solutions for these devices. Some research was already done in this area of research, see for example Giannopoulus, Kiefer, & Raubal (2013) and Ishikawa, Fujiwara, Imai, & Okabe (2008).

I would like to end this master thesis with the start of this master thesis in a slightly altered version:

Whereas “[…] the final test of map reading is the visualization of landscape from map" (Sylvester, 1952:52), the final test for the cartographer is the visualisation of landscape for the mind.

8. References

Allen, G. (1997). From knowledge to words to wayfinding: Issues in the production and comprehension of route directions. *Spatial Information Theory A Theoretical Basis for GIS Lecture Notes in Computer Science, 1329*, 263–372.

Arvor, D., Durieux, L., Andrés, S., & Laporte, M.-A. (2013). Advances in Geographic Object-Based Image Analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, *82*, 125–137. doi:10.1016/j.isprsjprs.2013.05.003

Asher, M. F., Tolhurst, D. J., Troscianko, T., & Gilchrist, I. D. (2013). Regional effects of clutter on human target detection performance. *Journal of vision*, *13*(5), 1–15. doi:10.1167/13.5.25

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*(6), 452–471. doi:10.1037/h0044049

Bartz, B. S. (1970). Experimental Use of the Search Task in an Analysis of Type Legibility in Cartography. *Cartographic Journal, The*, *7*(2), 103–112.

Blades, M., & Spencer, C. (1987). How do people use maps t o navigate through the world? *Cartographica: The International Journal for Geographic Information and Geovisualization*, *24*(3), 64–75. doi:10.3138/815T-6410-3764-7485

Board, C. (1978). Map reading tasks appropriate in experimental studies in cartographic communication. *... : The International Journal for Geographic Information ...*, *15*(1), 1–12.

Boer, A. (2012). *Abstracting the Reality: Usability Evaluation of Levels of Abstraction and Realism in Geographic Visualizations* (p. 123). Geographisches Institut der Universität Zürich.

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, *22*(1), 55–69. doi:10.1109/TIP.2012.2210727

Bravo, M. J., & Farid, H. (2008). A scale invariant measure of clutter. *Journal of vision*, *8*(1), 1–9. doi:10.1167/8.1.23

Brown, B. (2007). Working the problems of Tourism. *Annals of Tourism Research*, *34*(2), 364–383. doi:10.1016/j.annals.2006.09.003

Bunch, R. L., & Lloyd, R. (2006). The Cognitive Load of Geographic Information. *The Professional Geographer*, *58*(2), 209–220. doi:10.1111/j.1467-9272.2006.00527.x

Caduff, D., & Timpf, S. (2008). On the assessment of landmark salience for human navigation. *Cognitive processing*, *9*(4), 249–67. doi:10.1007/s10339-007-0199-2

Cameron, E. L., Eckstein, M., Tai, J., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks — identification, yes/no detection and localization. *Spatial Vision*, *17*(4), 295–325. doi:10.1163/1568568041920212

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision research*, *51*(13), 1484–525. doi:10.1016/j.visres.2011.04.012

Carrasco, M., Talgar, C. P., & Cameron, E. L. (2001). Characterizing visual performance fields: effects of transient covert attention, spatial frequency, eccentricity, task and set size. *Spatial vision*, *15*(1), 61–75.

Carter, J. (2005). The many dimensions of map use. *Proceedings, International Cartographic Conference*.

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*(1), 5–17. doi:10.1559/152304009787340197

Crampton, J. (1992). A Cognitive Analysis of Wayfinding Expertise. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *29*(3), 46–65. doi:10.3138/10TH-4830-2R77-8N21

DeCarlo, D., & Santella, A. (2002). Stylization and abstraction of photographs. *ACM Transactions on Graphics*, *21*(3), 769–776. doi:10.1145/566654.566650

Duchowski, A. T. (2007). *Eye Tracking Methodology*. London: Springer London. doi:10.1007/978-1-84628-609-4

Duncan, G., & Humphreys, J. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of experimental psychology. Human perception and performance*, *18*(2), 578–588.

Eckstein, M. P. (2011). Visual search: a retrospective. *Journal of vision*, *11*(5). doi:10.1167/11.5.14

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed., p. 779). London.

Frankenstein, J., Büchner, S., Tenbrink, T., & Hölscher, C. (2010). Influence of Geometry and Objects on Local Route Choices during Wayfinding. In C. Hölscher, T. Shipley, M. Olivetti Belardinelli, J. Bateman, & N. Newcombe (Eds.), *Spatial Cognition VII SE - 7* (Vol. 6222, pp. 41–53). Springer Berlin Heidelberg. doi:10.1007/978-3-642-14749-4_7

Garlandini, S., & Fabrikant, S. I. S. (2009). Evaluating the effectiveness and efficiency of visual variables for geographic information visualization. (K. Hornsby, C. Claramunt, M. Denis, & G. Ligozat, Eds.)*Spatial Information Theory*, *5756*, 195–211. doi:10.1007/978-3-642-03832-7_12

Giannopoulus, I., Kiefer, P., & Raubal, M. (2013). Mobile Outdoor Gaze-Based GeoHCI. In *GeoHCI Workshop at CHI 2013, April 27-28 2013*. Paris, France.

Golledge, R. G. (1992). Place recognition and wayfinding: Making sense of space. *Geoforum*, *23*(2), 199–214. doi:10.1016/0016-7185(92)90017-X

Harel, J., Koch, C., & Perona, P. (2007). Graph-Based Visual Saliency. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems 19* (pp. 545–552). Cambridge, MA: MIT Press.

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:http://dx.doi.org/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. doi:10.1016/j.intell.2003.12.001

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology*, *28*(1), 74–82. doi:10.1016/j.jenvp.2007.09.002

Itti, L., Dhavale, N., & Pighin, F. (2003). Realistic Avatar Eye and Head Animation Using a Neurobiological Model of Visual Attention. In B. Bosacchi, D. B. Fogel, & J. C. Bezdek (Eds.), *Proc. SPIE 48th Annual International Symposium on Optical Science and Technology* (Vol. 5200, pp. 64–78). Bellingham, WA: SPIE Press.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10-12), 1489–506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *... Analysis and Machine Intelligence, …*, *20*(11), 1254–1259.

Kiefer, P., Giannopoulos, I., & Raubal, M. (2013). Where Am I? Investigating Map Matching During Self-Localization With Mobile Eye Tracking in an Urban Environment. *Transactions in GIS*, n/a–n/a. doi:10.1111/tgis.12067

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & cognition*, *29*(5), 745–56.

Kraak, M.-J., & Ormeling, F. (2003). *Cartography : visualization of geospatial data* (2nd ed., p. 205). Harlow: Prentice Hall.

Lloyd, R. (1997). Visual Search Processes Used in Map Reading. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *34*(1), 11–32. doi:10.3138/F342-1217-827R-0777

Lloyd, R., & Bunch, R. L. (2003). Technology and Map-Learning: Users, Methods, and Symbols. *Annals of the Association of American Geographers*, *93*(4), 828–850. doi:10.1111/j.1467-8306.2003.09304004.x

Lloyd, R., Cammack, R., & Holliday, W. (1995). Learning environments and switching perspectives. *Cartographica: The International …*, *32*(2).

Lloyd, R., & Carolina, S. (2005). Attention on maps. *Cartographic Perspectives*, (52), 28–57.

Lobben, A. K. (2004). Tasks, Strategies, and Cognitive Processes Associated With Navigational Map Reading: A Review Perspective∗. *The Professional Geographer*, *56*(2), 270–281. doi:10.1111/j.0033-0124.2004.05602010.x

Lovelace, K. L., Hegarty, M., & Montello, D. R. (1999). Elements of Good Route Directions in Familiar and Unfamiliar Environments. In C. Freksa & D. Mark (Eds.), *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science SE - 5* (Vol. 1661, pp. 65–82). Springer Berlin Heidelberg. doi:10.1007/3-540-48384-5_5

Lynch, K. (1960). *The Image of the City*. Boston: MIT.

MacEachren, A. M., & Kraak, M.-J. (2001). Research Challenges in Geovisualization. *Cartography and Geographic Information Science*, *28*(1), 3–12. doi:10.1559/152304001782173970

Mack, M. L., & Oliva, A. (2004). Computational estimation of visual complexity. *12th Annual Object, Perception, Attention, and Memory Conference. Minneapolis, Minnesota*.

Michon, P. E., & Denis, M. (2001). When and Why Are Visual Landmarks Used in Giving Directions? In D. Montello (Ed.), *Spatial Information Theory SE - 20* (Vol. 2205, pp. 292–305). Springer Berlin Heidelberg. doi:10.1007/3-540-45424-1_20

Montello, D. R. (2005). Navigation. In P. Shah & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 257–294). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511610448

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision research*, *45*(2), 205–31. doi:10.1016/j.visres.2004.07.042

Nelson, E. S. (1994). Colour Detection on Bivariate Choropleth Maps: The Visual Search Process. *Cartographica*, *31*(4), 33–43. doi:10.3138/02M0-1746-8650-P464

Pazzaglia, F., & Meneghetti, C. (2010). Individual Differences in Spatial Language and Way-Finding: The Role of Cognition, Emotion and Motivation. In C. Hölscher, T. Shipley, M. Olivetti Belardinelli, J. Bateman, & N. Newcombe (Eds.), *Spatial Cognition, VII LNAI 6222* (Vol. 6222, pp. 1–3). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-642-14749-4_1

Peters, D., Wu, Y., & Winter, S. (2010). Testing Landmark Identification Theories in Virtual Environments. In C. Hölscher, T. Shipley, M. Olivetti Belardinelli, J. Bateman, & N. Newcombe (Eds.), *Spatial Cognition VII SE - 8* (Vol. 6222, pp. 54–69). Springer Berlin Heidelberg. doi:10.1007/978-3-642-14749-4_8

Presson, C. C., & Montello, D. R. (1988). Points of reference in spatial cognition: Stalking the elusive landmark*. *British Journal of Developmental Psychology*, *6*(4), 378–381. doi:10.1111/j.2044-835X.1988.tb01113.x

Raubal, M., & Winter, S. (2002). Enriching Wayfinding Instructions with Local Landmarks. In M. Egenhofer & D. Mark (Eds.), *Geographic Information Science SE - 17* (Vol. 2478, pp. 243–259). Springer Berlin Heidelberg. doi:10.1007/3-540-45799-2_17

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of vision*, *7*(2), 17.1–22. doi:10.1167/7.2.17

Sorrows, M., & Hirtle, S. (1999). The Nature of Landmarks for Real and Electronic Spaces. In C. Freksa & D. Mark (Eds.), *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science SE - 3* (Vol. 1661, pp. 37–50). Springer Berlin Heidelberg. doi:10.1007/3-540-48384-5_3

Sylvester, D. (1952, January 1). Map and landscape. Philip and Son.

Taylor, H. A. (2005). Mapping the Understanding of Understanding Maps. In P. Shah & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 295–333). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511610448

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of memory and language*, *35*, 371–391.

Toutenburg, H., & Heumann, C. (2008). *Deskriptive Statistik* (6th ed.). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-77788-5

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological review*, *95*(1), 15–48.

Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. *The American journal of psychology*, *105*(2), 341–362.

Ware, C. (2012). *Information Visualization : Perception for Design* (3rd ed.). Amsterdam: Elsevier.

Wolfe, J. M. (1994a). Guided Search 2.0 A revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202–38. doi:10.3758/BF03200774

Wolfe, J. M. (1994b). Visual search in continuous, naturalistic stimuli. *Vision Research*, *34*(9), 1187–1195. doi:10.1016/0042-6989(94)90300-X

Wolfe, J. M. (2007). Guided Search 4.0. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). Oxford: Oxford University Press.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of experimental psychology. Human perception and performance*, *15*(3), 419–33.

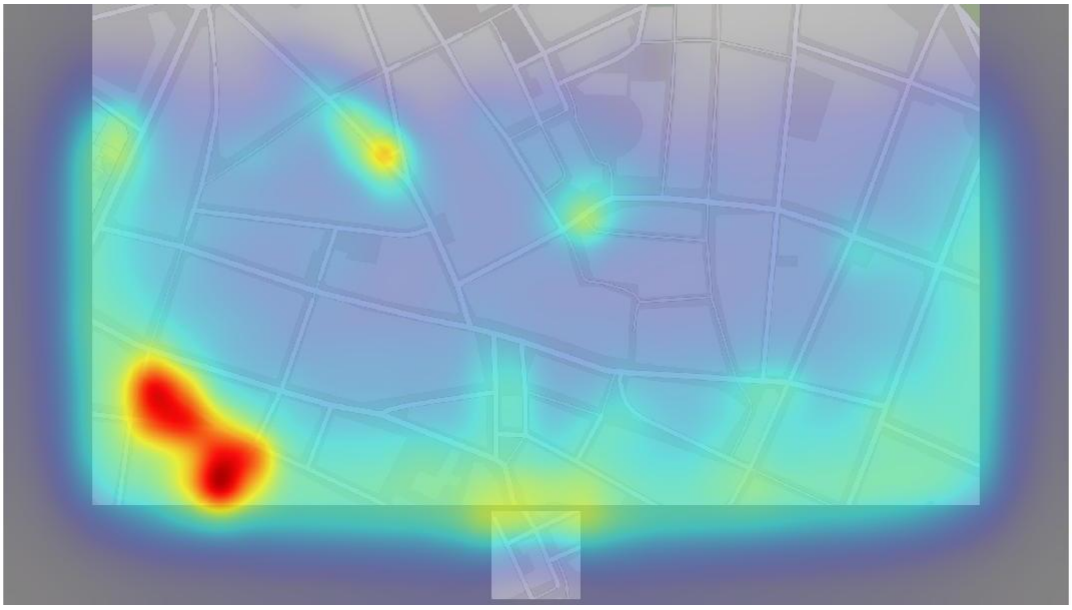
Wolfe, J. M., & Gancarz, G. (1997). Guided Search 3.0. In V. Lakshminarayanan (Ed.), *Basic and Clinical Applications of Vision Science SE - 30* (Vol. 60, pp. 189–192). Springer Netherlands. doi:10.1007/978-94-011-5698-1_30

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature reviews. Neuroscience*, *5*(6), 495–501. doi:10.1038/nrn1411

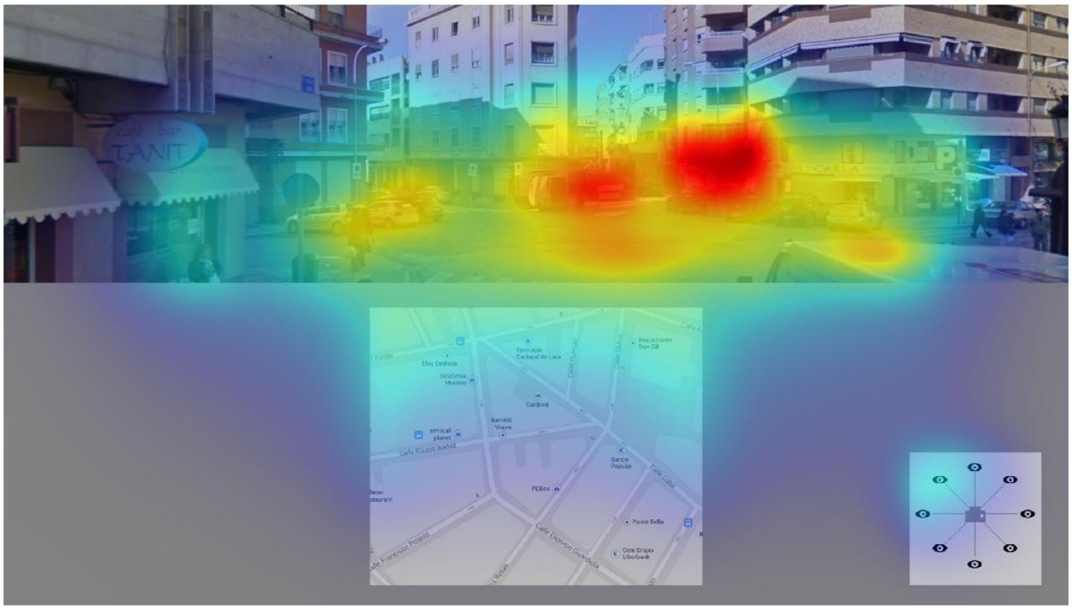
Appendix

A.1 Saliency Maps Examples

2D Layout Maps:



“Picture to 2D layout map”:



A.2 Santa Barbara Sense of Direction Scale:

Appendix B. Santa Barbara Sense of Direction Scale

Sex: F M
Age:_______

Today's Date:_______________
V. 2

This questionnaire consists of several statements about your spatial and navigational abilities, preferences, and experiences. After each statement, you should circle a number to indicate your level of agreement with the statement. Circle "1" if you strongly agree that the statement applies to you, "7" if you strongly disagree, or some number in between if your agreement is intermediate. Circle "4" if you neither agree nor disagree.

1. I am very good at giving directions.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

2. I have a poor memory for where I left things.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

3. I am very good at judging distances.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

4. My "sense of direction" is very good.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

5. I tend to think of my environment in terms of cardinal directions (N, S, E, W).

   strongly agree 1 2 3 4 5 6 7 strongly disagree

6. I very easily get lost in a new city.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

7. I enjoy reading maps.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

8. I have trouble understanding directions.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

9. I am very good at reading maps.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

10. I don't remember routes very well while riding as a passenger in a car.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

11. I don't enjoy giving directions.

strongly agree 1 2 3 4 5 6 7 strongly disagree

12. It's not important to me to know where I am.

strongly agree 1 2 3 4 5 6 7 strongly disagree

13. I usually let someone else do the navigational planning for long trips.

strongly agree 1 2 3 4 5 6 7 strongly disagree

14. I can usually remember a new route after I have traveled it only once.

strongly agree 1 2 3 4 5 6 7 strongly disagree

15. I don't have a very good "mental map" of my environment.

strongly agree 1 2 3 4 5 6 7 strongly disagree

A.3 Spatial Orientation Test:

Spatial Orientation Test

This is a test of your ability to imagine different perspectives or orientations in space. On each of the following pages you will see a picture of an array of objects and an "arrow circle" with a question about the direction between some of the objects. For the question on each page, you should imagine that you are standing at one object in the array (which will be named in the center of the circle) and facing another object, named at the top of the circle. Your task is to draw an arrow from the center object showing the direction to a third object from this facing orientation.

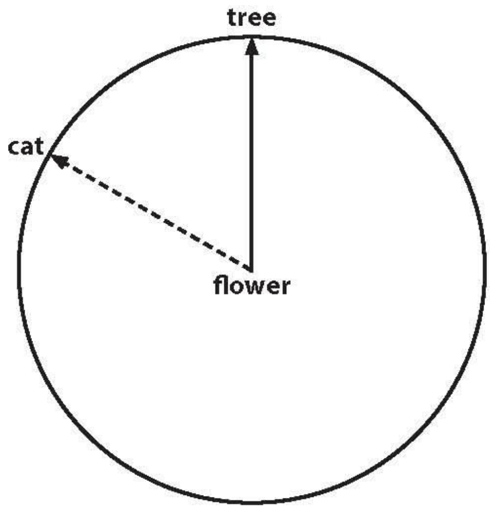
Look at the sample item on the next page. In this item you are asked to imagine that you are standing at the flower, which is named in the center of the circle, and facing the tree, which is named at the top of the circle. Your task is to draw an arrow pointing to the cat. In the sample item this arrow has been drawn for you. In the test items, your task is to draw this arrow. Can you see that if you were at the flower facing the tree, the cat would be in this direction? Please ask the experimenter now if you have any questions about what you are required to do.

There are 12 items in this test, one on each page. For each item, the array of objects is shown at the top of the page and the arrow circle is shown at the botom. Please do not pick up or turn the test booklet, and do not make any marks on the maps. Try to mark the correct directions but do not spend too much time on any one question.

You will have 5 minutes for this test.

Spatial Orientation Test **Name:** ____________________

Example:
Imagine you are standing at the **flower** and facing the **tree**.
Point to the **cat**.



A.4 Consent Form

The University of Zurich - **Participant Information Statement and Consent Form** Map Perception and Orientation: A Study with Eye Movement Analysis
Sep. 30 – Oct. 15, 2013
Participant No:

Purpose of study

You are invited to participate in a study regarding an evaluation of online maps (GoogleMaps). We hope to learn more about the perception of certain map elements (e.g. streets, buildings or texture in general) and their impact on self-location in an unfamiliar environment.

Description of study and risks

If you decide to participate, we will ask you to begin by filling out a short background questionnaire including demographic information. This will be followed by a session at the computer where you will be asked to complete a series of search tasks. During this process we will record your interactions with the computer using a webcam, audio recorder and eye tracking. The eye tracking device is non-contact, uses near infrared light and should not cause any discomfort.

The whole procedure should take approximately 60 minutes and there are no particular risks or benefits to you from participating in this experiment.

Confidentiality and disclosure of information

Any information that can be identified with you in connection with this study will remain confidential and will be disclosed only with your permission. If you give us permission by signing this document, we plan to publish the results of this research in scientific publications. In any publication, information will be provided in such a way that you cannot be identified.

Compensation

We do not provide any compensation for your participation in this experiment, nor are there any costs for you for your participation.

Feedback to participants

If you would like to be kept informed about the results of this research, please leave your name and contact details with the experiment leader. A copy of publications resulting from this research will be sent to you when available.

Your consent

Your decision whether or not to participate will not prejudice your future relations with University of Zurich. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.

If you have any questions, please feel free to ask us. If you have any additional questions later, Dr. Arzu Coltekin (044 6355440, arzu@geo.uzh.ch) or Floris Heim (floris.heim@uzh.ch) will be happy to answer them.

You will be given a copy of this form to keep.

Page 1 of 2

You are making a decision whether or not to participate. Your signature indicates that, having read the information provided above, you have decided to participate.

……………………………………. ………………………………

Signature of Research Participant Signature of Experimenter

……………………………………. ………………………………

Please PRINT name Please PRINT name

…………………………………….

Date and Place

REVOCATION OF CONSENT

Map Perception and Orientation: A Study with Eye Movement Analysis

I hereby wish to WITHDRAW my consent to participate in the research proposal described above and understand that such withdrawal WILL NOT jeopardize any treatment or my relationship with The University of Zurich.

……………………………………. ………………………………

Signature Date

…………………………………….

Please PRINT name

This section of Revocation of Consent should be forwarded to Dr. Arzu Coltekin, Geographic Information Visualization and Analysis, Dept. of Geography, University of Zurich, CH-8057, Zurich.

Personal Declaration:

I hereby declare that the submitted thesis is the result of my own, independent, work. All external sources are explicitly acknowledged in the thesis

Date: Signature: